University of Lille
Doctoral school of Matter, Radiation and Environmental Sciences
Laboratory of Atmospheric Optics

Thesis presented by
Yuyang CHANG

Retrieval of height-resolved aerosol components using high performance multiwavelength Raman-depolarization-fluorescence lidar

Defense on 5th December 2023 to obtain the doctoral degree

Discipline: Sciences of the Earth and the universe, space.
Specialty: Earth, fluid envelopes

Reviewers		
Adolfo COMERON	Professor	Universitat Politècnica de Catalunya, Spain
Iwona STACHLEWSKA (President of the Jury)	Professor	University of Warsaw, Poland
Examinators		
Daniel PEREZ-RAMIREZ	Professor	University of Granada, Spain
Oleg DUBOVIK	Professor	University of Lille, France
Alain MIFFRE	Associated professor	Lyon 1 University, France
Igor VESELOVSKII	Professor	Russian Academy of Sciences, Russia
Inviter		
Sergey KHAYKIN	Professor	Sorbonne University, France
Supervisor		
Philippe Goloub	Professor	University of Lille, France
Co-supervisor		
Qiaoyun Hu	Doctor	University of Lille, France

Université de Lille
École doctorale Sciences de la Matière, du Rayonnement et de l'Environnement
Laboratoire d'Optique Atmosphérique

Thèse présentée par

Yuyang CHANG

Développement d'une nouvelle technique de mesure du profil atmosphérique en aérosols à l'aide d'un lidar Raman-dépolarisation-fluorescence

Soutenue le 5 décembre 2023 pour obtenir le grade de docteur

Discipline : Sciences de la terre et de l'univers, espace
Spécialité : Terre, enveloppes fluides

Rapporteur		
Adolfo COMERON	Professeur	Universitat Politècnica de Catalunya, Espagne
Iwona STACHLEWSKA (Président du Jury)	Professeure	University of Warsaw, Pologne
Examinateur		
Daniel PEREZ-RAMIREZ	Professeur	University of Granada, Espagne
Oleg DUBOVIK	Directeur de recherche	Université de Lille, France
Alain MIFFRE	Maître de conférences	Lyon 1 University, France
Igor VESELOVSKII	Professeur	Russian Academy of Sciences, Russie
Invité		
Sergey KHAYKIN	Professeur	Sorbonne Université, France
Direction de thèse		
Philippe Goloub	Professeur	Université de Lille, France
Co-encadrement		
Qiaoyun Hu	Docteure	Université de Lille, France

Abstract

Vertical information on aerosol optical and microphysical properties is of significant importance to study aerosol evolution, transport, as well as their impacts on human health, local environment and global climate. This thesis developed an algorithm, the Basic algOrithm for REtrieval of Aerosol with Lidar (BOREAL), for retrieving heigh-resolved aerosol microphysical properties from combinations of extinction, backscattering and depolarization lidar measurements. Based on maximum likelihood estimation, the retrieval algorithm uses a nonlinear iteration approach to search for the best fit to both measurements and constraints. The retrieved aerosol microphysical properties include particle size distribution, volume concentration, effective radius, complex refractive index (CRI) and single scattering albedo (SSA).

The performance of BOREAL, retrieval accuracy and measurement sensitivity are assessed through simulated data. In general, retrieval accuracy is higher for fine-mode particles than coarse-mode particles. The simulations demonstrate the importance of exploiting *a priori* constraint to improve the retrieval accuracy of CRI and SSA. Apart from spherical particles, performance of retrieving non-spherical particles is also evaluated by integrating three different particle scattering models, i.e., the Sphere, Spheroid and Irregular-Hexahedral (IH) models, into BOREAL. The results show incorporating depolarization measurements into inversion is essential to better constrain and stabilize the retrieval. Besides, approximating non-spherical particles to spheres will evidently degrade retrieval quality in cases of lidar measurements. In addition, BOREAL is applied to real lidar observations of different aerosol types, including biomass burning, dust and continental polluted aerosols at the ATOLL observatory. Results are analyzed and compared with retrievals from AERONET and previous studies, which demonstrates the robustness of BOREAL for real data application and aerosol characterization.

Overall, this work contributes to Labex CaPPA and ACTRIS efforts to better quantify aerosol microphysical properties using lidar measurements.

Keywords: atmospheric aerosol; remote sensing; inverse problem; lidar measurements; aerosol microphysical properties, BOREAL.

Résumé

La connaissance de la répartition verticale des propriétés optiques et microphysiques des aérosols est cruciale pour étudier l'évolution et le transport des aérosols, ainsi que leurs impacts sur la santé humaine, l'environnement local et le climat mondial. Dans ce travail nous avons développé un algorithme BOREAL pour restituer les propriétés microphysiques des aérosols à partir de combinaisons de mesures lidar d'extinction, de rétrodiffusion et de dépolarisation spectrales. Basé sur une estimation de vraisemblance maximale, l'algorithme de restitution utilise une approche d'itération non linéaire pour rechercher la meilleure adéquation entre les mesures et les contraintes. Les propriétés microphysiques des aérosols restituées comprennent la distribution de taille des particules, leur concentration volumique, leur rayon efficace, l'indice de réfraction complexe (CRI) et l'albédo de diffusion simple (SSA).

Les performances de BOREAL, sa précision et la sensibilité des mesures sont évaluées à l'aide de données simulées. En général, la précision de la restitution est meilleure pour les particules de mode fin que pour les particules de mode grossier. Les simulations démontrent l'importance de l'exploitation de contraintes a priori pour améliorer la précision de la restitution du CRI et du SSA. Outre les particules sphériques, la performance de la restitution des particules non sphériques est également évaluée en intégrant trois modèles de diffusion de particules différents, à savoir les modèles Sphérique, Sphéroïdale et Irrégulier-Hexaédrique (IH), dans BOREAL. Les résultats montrent que l'intégration des mesures de dépolarisation dans l'inversion est essentielle pour mieux contraindre et stabiliser la restitution. De plus, l'approximation des particules non sphériques par des sphères dégrade manifestement la qualité de la restitution. Enfin, BOREAL est utilisé pour restituer les propriétés aérosols au cours d'événements de feux de biomasse, de poussières désertiques et les d'aérosols continentaux pollués détectés depuis la plateforme ATOLL. Les résultats sont analysés et comparés aux restitutions d'AERONET ainsi qu'aux résultats d'études précédentes, ce qui démontrer la robustesse de BOREAL pour l'application de données réelles et la caractérisation d'aérosols.

Ce travail contribue aux études menées dans le cadre du Labex CaPPA et d'ACTRIS en quantifiant les propriétés microphysiques des aérosols à partir des observations lidar.

Mots-clés : aérosols atmosphériques; télédétection; problème inverse; mesures lidar; propriétés microphysiques des aérosols, BOREAL

Acknowledgements

I would like to firstly thank the director of my thesis Philippe Goloub and co-supervisor Qiaoyun Hu not only for their wise guidance and thorough support on my scientific research, but also for their kind help in my administrative procedures and life in France. I'm grateful to Igor Veselovskii and Sergey Khaykin for their valuable comments and opinions on my work during the Comité de Suivi Individuel. I sincerely appreciate the efforts from the thesis reviewers Adolfo Comerón and Iwona Stachlewska, together with the examinators Daniel Pérez-Ramírez, Oleg Dubovik, Alain Miffre and Igor Veselovskii. I'm also thankful to Cheng Chen, Tatsiana Lapionak and Anton Y. Lopatin for their kind advice on my work. I appreciate technical help from the engineers of the Laboratoire d'Optique Atmosphérique (LOA): Fabrice Ducos, François Thieuleux and Jean-Jacques Marécaille. I genuinely thank Anne Priem, Marie-Lyse Liévin, Frédéric Parol, Philippe Dubuisson and Isabelle Jankowiak for their assistance in administrative procedures concerning doctoral funding, travel allowance, thesis application and defense, and laboratory entrance.

In particular, I sincerely thank the director of my master thesis, Jin Hong, from Anhui Institute of Optics and Fine Mechanics of Chinese Academy of Sciences, as well as Yanli Qiao and Zhengqiang Li for leading me to the field of remote sensing and recommending me to continue my career in France. Without their encouragement, I would not proceed this far.

Thanks to my colleagues, friends and fellow students Maria Fernanda Sanchez, William Boissiere, Robin Miri, Raphaël Peroni, Gabriel Chesnoiu, Smita Rani Panda, Mégane Ventura, Athulya Saiprakash, Zhen Liu, Siyao Zhai, Shikuan Jin, Abhinna Behera and Ioana Popovici for all the fun and for making my PhD career worthwhile.

Great thanks to my parents, Gang Chang and Ling Huang. Thanks for your continuous love and support on my oversea study. Special thanks to my girlfriend Shengnan Cao. Thank you for unconditionally supporting my career and releasing my homesick.

I'm also thankful to AERONET for establishing and maintaining the sites and providing the data relevant to this study. Finally, I'm genuinely grateful to the Labex CaPPA and the region Hauts-de-France for funding support.

List of Tables

List of Figures

List of Important Abbreviations

Abbr.	Full Name	Symbol
AOD	Aerosol optical depth	τ
PLDR	Particle linear depolarization ratio	δ_{a} or δ
SSA	Single scattering albedo	ϖ
VSD	Volume size distribution	$v(r)$
ACTRIS	Aerosol-Clouds and Trace gases Research InfraStructure	
AERONET	AErosol RObotic NETwork	
AULSTRAL	AUtomated Server for the TReatment of Atmospheric Lidars	
BBA	Biomass burning aerosol	
BC	Bi-modes with the coarse part dominating	
BD	Bimodal dust	
BF	Bi-modes with the fine part dominating	
BL	Boundary layer	
BOREAL	Basic algOrithm for REtrieval of Aerosol with Lidar	
CRI	Complex refractive index	
DA	Dust aerosol	
TD	Transported dust	
IH	Irregular-Hexahedral	
LILAS	LIlle Lidar Atmospheric Study	
LOA	Laboratoire d'Optique Atmosphérique	
LR	Lidar ratio	
MC	Mono-coarse	
MF	Mono-fine	
TD	Transported dust	
	A priori values of m_{R}, m_{I}	$m_{\mathrm{R,a}}$, $m_{\mathrm{I,a}}$
	A priori standard deviations of m_{R}, m_{I}	$\sigma_{m_{\mathrm{I,a}}}$, $\sigma_{m_{\mathrm{R,a}}}$
	Extinction coefficient	α
	Fitting error	$\varepsilon_{\mathrm{fit}}$
	Fluorescence backscattering coefficient	β_{F}
	Fluorescence capacity	G_{F}
	Geometric standard deviation of a lognormal VSD	σ_{g}
	Imaginary part of complex refractive index	m_{I}
	Mode radius of a lognormal VSD	r_{v}
	Particle effective radius	r_{eff}
	True values of m_{R}, m_{I}	m_{R}^{*}, m_{I}^{*}
	Total volume concentration	V_{t}

Contents

1 Introduction

1.1 Context

By definition, aerosol is a suspension system of solid or liquid particles in the atmosphere (Hinds and Zhu, 2022). Aerosol particles can be produced by both natural processes and anthropogenic activities, with radius ranging from several nanometers to a few tens of micrometers determined by their formation mechanisms (Lenoble et al., 2013). In general, particles with diameters greater than 1 μm (e.g., mineral dust and sea salt) are mostly formed by wind-driven processes and to the contrary, particles with diameters less than 1 μm are usually formed by combustion or chemical conversion of the gaseous precursors (e.g., sulfate, soot and particulate organic matters) (Liou, 2002). Studies on aerosols are provoked by several reasons. Firstly, aerosols can have a negative impact on human health. These particulate matters might contain various bacteria, viruses or carcinogenic constituents, and depending on the size they can be inhaled and populated in different parts of the human body (Elder et al., 2009). Secondly, aerosols play roles in local environment. Anthropogenic secondary sources like ammonia, nitrogen or sulfur oxides are main causes of eutrophication and acid deposition (Smith and Schindler, 2009; Likens et al., 1996). Besides, outbreak of mineral dust or volcanic ash can severely reduce traffic visibility.

What's more important is that aerosol is a crucial atmospheric agent that impacts the Earth radiation budget (ERB), a key driver of water cycle, atmosphere and ocean dynamics and thermodynamics, as well as global climate change (Hansen et al., 2005). The change of top-of-atmosphere (TOA) ERB due to the perturbation of aerosols excluding any long-term radiative response to a change in the global surface air temperature (GSAT) is referred to as the effective radiative forcing (ERF) of aerosols (Boucher et al., 2014; Sherwood et al., 2015). Aerosols can contribute to ERF by aerosol-radiation interaction (ERFari) and aerosol-cloud interaction (EFRaci), both consisting of the instantaneous radiative forcing (IRF) and adjustments, as shown in Figure 1.1. In ERFari, solar radiation is, on the one hand, directly scattered or absorbed by aerosols (instantaneous radiative forcing by aerosol-radiation interactions, or IRFari) and, on the other hand, adjusted due to IRFari-induced changes in clouds, lapse rate and water vapor. In ERFaci, aerosols serve as cloud condensation nuclei (CCN) for water clouds and nucleating particles (INPs) for ice clouds (Ullrich et al., 2017; Mahrt et al., 2018) to change cloud albedo (Twomey, 1959) (instantaneous radiative forcing by aerosol-cloud interactions, or IRFaci), as well as to influence cloud lifetime and thermodynamics (Albrecht, 1989) (adjustments).

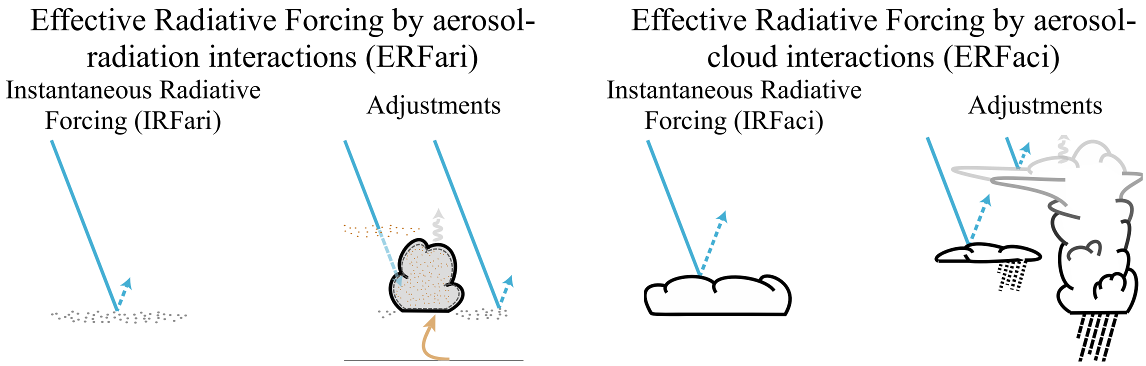


Figure 1.1. Schematic of how aerosols influence Earth radiation budget (ERD) by means of effective radiative forcing (ERF). (Adapted from (Boucher et al., 2014))

Simulation from Earth system models (ESMs) combined with aerosol-chemical transport models is an essential and unique way to estimate total ERF of aerosols (Liou, 2002). At the same time, combined use of ESM simulations and global observations has been proved effective to reduce the estimate uncertainty (Boucher et al., 2014). Aerosol optical and microphysical properties derived from observations and corresponding retrievals, on the one hand, are directly related to TOA, surface energy fluxes and IRF (Ma et al., 2014; Gryspeerdt et al., 2017) and, on the other hand, can be used to constrain, validate and improve ESMs (Lund et al., 2018; Bender et al., 2019; Seifert et al., 2015). Based on agreement between observation-based and modeling-based evidence, the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC-AR6) (Forster et al., 2021) gives virtually certain negative ERF of total aerosols with medium confidence, compared with the most positive ERFs of carbon dioxide and other well-mixed green-house gases (GHGs) with high confidence, as shown in Figure 1.2. The larger uncertainty for aerosols compared to GHGs results from their great variability in tempo-spatial distribution and differences in size, shape, phase and chemical composition. To the contrary, observation- and retrieval- based studies aiming at characterizing aerosol properties and improving understanding of interactions between aerosols and other components, especially in regional scales, are still insufficient (Forster et al., 2021).

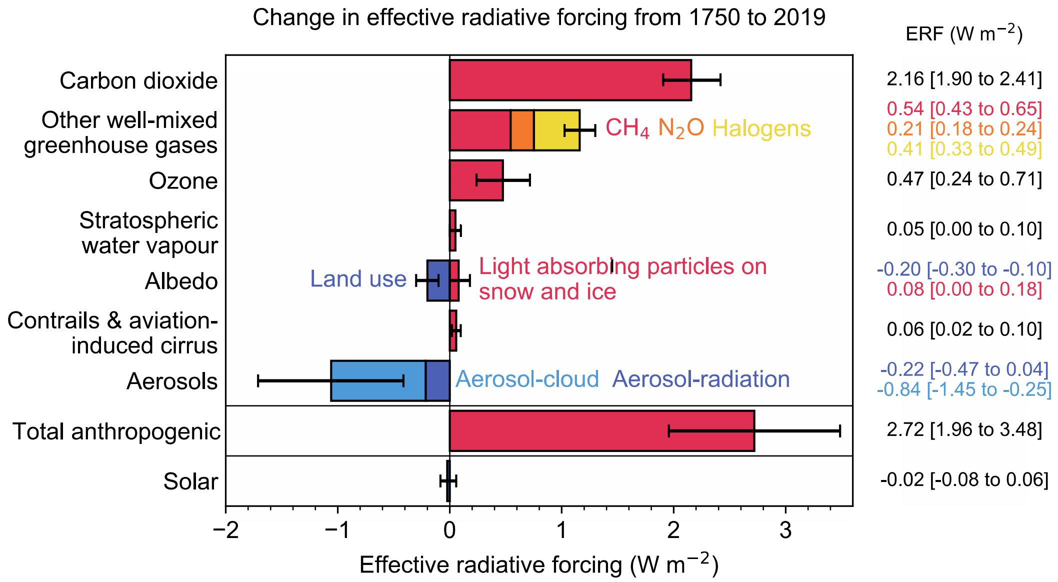


Figure 1.2. Change in effective radiative forcing (ERF) from 1750 to 2019 by contributing forcing agents. Bars represent best estimates and very likely (5-95%) ranges are given by error bars. Aerosols are broken down into contributions from aerosol-cloud interactions (ERFaci) and aerosol-radiation interactions (ERFari). (cited from Forster et al., (2021))

There are two main categories in observation technique, namely in-situ measurements and remote sensing. In-situ measurements collect aerosol samples at the interested point and make direct measurements. The instruments used for in-situ measurements are usually operated in a well-controlled environment and capable of providing rich measurements of aerosol optical and microphysical properties as well as chemical composition. When operated in an aircraft, in-situ measurements can collect aerosol samples at different altitudes so as to obtain vertical profiles (Osborne et al., 2008; Reid et al., 2008). Nevertheless, the sampling and collecting processes of in-situ measurements may alter the state and ambient conditions of the aerosol to be measured (Spanu et al., 2020). Moreover, it is hard for in-situ measurements to provide large spatial coverage. Remote sensing technique detects the radiation interacting with the object without making direct contact. Compared to in-situ measurements, remote sensing does not change the object's original state and realizes continuous observations in a large tempo-spatial scale (Lenoble et al., 2013). According to if an active radiation source is equipped or not, remote sensing can be categorized into active (with the active radiation source) and passive (without the active ration source) types; based on the platform where the remote sensor is operated, it is able to be divided into ground-based and space-borne types. Each type of remote sensing methods has its advantages and disadvantages and different methods are often jointly employed to maximize the information of the measurements.

There have been considerable efforts and collaborations to observe and study aerosol in international, national and regional scales. Space-borne projects and missions, such as the Polar System – Second Generation platform by the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) (Marbach et al., 2015), A-Train by the National Aeronautics and Space Administration (NASA) and its international partners (Schoeberl, 2002), GOSAT series satellites by the Japan Aerospace Exploration Agency (JAXA) and China High-Resolution Earth Observation System by Chinese National Space Administration (CNSA) (Li et al., 2018), have been completed, ongoing or in preparation. For ground-based observations, there have been more extensive collaborations to connect individual observatories as networks. The AErosol RObotic NETwork (AERONET) is a word-wide ground-based remote sensing aerosol network firstly established by NASA and PHOtométrie pour le Traitement Opéraitonnel de Normalisation and Satellitaire (PHOTONS) and developed by multiple international collaborators (Holben et al., 1998). The Aerosol-Clouds and Trace gases Research InfraStructure (ACTRIS) is a pan-European community devoting to developing research infrastructure for short-lived atmospheric constituents [https://www.actris.eu, last access: October 16, 2023]. The European Aerosol Research Lidar NETwork (EARLINET) was founded in 2000 to establish an aerosol climatology for Europe with lidar (abbreviation for LIght Detection And Ranging, which will be described in detail in the following Chapters) (Pappalardo et al., 2014; Wandinger et al., 2016). With expertise in instrumental design, measurement technology, aerosol retrieval methods and atmospheric modeling, the Laboratoire d'Optique Atmosphérique (LOA), joint laboratory of the Centre National de la Recherche Scientifique (CNRS) and University of Lille, France, acts as a major contributor to most of the above-mentioned projects and communities.

1.2 State of the art of aerosol retrieval methods

As mentioned above, retrieval methods are needed for deriving quantitative aerosol properties from observations. A promising retrieval method should make good use of the sensitivities of various measurements, such as wavelength, viewing angle, polarization and depolarization, to different aerosol properties. In this section, we introduce the state of the art of aerosol retrieval methods based on different observational platforms and measuring principles.

1.2.1 Retrieval methods for passive space-borne observations

Characteristics of the space platform (e.g., orbit, maximum loading, inner layout…) necessitate elaborate instrumental design. Retrieval algorithms for space-borne instruments should be customized depending on information provided by the measurements (e.g., spectral, polarimetric, directional and so on) and based on sensitivity study. In general, aerosol properties are retrieved by comparing the measurements with look-up table (LUT) values calculated from pre-defined aerosol models, after some preprocessing steps like cloud-screening or subtraction of surface contribution. Regarding to different surface types, surface contribution may be estimated from models (Lucht et al., 2000), suppressed by selection of measurement wavelengths (Levy et al., 2007), or deducted by making use of multi-directional measurements (Diner et al., 2008). Commonly, the more dimensions of the measurements, the more detailed aerosol retrieval products. For example, total aerosol optical depth (AOD_t) is retrieved from 2 spectral bands of the Advanced Very High Resolution Radiometer (AVHRR) (Smirnov et al., 2002); the fine-mode fraction to AOD_t can be derived if more channels (8 in 0.41-2.13 μm) of the MODerate resolution Imaging Spectroradiometers (MODIS) are exploited (Remer et al., 2005); furthermore, parameterized size distributions of fine- and coarse- mode aerosols are able to be retrieved using multi-angle, multi-spectral polarimetric measurements from the POLarization and Directionality of the Earth's Reflectance (POLDER) instrument (Tanré et al., 2011). In addition, there are also numeric inversion methods with higher complexity for instruments which can take height-resolved measurements such as occultation and limb profilers (McCormick et al., 1979; Rault and Loughman, 2013).

1.2.2 Retrieval methods for ground-based observations

1.2.2.1 Sun-sky photometer

The sun-sky photometer is the basic instrument of AERONET to standardize ground-based aerosol measurements and provide global data validation (Holben et al., 1998). A sun-sky photometer makes two types of measurements by rotating its robotic arms: the direct sun measurement to acquire transmission solar irradiance as well as AOD, and the scanning sky measurement to acquire angular distribution of sky radiance. The accuracies of measured AOD and radiance are ~ 0.01 and ~ 5%, respectively. Both measurements are made in multiple spectral bands from 340 nm to 1020 nm. Some instrumental types also provide polarimetric measurements implemented by a filter wheel containing sets of polarizers. The operational retrieval algorithm of the sun-sky photometer/AERONET is based on the study of Dubovik and King (2000) where the maximum likelihood estimation (MLE) approach is used to simultaneously invert the sun and sky measurements into columnar aerosol microphysical properties including size distribution and spectrally dependent real part (m_R) and imaginary part (m_I) of complex refractive index (CRI). MLE finds the values of the retrieval variables that maximize the likelihood function as the best estimate, which is equivalent to a multi-term least-square problem. Each term has the quadratic form representing difference between the measurements and modeled values weighted by corresponding measurement errors. In this way, different types of measurements with various error levels are considered simultaneously. Apart from the *real* measurements (i.e., sun and sky measurements), the method regards additional smoothing constraints on size distribution and CRI as *virtual* measurements. Correspondingly, the errors of these *virtual* measurements measure the strength of the constraints. In the current operational protocol, the algorithm inverts sun and sky measurements at 440, 670, 870 and 1020 nm. Inversion assumptions include plane-parallel atmosphere, aerosol vertical distributions from MERRA-2 global assimilation (Gelaro et al., 2017), surface BRED from Cox-Munk (Cox and Munk, 1954) and Li-Ross (Lucht et al., 2000) models, water vapor amount from 940 nm channel retrieval and gaseous absorption from climatology data. In addition, aerosol particles are assumed to be composed of spherical particles and randomly orientated spheroidal particles with a fixed axis ratio distribution (ASD) (Volten et al., 2001). The optical properties of the latter are estimated by the Spheroid model (Dubovik et al., 2006) and the fraction of the spherical particles to the total volume concentration, referred as the sphere volume fraction (SVF), is retrieved. Some studies also investigated the feasibility of inverting additional polarimetric data (Li et al., 2009; Fedarenka et al., 2016) and demonstrated that the information content provided by polarimetric measurements can greatly improve the retrieval of fine mode dominated aerosols. However, most polarimetric measurements are conducted in principal planes, which means there are less chances to have quality-assured polarimetric data than radiance and irradiance data (Fedarenka et al., 2016).

1.2.2.2 Lidar

Lidar, the abbreviation of Light detection and ranging, is one of the backbones of active remote sensing for atmospheric profiling, with laser as its emitted source. The instrumental setup and data processing of lidar will be described in more detail in Sect. 2.2. The first application of lidar to atmospheric observations came in 1960s (Fiocco and Smullin, 1963) and since then, with continuous progress in optical and electronic technology, various lidars that profile atmospheric optical properties in layers at different altitudes have been developed. However, it was not until 1990s when technology of simultaneously acquiring aerosol backscattering and extinction coefficients at multiple wavelengths with high accuracy showed up that retrieving aerosol microphysical properties from lidar came to realize (Ansmann et al., 1992; Piironen and Eloranta, 1994). Currently, prevailing lidar-aerosol retrieval algorithms are mainly based on linear inversion with regularization which retrieves size distribution and spectrally independent m_R, m_I. The core idea of linear inversion with regularization is to firstly discretize the size distribution, a continuous function of particle size, to a linear combination of a set of base functions. Correspondingly, relationships between the measurements and coefficients of the base functions can be described with a linear operator (or referred to as the kernel matrix). The coefficients are then solved by implementing operator inversion and regularization is usually needed to deal with ill-posed problems. One representative regularization method is the Twomey-Tikhonov regularization which exploits a regularization term to ensure the smoothness of the retrieved size distribution curve (Müller et al., 1999; Veselovskii et al., 2002). Another regularization method is based on truncated singular value decomposition (TSVD) in which a set of B-splines with variable numbers and orders is used as the base functions. It includes three regularization parameters (*n, d, k*) representing the number, order of the B-splines and the level of the truncation, respectively (Böckmann, 2001; Böckmann et al., 2005). A third one is based on the Landwever iteration using three regularization parameters as well, with the first two the same as those in TSVD, while the last one representing the iteration number (Böckmann and Kirsche, 2006). In general, a complete retrieval process contains three steps: (1) Identification of the searching domain, in which multi-dimensional grids composed of discrete values of variables which impact the kernel matrix, such as m_R, m_I, minimum and maximum particle radii (r_{min}, r_{max}), is constructed. The ranges of the grid values are representative of typical aerosol types; (2) Regularization inversion, in which the linear inversion with regularization is conducted for every grid point of (m_R, m_I, r_{min}, r_{max}) to derive a set of individual solutions; (3) Identification of the solution space, in which a family of individual solutions which minimize the discrepancy between the real and recalculated measurements is determined and designated as the solution space. An averaging process to the solution space can be further performed to derive the final solution from which particle effective radius, volume, surface area and number concentrations are calculated. The minimum measurement requirement for these approaches is 3β (i.e., backscattering coefficients at 355, 532 and 1064 nm) + 2α (i.e., extinction coefficients at 355 and 532 nm). In the past two decades, there were further studies on regularization methods, including retrieving bimodal distributed aerosols (Veselovskii et al., 2004), retrieving non-spherical dust aerosols (Veselovskii et al., 2010; Müller et al., 2013; Böckmann and Osterloh, 2014; Tesche et al., 2019), and efforts to improve the identification of the solution space and automated retrieval (Kolgotin et al., 2016; Müller et al., 2016, 2019). At the same time, one got to realize the limitations of using regularization methods for lidar-aerosol retrieval. First of all, inversion has to be performed for every grid point defined in Step 1, which leads to the retrieval less efficient and moreover, increases difficulty in identifying solution space if the parameters constituting the grid have large ranges and small intervals. The intrinsic reason for the latter is the underdetermination of the $3\beta + 2\alpha$ inversion which often generates numerous individual solutions that can be quite different from each other but reproduce optical properties very close to input measurements (Chemyakin et al., 2016). Although several ways of improving the identification of solution space have been proposed, such as making use of additional constraints on size distribution shape and CRI range (Müller et al., 2014, 2016), accounting for regression relationships for optical properties (Kolgotin et al., 2016, 2018), and some protocols of manually choosing solution space based on previous experiences (Veselovskii et al., 2002; Müller et al., 2016), the regularization method still lacks simple and straightforward manner to incorporate *a priori* constraints from extra sources which help in identifying the solution space. Secondly, they have difficulties in inverting measurements which is not linearly related to the size distribution, for example the particle linear depolarization ratio (PLDR).

Instead of directly retrieving size distribution, the linear estimation (LE) method retrieves bulk properties which can be expressed as analytical functional integrals with the size distribution, such as effective radius and concentrations of different orders (e.g., volume, surface area and number) (Donovan and Carswell, 1997; Graaf et al., 2009; Veselovskii et al., 2012). Like other linear retrieval methods, the first step of the LE method is the discretization and construction of the kernel matrix. Then, spectrally independent CRI can be estimated by minimizing the difference between real and represented measurements. The latter is a linear functional of the real measurements and kernel matrix. Next, the part of the size distribution projected onto the kernel row space is retrieved through principal component analysis, from which bulk properties are finally calculated. The LE method has been proved to be effective in resisting measurement noise and fast to invert massive lidar time series into particle bulk properties even if the measurements are reduced (e.g., $3\beta + 1\alpha$) (Veselovskii et al., 2012). However, a main issue is to estimate the influence of the non-retrievable part of the size distribution (i.e., the part perpendicular to the kernel row space) on retrieval accuracy.

In addition, there are also studies focusing on retrieving CRI, effective radius and particle concentrations with LUT methods (Chemyakin et al., 2014).

1.2.3 Other retrieval methods

1.2.3.1 Retrieval from combined observations of multiple instruments

It is a quite natural thought that combining multiple sources of observations provided by various types of instruments could extend the information content of the retrieval. Progress in combined retrieval is attributed to two aspects. On the one hand, development of joint-observational platforms, including satellites loaded with multiple sensors (e.g., Terra with MODIS and MISR), constellations of satellites following the same orbit (e.g., A-Train) and ground-based networks (e.g., AERONET and EARLINET) greatly facilitates the observations of the same scenes. On the other hand, appearance of advanced retrieval algorithms, like the Generalized Retrieval of Aerosol and Surface Properties (GRASP) algorithm (Dubovik et al., 2021), Lidar/Radiometer Inversion Code (LiRIC) algorithm (Chaikovsky et al., 2016) and Generalized Aerosol Retrieval from Radiometer and Lidar Combined data (GARRLiC) algorithm (Lopatin et al., 2013) provides a generalized way to incorporate multi-source data characterized by different measurement errors into the inversion scheme. Basically, there are three categories of combined retrievals: (1) joint retrieval, i.e., the *a priori* constraints on a retrieval from measurements of one instrument come from retrieval products of another instrument which went through the identical measurement scene (e.g., Chaikovsky et al., 2016); (2) synergy retrieval, i.e., multiple sources of measurements are inverted simultaneously with inversion algorithms being of high complexity (e.g., Lopatin et al., 2013; Dubovik et al., 2021) and (3) assimilation, i.e., remote sensing observations combine with climate models (e.g., Chen et al., 2022).

1.2.3.2 Retrieval based on artificial neural network (ANN)

With more and more remote sensing instruments that have being put into use, the amounts of remote sensing data and meteorological data are rapidly increasing and accumulating. On the other hand, the information content of a single measurement might be not enough or the underlying physical process is still not fully understood so that it is difficult to “mathematically retrieve” the interested parameters. Therefore, the artificial neural network (ANN) approach is getting more and more attention in the field of remote sensing. ANN is a powerful tool to process large amounts of data when the relationship between the input and output is difficult to be described by specific mathematical models. It is a subfield of machine learning and is made up of node layers, including the input, output and hidden layers (Figure 1.3). Each node is an artificial neuron with weight and threshold values and connects to each neuron in the next layer. By "learning" the patterns' features from the training datasets, the ANN adjusts the weights between neuron connections based on the amount of error in the output compared to the expected result, so as to improve the model (Nielsen, 2015).

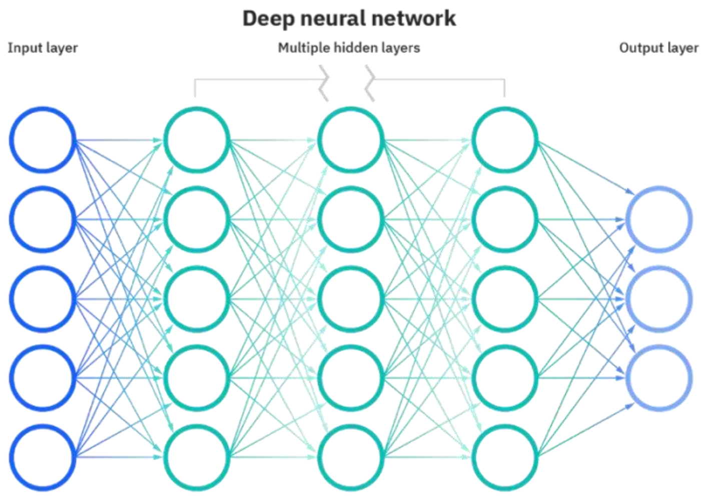


Figure 1.3. Schematic of the structure of an artificial neural network (ANN). An ANN with more than 3 layers (including the input and output layers) is called as the deep neural network (https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/#, last access: October 16, 2023)

Over the last decades, ANNs have been used in analysis of data from satellites (Ali et al., 2013), radars (Orlandini and Morlini, 2000), microwave radiometers (Roberts et al., 2010), nephelometers (Berdnik and Loiko, 2016), multiangle spectropolarimeters (Di Noia et al., 2015) and multi-source datasets (Gupta and Christopher, 2009). With regard to lidar application, Nicolae et al. (2018) proposed the Neural network Aerosol Typing Algorithm based on LIdar data (NATALI) algorithm based on three ANN architectures and aiming at retrieving aerosol type from multiwavelength Mie-Raman-depolarization lidar measurements. The datasets for training, testing and validating the ANNs were generated from a pre-defined aerosol model specifying the shape and microphysical properties of a pure aerosol type and the mixing rules of multiple types. NATALI follows two retrieval schemes: (a) a high-resolution scheme allowing the identification of 14 aerosol mixtures if the depolarization measurement is available and otherwise, (b) a low-resolution scheme allowing the identification of 5 predominant aerosol types.

1.3 Objectives and layout of the thesis

Continuous monitoring of vertical distributions of aerosols is of significant importance to understand their transport as well as tempo-spatial evolution, which in turns reduces the uncertainty of estimating regional and global aerosol climate effect. The main objective of the thesis is to improve the characterization of aerosol vertical properties using measurements from a state-of-the-art multi-wavelength Mie-Raman depolarization fluorescence lidar, LIlle Lidar Atmospheric Study (LILAS), developed and operated by the LOA as part of the ACTRIS research infrastructure. To achieve this goal, a lidar-aerosol retrieval algorithm is needed to derive height-resolved microphysical properties of atmospheric aerosols, such as volume size distribution (VSD), total volume concentration (V_t), effective radius (r_{eff}) and complex refractive index ($\mathrm{CRI} = m_R - im_I$). Because of the night working condition for the Raman channels which prevents from synergy retrieval with passive remote sensors using GARRLiC/GRASP (Sect. 1.2.3), and limits of previous linear lidar stand-alone retrieval algorithms (Sect. 1.2.2.2), this study develops a novel lidar-aerosol retrieval algorithm, the Basic algOrithm for REtrieval of Aerosol with Lidar (BOREAL), which is based on a statistical optimal method – maximum likelihood estimation (MLE). A highlight of BOREAL is that it is able to account for various a priori constraints, types of measurements and different forward models in a simple and straightforward way. We make extensive and comprehensive tests of algorithmic performance using synthetic aerosol models that mimic real aerosol species. In particular, the capability of retrieving mineral dust aerosol is studied by combining BOREAL with different particle scattering models. We also apply BOREAL to different aerosol events observed by LILAS during field campaigns and regular operations in Lille, including biomass burning aerosols (BBAs), dust aerosols (DAs) and continental polluted aerosols. Furthermore, we seal the code as an independent module and integrate it into the AUtomated Server for the TReatment of Atmospheric Lidars (AUSTRAL) platform, which render BOREAL able to process massive data not only from LILAS but also from various other lidars in an efficient, unsupervised and automated way.

This thesis consists of six chapters. In Chapter 1, context of aerosol study is firstly introduced, followed by a review on the state of the art of aerosol retrieval algorithms according to different types of remote sensing techniques and an explanation of the thesis' objectives. In Chap. 2, fundamental backgrounds about basic theory and instrumentation are given. The

Theory section describes modeling of basic processes of aerosol-radiation interaction based on which introductions of specific scattering models and properties of typical aerosol species are provided. The *Instrumentation* section briefly describes the LILAS system and data processing related to the thesis. In Chap. 3, a detailed description of BOREAL's mathematical principle and algorithmic implementation is presented. In Chap. 4, the performance of BOREAL is extensively tested by sensitivity study and inverting synthetic data. In particular, the feasibility of retrieving non-spherical particles like mineral dust is evaluated by combining BOREAL with different particle scattering models. In Chap. 5, BOREAL is applied to retrieve different aerosol events, followed by discussions of the retrieval results. Finally, in Chap. 6, conclusions are drawn and perspectives are proposed.

2 Fundamental backgrounds

In this chapter, backgrounds related to the thesis are presented. It starts with basic theory of interactions between particles and electromagnetic waves, from which variables characterizing optical and microphysical properties of a single particle and particle ensembles are given. Then, particle scattering models used by BOREAL as forward models are introduced. The second part of this chapter describes the lidar system, including basic instrumental setup, derivation of the lidar equation and retrieval of optical profiles based on elastic, Raman, depolarization and fluorescence lidars. In particular, we introduce the LILAS system, of which the optical data serve as the source of the input of BOREAL.

2.1 Theory – interactions of atmospheric particles with radiation

2.1.1 Scattering by a single particle

According to the Maxwell Equations, the electric field of a plane harmonic wave propagating in vacuum along the z-axis of the Cartesian coordinate system, at position z and time t, can be expressed as

$$\mathbf{E} = \mathbf{E}_0 \exp(ikz - i\omega t)\,, \tag{2.1}$$

where i is the imaginary unit, $k = 2\pi/\lambda$ the wavenumber in vacuum determined by wavelength λ, ω the angular frequency and $\mathbf{E}_0$ the complex amplitude vector. The vibrational direction of an electric field vector, $\mathbf{E}$, is always orthogonal to its propagating direction, which means for a reference plane determined by the propagating direction and an arbitrary orthogonal direction, it can be decomposed to components parallel (E_l) and perpendicular (E_r) to the reference plane:

$$E_\mathrm{l} = a_\mathrm{l} \cos(\xi + \delta_\mathrm{l})\,, \tag{2.2}$$

$$E_\mathrm{r} = a_\mathrm{r} \cos(\xi + \delta_\mathrm{r})\,, \tag{2.3}$$

where $\xi = kz - \omega t$. Cancelling ξ gives

$$\left(\frac{E_\mathrm{l}}{a_\mathrm{l}}\right)^2 + \left(\frac{E_\mathrm{r}}{a_\mathrm{r}}\right)^2 - 2\frac{E_l}{a_\mathrm{l}}\frac{E_\mathrm{r}}{a_\mathrm{r}}\cos\delta = \sin^2\delta, \tag{2.4}$$

an ellipse where the phase difference is $\delta = \delta_{\mathrm{l}} - \delta_{\mathrm{r}}$. As shown in Figure 2.1, the polarization state $(\delta, a_{\mathrm{r}}/a_{\mathrm{l}})$ can be uniquely described by the orientation angle χ and ellipticity $\tan\vartheta = b/a$ of the ellipse like

$$\begin{aligned} &\frac{a_{\mathrm{r}}}{a_{\mathrm{l}}} = \tan\varrho\,, \\ &\tan 2\chi = \tan 2\varrho \cos\delta\,, \\ &\sin 2\vartheta = \sin 2\varrho \sin\delta\,. \end{aligned} \tag{2.5}$$

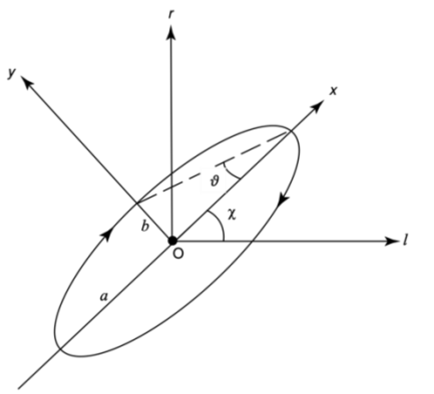


Figure 2.1. Geometric representation of elliptical polarization of which the propagating direction is into the paper (adapted from Liou (2002)).

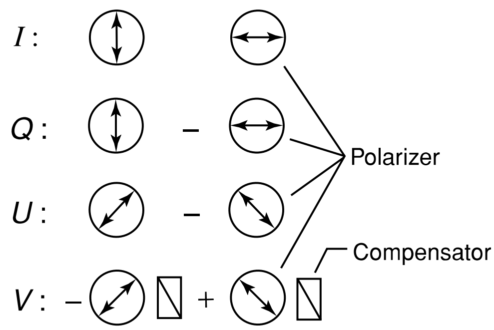


Figure 2.2. Combinations of polarizers and compensators to measure the Stokes components (cited from Liou (2002))

The polarization ellipse represents the complex electric vector with the orientation angle χ and ellipticity which are easier to measure. Another representation preferable in the field of remote sensing is the Stokes vector, which is composed by four real quantities defined as

$$\begin{aligned} &I = E_{\mathrm{l}}E_{\mathrm{l}}^{*} + E_{\mathrm{r}}E_{\mathrm{r}}^{*} = a_{\mathrm{l}}^{2} + a_{\mathrm{r}}^{2}, \\ &Q = E_{\mathrm{l}}E_{\mathrm{l}}^{*} - E_{\mathrm{r}}E_{\mathrm{r}}^{*} = a_{\mathrm{l}}^{2} + a_{\mathrm{r}}^{2}, \\ &U = E_{\mathrm{l}}E_{\mathrm{r}}^{*} + E_{\mathrm{r}}E_{\mathrm{l}}^{*} = 2a_{\mathrm{l}}a_{\mathrm{r}}\cos\delta\,, \\ &V = i(E_{\mathrm{r}}E_{\mathrm{l}}^{*} - E_{\mathrm{l}}E_{\mathrm{r}}^{*}) = 2a_{\mathrm{l}}a_{\mathrm{r}}\sin\delta\,. \end{aligned} \tag{2.6}$$

For a plane wave, each component of the Stokes vector has the unit of irradiance. Making use of Eq. (2.5), the Stoke components can also be written as

$$\begin{aligned} &I^{2} = Q^{2} + U^{2} + V^{2}, \\ &Q = I\cos 2\beta \cos 2\chi\,, \\ &U = I\cos 2\beta \sin 2\chi\,, \\ &V = I\sin 2\beta\,, \end{aligned} \tag{2.7}$$

which clearly shows that I – the total irradiance of an arbitrary elliptically polarized light can be decomposed to the irradiances of: Q – linear polarized light in the plane 0 or 90° with respect to the reference plane; U – linear polarized light in the plane 45° or 135° with respect to the reference plane; and V – right- or left-handed circular polarized light. In practice, the Stokes vector can be measured with combinations of polarizers and compensators, as shown in Figure 2.2.

Another advantage of employing the Stokes vector in remote sensing application is that partially polarized light, which is often dealt with in this field, can be conveniently expressed with it as

$$[I, \quad Q, \quad U, \quad V]_{\mathrm{parp}}^{T} = [I, \quad 0, \quad 0, \quad 0]_{\mathrm{unp}}^{T} + [I, \quad Q, \quad U, \quad V]_{\mathrm{p}}^{T}, \tag{2.8}$$

where the first and second terms on the right-hand side represent the unpolarized and polarized components, respectively.

Since matter is composed of discrete electric charges: electrons and protons, when an obstacle is illuminated by an electromagnetic wave, electric charges in the obstacle oscillate with it and radiate electromagnetic energy in all directions. Such process is referred to as "scattering". Figure 2.3 shows the relation between the incident electronic field $\mathbf{E}_{\mathrm{inc}}$ and the scattered field $\mathbf{E}_{\mathrm{sca}}$. The reference frame is determined by the origin at the particle center and base unit vectors $\hat{\mathbf{e}}_x, \hat{\mathbf{e}}_y, \hat{\mathbf{e}}_z$. The propagating directions of $\mathbf{E}_{\mathrm{inc}}$ and $\mathbf{E}_{\mathrm{sca}}$ coincide with $\hat{\mathbf{e}}_z$ and $\hat{\mathbf{e}}_r$, respectively. If one denotes the components of $\mathbf{E}_{\mathrm{inc}}$ parallel and perpendicular to the scattering plane, s, namely the plane determined by $\hat{\mathbf{e}}_z$ and $\hat{\mathbf{e}}_r$, as $\mathbf{E}_{\parallel\mathrm{inc}} = E_{\parallel\mathrm{inc}}\hat{\mathbf{e}}_{\parallel s}$ and $\mathbf{E}_{\perp\mathrm{inc}} = E_{\perp\mathrm{inc}}\hat{\mathbf{e}}_{\perp s}$, and likewise, denotes the components of $\mathbf{E}_{\mathrm{sca}}$ parallel and perpendicular to the scattering plane as $\mathbf{E}_{\parallel\mathrm{sca}} = E_{\parallel\mathrm{sca}}\hat{\mathbf{e}}_{\parallel s}$ and $\mathbf{E}_{\perp\mathrm{sca}} = E_{\perp\mathrm{sca}}\hat{\mathbf{e}}_{\perp s}$, respectively, due to the linearity of the boundary conditions, the relation between the incident and scattered fields can be expressed as

$$\begin{bmatrix} E_{\parallel\mathrm{sca}} \\ E_{\perp\mathrm{sca}} \end{bmatrix} = \frac{\exp(ikR - ikz)}{-ikR} \mathbf{S} \begin{bmatrix} E_{\parallel\mathrm{inc}} \\ E_{\perp\mathrm{inc}} \end{bmatrix}, \tag{2.9}$$

where R is the distance from the scattered field to the particle center and $\mathbf{S}$ the 2×2 amplitude scattering matrix (Bohren and Huffman, 2004). The equation holds for the far-field region ($kR \gg 1$).

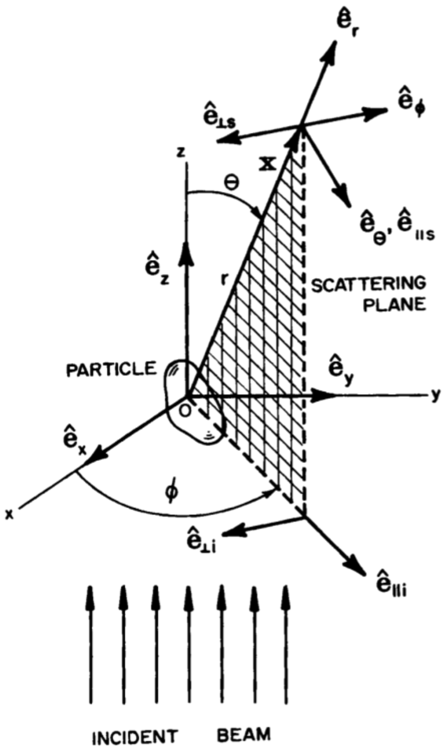


Figure 2.3. Geometry of scattering by an arbitrary particle (cited from Bohren and Huffman [1983])

The scattering process of a single particle can also be expressed in a form of Stokes components as

$$\begin{bmatrix} I_{\mathrm{sca}} \\ Q_{\mathrm{sca}} \\ U_{\mathrm{sca}} \\ V_{\mathrm{sca}} \end{bmatrix} = \frac{1}{k^2R^2}\mathbf{F}\begin{bmatrix} I_{\mathrm{inc}} \\ Q_{\mathrm{inc}} \\ U_{\mathrm{inc}} \\ V_{\mathrm{inc}} \end{bmatrix}, \tag{2.10}$$

where $\mathbf{F}$ is the 4×4 transformation matrix (Liou, 2002) equivalent to $\mathbf{S}$. For a spherical particle, $\mathbf{F}$ is expressed as

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} & 0 & 0 \\ F_{12} & F_{11} & 0 & 0 \\ 0 & 0 & F_{33} & F_{34} \\ 0 & 0 & -F_{34} & F_{33} \end{bmatrix} \tag{2.11}$$

where

$$F_{11} = \frac{1}{2}(\{\mathbf{S}\}_{11}^2 + \{\mathbf{S}\}_{22}^2), \qquad F_{12} = \frac{1}{2}(\{\mathbf{S}\}_{11}^2 - \{\mathbf{S}\}_{22}^2),$$
$$F_{33} = \frac{1}{2}(\{\mathbf{S}\}_{11}^*\{\mathbf{S}\}_{22} + \{\mathbf{S}\}_{11}\{\mathbf{S}\}_{22}^*), \quad F_{34} = \frac{i}{2}(\{\mathbf{S}\}_{22}\{\mathbf{S}\}_{11}^* - \{\mathbf{S}\}_{11}\{\mathbf{S}\}_{22}^*).$$

The elements of $\mathbf{F}$ are real dimensionless numbers depending on the λ of incident light, scattering angle Θ, and particle intrinsic properties such as size, shape and complex refractive index (CRI). The CRI is a wavelength-dependent dimensionless complex quantity determining the scattering properties of the particle. Its imaginary part $m_{\mathrm{I}}(\lambda)$ is related to particle absorption and for a non-absorbing particle, the real part $m_{\mathrm{R}}(\lambda)$ is the ratio of the light phase velocity in ambient medium to the group velocity in the particle. For a non-spherical particle, **F** also depends on particle orientation and if the particle is not rotational symmetric, the azimuth angle ϕ has to be accounted for. In this study, we employ the normalized form, referred to as the phase matrix:

$$\mathbf{P} = \frac{1}{Ck^2R^2}\mathbf{F} \tag{2.12}$$

with the normalization coefficient

$$C = \frac{1}{4\pi k^2R^2}\int_{4\pi} F_{11}\mathrm{d}\Omega \tag{2.13}$$

so that $\int_{\Omega} P_{11}\mathrm{d}\Omega = 4\pi$.

The reduction of electromagnetic flux due to the presence of a particle along the path is called extinction, denoted as Φ_{ext}. If the particle is in a non-absorbing medium, Φ_{ext} equals the sum of scattering Φ_{sca} and absorption Φ_{abs} (Bohren and Huffman, 2004). The extinction, absorption and scattering cross sections are defined as

$$C_{\mathrm{ext}} = \frac{\Phi_{\mathrm{ext}}}{I_{\mathrm{inc}}}, \quad C_{\mathrm{abs}} = \frac{\Phi_{\mathrm{abs}}}{I_{\mathrm{inc}}}, \quad C_{\mathrm{sca}} = \frac{\Phi_{\mathrm{sca}}}{I_{\mathrm{inc}}}, \tag{2.14}$$

respectively. The cross sections depend on not only intrinsic properties of the particle, but also the polarized state of incident light. In this study, we only consider unpolarized light (e.g., solar radiation) and linearly polarized light (e.g., laser light). From Figure 2.3 and Eq. (2.9), for a linear-polarized incident light beam whose vibrational direction is parallel to $\hat{\mathbf{e}}_x$, the scattered irradiance, $I_{\mathrm{sca}}(\Theta,\phi)$, in any direction is

$$\begin{aligned} I_{\mathrm{sca}}(\Theta,\phi) &= \frac{1}{k^2R^2}|\mathbf{d}|^2 I_{\mathrm{inc}}, \\ \mathbf{d} &= (\{\mathbf{S}\}_{11}\cos\phi + \{\mathbf{S}\}_{12}\sin\phi)\hat{\mathbf{e}}_{\parallel s} + (\{\mathbf{S}\}_{21}\cos\phi + \{\mathbf{S}\}_{22}\sin\phi)\hat{\mathbf{e}}_{\perp s}. \end{aligned} \tag{2.15}$$

Therefore, by definition, the scattering cross section can be expressed as

$$C_{\mathrm{sca,lin}} = \int_0^{2\pi}\int_0^{\pi}\frac{|\mathbf{d}|^2}{k^2}\sin\Theta\,\mathrm{d}\Theta\,\mathrm{d}\phi = 2\pi\int_0^{\pi}\frac{F_{11}}{k^2}\sin\Theta\,\mathrm{d}\Theta\,, \tag{2.16}$$

where the subscript "lin" denotes "linear-polarized" and in the last equation the relation $F_{11} = (|\{\mathbf{S}\}_{11}|^2 + |\{\mathbf{S}\}_{12}|^2 + |\{\mathbf{S}\}_{21}|^2 + |\{\mathbf{S}\}_{22}|^2)/2$ is used. By selecting $\hat{\mathbf{e}}_{\mathrm{x}}$, Eq. (2.16) holds for linearly polarized light in any direction. For unpolarized incident light, by making use of Eq. (2.10), the scattering cross section, $C_{\mathrm{sca,unp}}$, can be expressed as

$$C_{\mathrm{sca,unp}} = \int_{4\pi}\frac{F_{11}}{k^2}\mathrm{d}\Omega = 2\pi\int_0^{\pi}\frac{F_{11}}{k^2}\sin\Theta\,\mathrm{d}\Theta\,. \tag{2.17}$$

Thus, the scattering cross sections for linear-polarized and unpolarized light beams are the same. It also can be proved that $C_{\mathrm{ext,unp}} = C_{\mathrm{ext,lin}}$ and $C_{\mathrm{abs,unp}} = C_{\mathrm{abs,lin}}$ (Bohren and Huffman, 2004). These important conclusions mean that particle scattering models developed for unpolarized incident light (e.g., sun-sky photometer measurements) can be directly used in the situations of linear-polarized incident light (e.g., lidar measurements).

By comparing Eq. (2.17) with Eq. (2.12), (2.13), we have

$$\begin{aligned} &C = \frac{C_{\mathrm{sca}}}{4\pi R^2}, \\ &\frac{C_{\mathrm{sca}} P_{11}(\Theta)}{4\pi} = \frac{F_{11}(\Theta)}{k^2}, \\ &\int_{4\pi} P_{11}(\Theta) \mathrm{d}\Omega = 4\pi, \end{aligned} \tag{2.18}$$

where $P_{11}(\Theta)$ is also referred to the phase function and describes the angular distribution of the scattered energy; $\frac{C_{\mathrm{sca}} P_{11}(\Theta)}{4\pi}$ is called the differential scattering cross section and indicates the scattering irradiance in a unit solid angle along the direction, Θ. Furthermore, to qualitatively describe the angular distribution of the scattered energy, the asymmetry factor, g, is defined as

$$g = \frac{1}{4\pi} \int_{\Omega} P_{11}(\Theta) \cos\Theta \, \mathrm{d}\Omega \,. \tag{2.19}$$

If the particle scatters more light toward the forward ($\Theta < 90°$) than backward direction ($\Theta > 90°$), $g > 0$; otherwise, $g < 0$. For isotropic scattering or scattering symmetric about $\Theta = 90°$, $g = 0$. Finally, the single-scattering albedo (SSA) is defined as

$$\varpi = \frac{C_{\mathrm{sca}}}{C_{\mathrm{ext}}}, \tag{2.20}$$

which represents the percentage of a light beam that undergoes the scattering process.

Different theories have been developed to describe the scattering properties of particles with different sizes and shapes. The scattering by particles with the size parameter $x = 2\pi r/\lambda \ll 1$ can be approximated to Rayleigh scattering (van de Hulst, 1957); for the region where $x \sim 1$, the scattering properties of spherical particles can be accurately calculated by the Mie theory, while those of non-spherical particles usually resort to numeric methods, e.g., the T-matrix method (Mishchenko et al., 2000); for the region where $x \gg 1$, geometric optics approximation is an efficient and accurate way to calculate particle scattering properties (van de Hulst, 1957).

2.1.2 Scattering by particle ensembles

Instead of a single particle, atmospheric aerosols consist of particles with various sizes. The particle number size distribution (NSD) of an ensemble of particles is defined as

$$n(r) = \frac{\mathrm{d}N}{\mathrm{d}r}, \tag{2.21}$$

where r represents particle size. For spherical particles, r is the radius and for non-spherical particles, r could be the volume- or surface-area- equivalent radius. The particle number per unit volume between r and $r + \mathrm{d}r$ is denoted as $\mathrm{d}N$. The size distribution can also be described by the forms of surface-area size distribution (SSD) and volume size distribution (VSD) as

$$s(r) = 4\pi r^2 n(r), \quad v(r) = \frac{4}{3}\pi r^3 n(r). \tag{2.22}$$

Correspondingly, integrating over the entire size range gives the total number (N_t), surface (S_t) and volume (V_t) concentrations

$$N_\mathrm{t} = \int_{r_\mathrm{min}}^{r_\mathrm{max}} n(r)\mathrm{d}r\,, \quad S_\mathrm{t} = \int_{r_\mathrm{min}}^{r_\mathrm{max}} s(r)\mathrm{d}r\,, \quad V_\mathrm{t} = \int_{r_\mathrm{min}}^{r_\mathrm{max}} v(r)\mathrm{d}r, \tag{2.23}$$

where r_min and r_max represent the minimum and maximum particle radii of the ensemble. Because scattering by an ensemble of small particles depends on particle surface area or volume rather than number (Bohren and Huffman, 2004), and computation of scattering properties shows much smoother variability for logarithmic equidistant steps than for equidistant steps (Dubovik et al., 2006), throughout this study, we exploit VSD defined in the logarithmic scale:

$$v(r) = \frac{4}{3}\pi r^3 \frac{\mathrm{d}N}{\mathrm{dln}r} = \frac{\mathrm{d}V}{\mathrm{dln}r}. \tag{2.24}$$

It has been widely proved (Hansen and Travis, 1974; Kaufman et al., 2003) that another key parameter related to the radiative properties of a given size distribution is the effective radius which is defined as

$$r_\mathrm{eff} = \frac{\int_{\ln r_\mathrm{min}}^{\ln r_\mathrm{max}} v(r)\,\mathrm{dln}r}{\int_{\ln r_\mathrm{min}}^{\ln r_\mathrm{max}} \frac{1}{r} v(r)\,\mathrm{dln}r}. \tag{2.25}$$

Different functions have been proposed to parameterize size distribution of a particle ensemble (e.g., Junge, 1955; Davies, 1974). For example, the lognormal VSD is preferable in many modeling studies (Hess et al., 1998; Chin et al., 2002) and has been proved by both in-situ measurements (Di Biagio et al., 2019) and remote sensing retrievals (Dubovik et al., 2002), which is expressed as

$$\frac{\mathrm{d}V}{\mathrm{dln}r}=\frac{V_t}{\sqrt{2\pi}\ln\sigma_{\mathrm{g}}}\exp\left[-\frac{(\ln r-\ln r_{\mathrm{v}})^2}{2\ln^2\sigma_{\mathrm{g}}}\right], \tag{2.26}$$

where r_v is the mode radius (or volume median radius) and σ_{g} the geometric standard deviation. It can be proved that if a VSD conforms to the lognormal distribution, the corresponding SSD and NSD also can be expressed by the lognormal distribution. Furthermore, the following relationship holds for a lognormal VSD:

$$r_{\mathrm{eff}}=r_{\mathrm{v}}\exp\left(-\frac{1}{2}\ln^2\sigma_{\mathrm{g}}\right). \tag{2.27}$$

The scattering properties of a particle ensemble are called the bulk scattering properties. In practice, it is usually assumed that interferences from different particles are negligible (incoherent scattering) so that the intensity scattered by each scatterer is additive (Mishchenko et al., 2000). Accordingly, the bulk cross sections, also known as the extinction, scattering and absorption coefficients, are expressed as

$$\langle C_i\rangle=\int_{\ln r_{\min}}^{\ln r_{\max}}\frac{3}{4\pi r^3}C_i(r)v(r)\mathrm{dlnr}. \quad (i=\mathrm{ext},\mathrm{sca},\mathrm{abs}) \tag{2.28}$$

Note that the integral is expressed in the form of VSD. Likewise, the bulk SSA, $\langle\varpi\rangle$, bulk asymmetry factor, $\langle g\rangle$, and bulk phase matrix, $\langle\boldsymbol{P}\rangle$, can be calculated by

$$\langle\varpi\rangle=\frac{\langle C_{\mathrm{sca}}\rangle}{\langle C_{\mathrm{ext}}\rangle}, \tag{2.29}$$

$$\langle g\rangle=\frac{\int_{\ln r_{\min}}^{\ln r_{\max}}\frac{3}{4\pi r^3}C_{\mathrm{sca}}(r)g(r)v(r)\mathrm{dlnr}}{\langle C_{\mathrm{sca}}\rangle}, \tag{2.30}$$

$$\langle\boldsymbol{P}\rangle=\frac{\int_{\ln r_{\min}}^{\ln r_{\max}}\frac{3}{4\pi r^3}C_{\mathrm{sca}}(r)\boldsymbol{P}(r)v(r)\mathrm{dlnr}}{\langle C_{\mathrm{sca}}\rangle}, \tag{2.31}$$

respectively. The meaning of Eq. (2.31) is to perform the calculation for every element of the matrix. In addition, the bulk differential scattering cross section is expressed as

$$\left\langle\frac{\mathrm{d}C_{\mathrm{sca}}(r)}{\mathrm{d}\Omega}\right\rangle=\frac{1}{4\pi}\int_{\ln r_{\min}}^{\ln r_{\max}}\frac{3}{4\pi r^3}C_{\mathrm{sca}}(r)P_{11}(\Theta,r)v(r)\mathrm{dlnr}. \tag{2.32}$$

In particular, $\langle\frac{\mathrm{d}C_{\mathrm{sca}}(r)}{\mathrm{d}\Omega}\rangle$ is called the backscattering coefficient and denoted as β if $\Theta=\pi$.

One evident contrast between the single-particle scattering and bulk scattering is that the later shows smoother variability due to the size-averaging effect. Figure 2.4 shows contours of $-P_{12}/P_{11}$ (%) versus Θ and x for monodisperse spheres and polydisperse spheres with a narrow Junge distribution (Mishchenko et al., 2000). It can be seen that the strong oscillations

(Fig. 2.4 (a)) due to the interference of light diffracted and reflected/transmitted by a particle (Hansen and Travis, 1974) are smoothed out by the averaging effect (Figure 2.4 (b)).

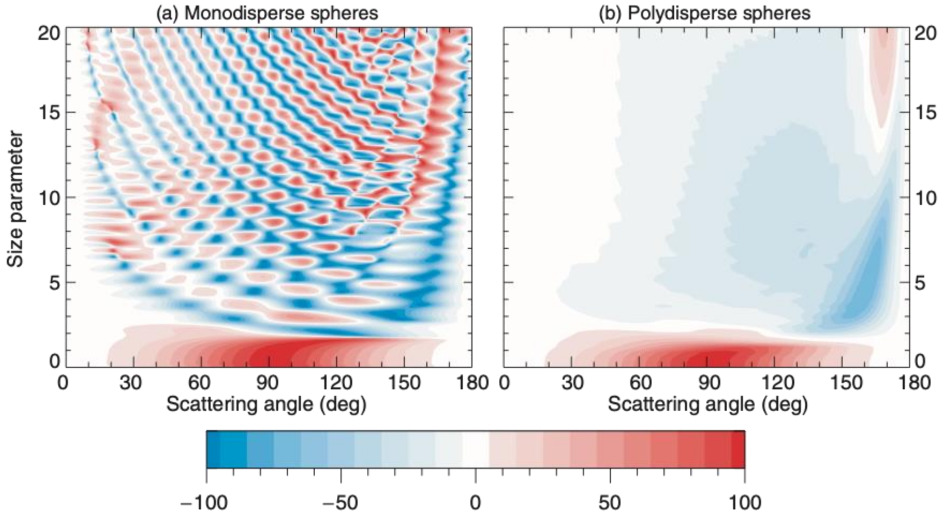


Figure 2.4. Color contour plots of $-P_{12}/P_{11}$ (%) of (a) monodisperse spheres and (b) polydisperse spheres. The complex refractive index (CRI) is 1.53 – 0.008i. (Adapted from Mishchenko et al. (2000))

2.1.3 Scattering models

In the context of this thesis, a scattering model refers to a database of particle scattering properties calculated based on particle size and shape. Besides, they are also functions of $\lambda, m_{\mathrm{R}}, m_{\mathrm{I}}$ and Θ. Three scattering models are considered in this study: i.e., the Sphere, Spheroid and Irregular-Hexahedral models. The latter two are developed to describe the scattering of non-spherical particles.

2.1.3.1 Spheroid model

The Spheroid model is a sub database of the Spheroid-package publicly available on https://www.grasp-open.com/products/ (last access: October 16, 2023), which is based on the study of Dubovik et al. (2006). A spheroid is formed by rotating an ellipse around its minor axis (oblate spheroid) or major axis (prolate spheroid), as shown in Figure 2.5. It is described by the axis ratio $\zeta = c/a$ (where c is the axis of spheroid rotational symmetry and a is the axis perpendicular to c) and volume-equivalent radius r – the radius of the sphere having the same volume as the spheroid. The Spheroid model approximates irregular particles to an ensemble of randomly orientated spheroids characterized by a VSD and an axis ratio distribution (ARD), which are independent of each other. The latter is defined as

$$n(\zeta) = \frac{\mathrm{d}N(\zeta)}{\mathrm{d}\ln\zeta}, \tag{2.33}$$

where $\mathrm{d}N(\zeta)$ represents the fraction of spheroids between $\ln\zeta$ and $\ln\zeta + \mathrm{d}\ln\zeta$.

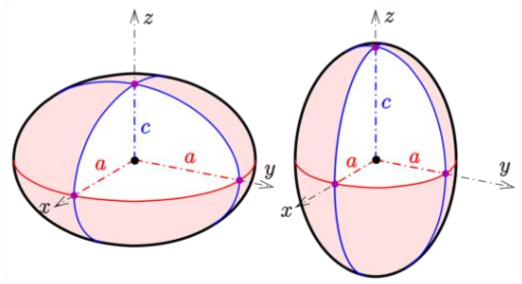


Figure 2.5. Oblate and prolate spheroids characterized by the axis ratio ζ and volume-equivalent radius r (cited from https://en.wikipedia.org/wiki/Spheroid, last access: October 16, 2023)

The advantages of the spheroid model lie in two aspects: (1) compared to spheres, approximating dust particles to spheroids better reproduces laboratory measurements of the phase matrix and, at the same time, remains reasonable computational burden due to relatively simple geometric shape of a spheroid; (2) the averaging effect of a randomly orientated non-spherical particle ensemble weakens the influence of detailed morphology of a single particle (Mishchenko et al., 2002). To perform bulk property calculation in a convenient way, instead of single-particle scattering properties, the database stores the extinction, scattering and angular scattering kernels – $K_{\mathrm{ext}}, K_{\mathrm{sca}}, K_{ii}(\Theta)$ corresponding to (ζ, r) grids so that the bulk optical properties can be derived by

$$\langle C_i \rangle = \sum_q \sum_p K_{i,pq} n(\zeta_p)\, v(r_q), \quad (i = \mathrm{ext}, \mathrm{sca}) \tag{2.34}$$

$$\langle C_{\mathrm{sca}} \rangle \langle P_{ij}(\Theta) \rangle = \sum_q \sum_p K_{ij,pq}(\Theta) n(\zeta_p)\, v(r_q), \quad (i, j = 1, 2, \dots, 4) \tag{2.35}$$

where $n(\zeta_p)$ and $v(r_q)$ are values of ARD and VSD at grid point ζ_p and r_q, respectively. The calculation of the kernels approximates $n(\zeta)$ as piecewise constant function and $v(r)$ as piecewise linear function, which will be described in more detail in Sect. 3.4.1.

The calculation covers the range of the size parameter, x, from 0.012 to 625 and different computational methods are used to maximize the efficiency and accuracy. Specifically, the advanced T-matrix method (Mishchenko et al., 2002) is used for smaller x and the approximate geometric-optics-integral-equation method (Yang and Liou, 1996) is used for larger x. The calculation is conducted at $\lambda = 340\,\mathrm{nm}$. Scattering properties at other wavelengths are deducted from $\lambda = 2\pi r/x$ according to the scale invariance rule (Mishchenko et al., 2002). Table 2.1 lists the ranges of the input variables to the Spheroid model.

The Spheroid model has been adopted as the forward model in the operational retrieval algorithms of AERONET and shows superior performance compared to the Sphere model when retrieving non-spherical particles like dust aerosols. However, its accuracy deteriorates at large scattering angles ($\Theta > 175°$) due to the limits of used computational methods (Dubovik et al., 2006). Comparisons between model simulations for the back scattering direction and lidar measurements also reveal apparent discrepancy (Haarig et al., 2022).

Table 2.1. Ranges of the input parameters for the Sphere model, Spheroid model, Irregular-Hexahedral model for the short-wave region (IH-SW) and Irregular-hexahedral model for the long-wave (IH-LW).

Input variable	**Sphere**		**Spheroid**		**IH-SW**		**IH-LW**	
	Min	**Max**	**Min**	**Max**	**Min**	**Max**	**Min**	**Max**
x **or** kD	0.012	625	0.012	625	≪1	11800	≪1	1470
m_R	1.3	1.7	1.3	1.7	1.37	1.7	0.4	3.2
m_I	~0	0.5	~0	0.5	0.0001	0.1	0.001	4.0
Θ	0	180°	0	180°	0	180°	0	180°
ζ **or** Ψ	1	1	0.3	3.0	0.695	0.785	0.695	0.785

2.1.3.2 Sphere model

The Sphere model is a sub database of the Spheroid-package publicly available on https://www.grasp-open.com/products/ (last access: October 16, 2023). In fact, it is the case of the Spheroid model with axis ratio $\zeta = 1$. Therefore, the scattering properties in the Sphere model are calculated with the fast Lorenz-Mie code for the same ranges and numbers of parameters as the Spheroid model (see Table 2.1).

2.1.3.3 Irregular-Hexahedral (IH) model

The irregular-hexahedral (IH) model was recently proposed by Saito et al. (2021) to mimic scattering properties of dust and volcanic ash aerosols. An irregular hexahedral particle refers to a hexahedron of which the faces are randomly tilted (Bi et al., 2010), as shown in Figure 2.6.

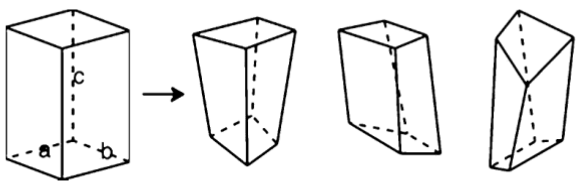


Figure 2.6. Illustration of generating irregular hexahedrons by randomly tilting the faces of a hexahedron (cited from Bi et al. (2010)).

The IH model is based on 20 irregular-hexahedral (IH) particles of which the aspect ratio (defined as the ratio of the largest particle dimension to the smallest particle dimension) varies between 1.14 to 4.02. The sphericity of each single IH particle is characterized by the degree of sphericity defined as

$$\psi_i = \frac{\pi^{1/3}(6V_i)^{2/3}}{4A_i}, \tag{2.36}$$

where the subscript i represents the particle ID; V_i and A_i are the volume and projected area of the i-th particle, respectively. The IH model approximates real non-spherical particles as a randomly orientated mixture of the 20 IH particles with a certain mixing ratio. Correspondingly, the ensemble-weighted degree of sphericity, serving as a measure of the sphericity of the particle ensemble, is calculated by

$$\Psi = \frac{\pi^{\frac{1}{3}}\left(\sum_{i=1}^{20} 6f_{\mathrm{mix},i}V_i\right)^{\frac{2}{3}}}{\sum_{i=1}^{20} 4f_{\mathrm{mix},i}A_i}, \tag{2.37}$$

where $f_{\mathrm{mix},i}$ is the mixing ratio for the i-th IH particle, which is independent of particle size. An illustration of the relation between ψ_i, $f_{\mathrm{mix},i}$ and Ψ is shown in Figure 2.7, from which one can see ψ_i increases with the increase of i, and the higher $f_{\mathrm{mix},i}$ and ψ_i, the higher Ψ. The model provides six combinations of $f_{\mathrm{mix},i}$ corresponding to ensembles with six Ψ ranging from 0.695 to 0.785. It then calculates the cross sections and phase matrix elements for each Ψ and for ranges of λ, m_r, m_i, and D (the diameter of the circumscribed sphere of the particle) using the following equations:

$$C_j(\lambda, m_\mathrm{R}, m_\mathrm{I}, \Psi, D) = \sum_{i=1}^{20} f_{\mathrm{mix},i} C_{j,i}(\lambda, m_\mathrm{R}, m_\mathrm{I}, \psi_i, D), \quad (j = \mathrm{ext}, \mathrm{sca}, \mathrm{abs}) \tag{2.38}$$

$$\mathbf{P}(\lambda, m_\mathrm{R}, m_\mathrm{I}, \Psi, D, \theta) = \frac{\sum_{i=1}^{20} f_{\mathrm{mix},i} C_{\mathrm{sca},i}(\lambda, m_\mathrm{R}, m_\mathrm{I}, \psi_i, D)\mathbf{P}_i(\lambda, m_\mathrm{R}, m_\mathrm{I}, \psi_i, D, \theta)}{C_\mathrm{sca}(\lambda, m_\mathrm{R}, m_\mathrm{I}, \Psi, D)}. \tag{2.39}$$

Note that unlike the Spheroid model, the IH model does not provide Kernel functions which means appropriate quadrature methods should be taken by the users when calculating the bulk scattering properties.

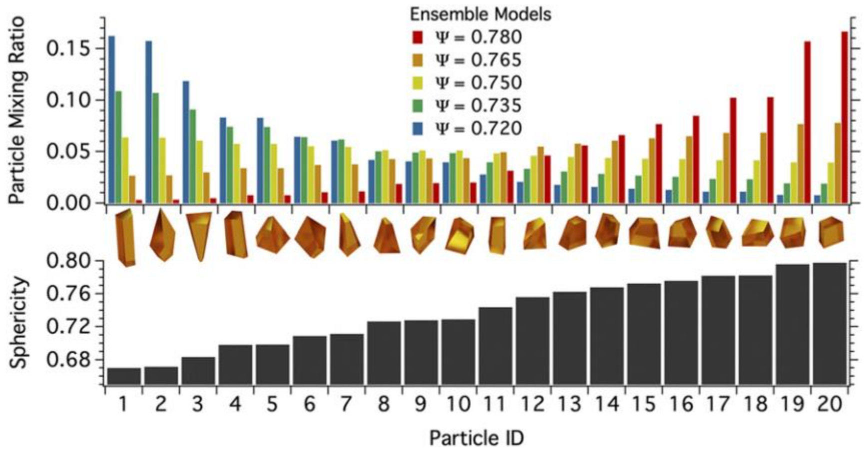


Figure 2.7. (top) ensemble-weighted degree of sphericity Ψ calculated from different combinations of mixing ratios; (bottom) single-particle degree of sphericity ψ_i versus particle ID i (cited from Saito et al. (2021)).

The scattering properties are calculated with respect to shortwave (SW) and longwave (LW) domains and the ranges of the input parameters are shown in Table 2.1. Note that the size parameter used in the IH model is defined as $kD = 2\pi D/\lambda$. According to the ranges of the size parameter, it employs different computational methods. Specifically, for $kD \ll 1$, the Rayleigh scattering approximation (Bohren and Huffman, 2004) with a V/A-equivalent sphere is used; for the small-to-moderate kD the invariant-imbedding T-matrix method (IITM) (Johnson, 1988) is used; and for large kD, the model combines the geometric-optics-integral-equation method and improved-geometric-optics method (Yang and Liou, 1996) – the former is mostly used for quasi-backscattering regions to achieve accurate convergence and the latter is mainly used for forward- to side-scattering regions to improve the computational efficiency.

2.2 Instrumentation – lidar system

Lidar (LIght Detection And Ranging) is an active remote sensing technique using laser as the radiative source. By measuring the time series of returned signals, it is capable of acquiring range-resolved information. With different instrumental configurations, lidar can be used to detect profiles of temperature, pressure, wind, humidity, clouds, trace gases and aerosols. In this study, we focus on aerosol detection lidars working for the range from the boundary layer (BL) to upper troposphere.

2.2.1 Instrumental setup

Typical lidar setup consists of a transmitter and a receiver, as shown in the left panel of Figure 2.8. A laser in the transmitter generates short light pulses which are sent to the atmosphere through a beam expander to reduce its divergence. The backscattered light entering the receiver is firstly collected by a telescope and then directed to a detector after passing through an optical analyzer. The optical analyzer could simply be interference filters to select at interested wavelengths or more sophisticated devices such as grating spectrometers, interferometers or polarizers. The detector converts the received optical signals to electrical signals for further analysis. Usually, the conversion can be done in two modes: photon-counting and analog. The former is preferable for weaker backscattered signal while the latter for the stronger.

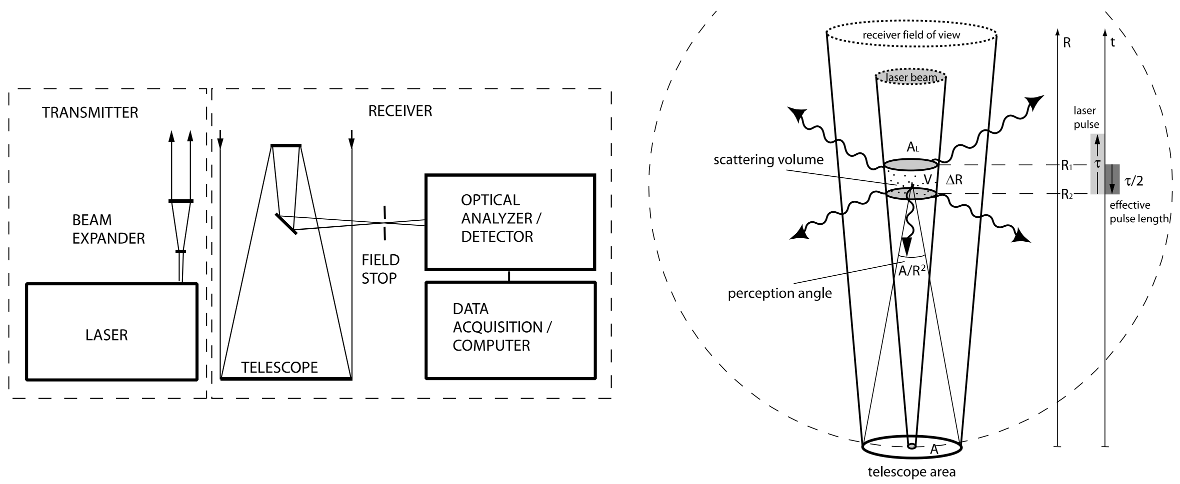


Figure 2.8. Schematic of (left) basic lidar setup (biaxial arrangement) and (right) lidar viewing geometry (coaxial arrangement). (cited from Wandinger (2005))

2.2.2 Lidar equation

To derive the relation between the power received by the telescope and that emitted to the atmosphere, consider the viewing geometry illustrated in the right panel of Figure 2.8. If the duration of an emitted pulse is τ and the detector records the signal at an instant time t after the front of the pulse was emitted, the received flux (or power, in W) is from the scattering by a bulk volume of particles with length $\Delta R = c\tau/2$ (c represents the speed of light) centered at $R = c(2t - \tau)/4$. According to Eq. (2.18), Eq, (2.32), the power collected by the telescope corresponding to distance R can be written as

$$P(R) = P_0 \eta O(R) \frac{c\tau}{2} \frac{A}{R^2} \beta(R) \exp\left[-2\int_0^R \alpha(z) \mathrm{d}z\right], \tag{2.40}$$

where P_0 is the power of emitted laser, η the system efficiency and A the telescope area. The overlap function, $O(z)$, varying between 0 and 1 depending on R, indicating the fraction of the backscattered power going into the receiver field of view. $\beta(R)$ is the total backscattering coefficient (Eq. (2.32)) and $\alpha(R)$ the total extinction coefficient (i.e., $\langle C_{\mathrm{ext}} \rangle$ in Eq. (2.28)). The factor 2 means the extinction happens along the path forth and back.

2.2.3 Derivation of optical profiles

State-of-the-art atmospheric detection lidars could receive and distinguish three types of returned signals generated from different light-particle interaction processes, i.e., elastic scattering, Raman scattering and fluorescence. In elastic scattering, incident light interacts with matter without energy transfer, making it scatter photons being of the same energy as the incident ones. However, a tiny fraction (approximately 1 in 1 million) of the scattered photons has lower or higher energy compared to the incident photons due to energy exchange with the matter, which is called Raman scattering (Harris and Bertolucci, 1989). Fluorescence is a process where molecules are excited by incident photons of a certain energy (usually in the UV-VIS spectrum), emit lower-energy photons and return back to their ground state in a very short time period ($10^{-9} - 10^{-6}$ s). Figure 2.9 illustrates the scattering and fluorescence processes for excitation laser at 355 nm. State-of-the-art lidars derive atmospheric optical profiles by combining signals from elastic, Raman and fluorescence channels.

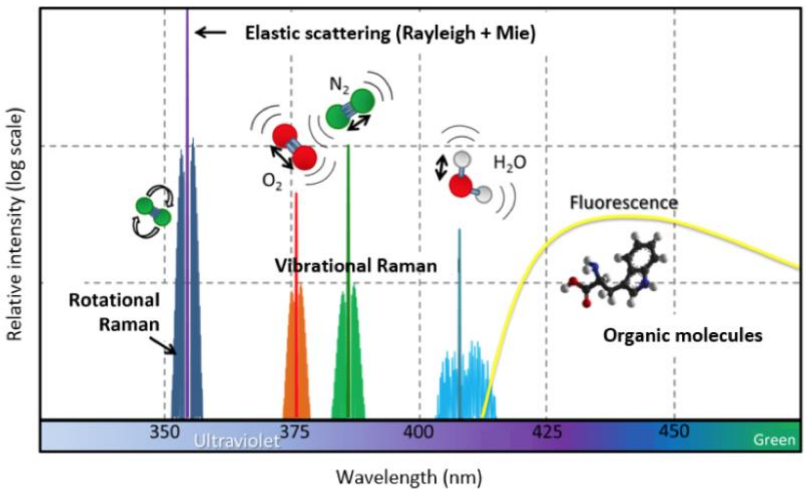


Figure 2.9. Scattering and fluorescence processes for 355 nm excitation [Felidj et al., 2016]

2.2.3.1 Elastic channels

For convenience, hereafter the overlap is assumed to be complete ($O(r) \equiv 1$) and the range-corrected lidar signal

$$P_{\mathrm{c}}(R) = R^2 P(R) = G[\beta_{\mathrm{a}}(R) + \beta_{\mathrm{m}}(R)] \exp\left\{-2\int_0^R [\alpha_{\mathrm{a}}(z) + \alpha_{\mathrm{m}}(z)]\mathrm{d}z\right\} \tag{2.41}$$

is used, where $G = \frac{1}{2}P_0\eta c\tau A$ is the system factor that only depends on instrumental performance. Note that the total backscattering and extinction coefficients are decomposed to aerosol contributions (i.e., $\beta_{\mathrm{a}}(R), \alpha_{\mathrm{a}}(R)$) and molecule contributions (i.e., $\beta_{\mathrm{m}}(R), \alpha_{\mathrm{m}}(R)$). Eq. (2.41) can be converted to a Bernoulli differential equation, of which the solution is (Klett, 1985)

$$\begin{aligned} \beta_{\mathrm{a}}(R) + \beta_{\mathrm{m}}(R) &= \frac{P_{\mathrm{c}}(R)\exp\left\{-2\int_{R_0}^{R}[L_{\mathrm{a}}(z) - L_{\mathrm{m}}]\beta_{\mathrm{m}}(z)\mathrm{d}z\right\}}{\dfrac{P_{\mathrm{c}}(R_0)}{\beta_{\mathrm{a}}(R_0) + \beta_{\mathrm{m}}(R_0)} - 2\int_{R_0}^{R} L_{\mathrm{a}}(z)P_{\mathrm{c}}(z)T(z,R_0)\mathrm{d}z}, \\ T(z,R_0) &= \exp\left\{-2\int_{R_0}^{z}[L_{\mathrm{a}}(z') - L_{\mathrm{m}}]\beta_{\mathrm{m}}(z')\mathrm{d}z'\right\}, \end{aligned} \tag{2.42}$$

where $L_i = \alpha_i/\beta_i$ $(i = \mathrm{a}, \mathrm{m})$ is referred to as the lidar ratio (LR). Since the components and distribution of molecules are much more stable in the atmosphere compared to aerosols, β_{m} can be accurately estimated from available meteorological data and L_{m} equals $8\pi/3$ (sr) throughout the whole atmosphere (Ansmann and Müller, 2005). However, L_{a} strongly depends on aerosol microphysical properties which often present great variability with height and is a priori unknown. The reference height R_0 should be properly selected so that $\beta_{\mathrm{a}}(R_0)$ is negligible and R_0 is greater than R (backward integration) to keep numerical stability (Klett, 1985). Deriving aerosol extinction and backscattering coefficient profiles using Eq. (2.42) with an assumed L_{a} profile and backward integration is called the Klett inversion. The intrinsic difficulty in the Klett inversion stems from the fact that there are two unknowns: $\beta_{\mathrm{a}}(R)$ and $\alpha_{\mathrm{a}}(R)$, in one equation.

2.2.3.2 Raman channels

In contrast to the Klett inversion, Raman technique is able to solve $\beta_{\mathrm{a}}(z)$ and $\alpha_{\mathrm{a}}(z)$ separately without critical assumption. In a Raman lidar, apart from the elastic backscattering at the emitted wavelength λ_0, the molecular (e.g., nitrogen) Raman scattering at a Raman-shifted wavelength λ_{Ra} is also measured. Because aerosol particles do not cause Raman effect, Eq. (2.41) for a Raman channel can be written as

$$P_{\mathrm{c}}(R,\lambda_{\mathrm{Ra}}) = G\beta_{\mathrm{m,Ra}}(R,\lambda_0)\exp\left\{-\int_0^R[\alpha(z,\lambda_0)+\alpha(z,\lambda_{\mathrm{Ra}})]\mathrm{d}z\right\}, \tag{2.43}$$

where the molecular Raman backscattering $\beta_{\mathrm{m,Ra}}(R,\lambda_0)$ can be calculated from its number concentration and differential Raman scattering cross section using Eq. (2.32). Accordingly, aerosol extinction coefficient at λ_0 can be calculated from Eq. (2.43) as

$$\alpha_{\mathrm{a}}(R,\lambda_0) = \frac{\frac{\mathrm{d}}{\mathrm{d}P_{\mathrm{c}}}\ln\frac{N_{\mathrm{t,m}}(R)}{P_{\mathrm{c}}(R,\lambda_{\mathrm{Ra}})} - \alpha_{\mathrm{m}}(R,\lambda_0) - \alpha_{\mathrm{m}}(R,\lambda_{\mathrm{Ra}})}{1+\left(\frac{\lambda_0}{\lambda_{\mathrm{Ra}}}\right)^{å}}, \tag{2.44}$$

where $N_{\mathrm{t,m}}$ is the number concentration of the molecular scatterer (e.g., nitrogen) and å is the aerosol extinction Angstrom exponent (EAE) between λ_0 and λ_{Ra}, which is usually taken as zero for rotational Raman signals. The derivation of $\beta_{\mathrm{a}}(R)$ makes use of both elastic and Raman signals and requires the signals at a reference height R_0, which gives

$$\beta_{\mathrm{a}}(R,\lambda_0) = [\beta_{\mathrm{a}}(R_0)+\beta_{\mathrm{m}}(R_0,\lambda_0)]\frac{R(R_0,\lambda_{\mathrm{Ra}})R(R,\lambda_0)N_{\mathrm{t,m}}(R)}{R(R_0,\lambda_0)R(R,\lambda_{\mathrm{Ra}})N_{\mathrm{t,m}}(R_0)}$$
$$\times\frac{\exp\left\{-\int_{R_0}^R[\alpha_{\mathrm{a}}(z,\lambda_{\mathrm{Ra}})+\alpha_{\mathrm{m}}(z,\lambda_{\mathrm{Ra}})]\mathrm{d}z\right\}}{\exp\left\{-\int_{R_0}^R[\alpha_{\mathrm{a}}(z,\lambda_0)+\alpha_{\mathrm{m}}(z,\lambda_0)]\mathrm{d}z\right\}} - \beta_{\mathrm{m}}(R,\lambda_0). \tag{2.45}$$

The reference height is recommended to choose in the upper troposphere where aerosol backscattering is negligible compared to molecules. As mentioned, $\beta_{\mathrm{m}}(R,\lambda_0)$ can be estimated from meteorological profiles or standard-atmosphere data.

2.2.3.3 Fluorescence channels

Different types of aerosols show distinct fluorescence spectra which provide extra information for aerosol characterization. The study by Pan (2015) showing that many aerosol types emit fluorescence signals in a range of 400-650 nm when excited at 355 nm indicates the potential of deriving fluorescence profiles with lidar. The work by Veselovskii et al. demonstrates the feasibility of separating the fluorescence signal within 444-487 nm from the 387 nm Raman signal using a dichroic mirror. Similar to the Raman channel, the range-corrected fluorescent signal can be expressed as

$$P_{\mathrm{c}}(R,\lambda_{\mathrm{F}}) = G\beta_{\mathrm{F}}(R)\exp\left\{-\int_0^R[\alpha(z,\lambda_0)+\alpha(z,\lambda_{\mathrm{F}})]\mathrm{d}z\right\}, \tag{2.46}$$

where the fluorescence backscattering coefficient β_{F} for the selected band is

$$\beta_{\mathrm{F}} = \int_{\lambda_{\min}}^{\lambda_{\max}}\int_{\ln r_{\min}}^{\ln r_{\max}}\frac{3}{4\pi r^3}\frac{\mathrm{d}C_{\mathrm{F}}(\lambda,r)}{\mathrm{d}\Omega}v(r)\mathrm{dlnr}\,\mathrm{d}\lambda \tag{2.47}$$

with the fluorescence differential cross section $\frac{\mathrm{d}C_\mathrm{F}(\lambda,r)}{\mathrm{d}\Omega}$. The extinction within the filter transmission band can be taken at the central wavelength λ_F of the band (Veselovskii et al., 2020) and β_F can be calculated from the ratio of Eq. (2.48) to Eq. (2.43):

$$\beta_\mathrm{F}(R) = \eta^* \frac{P_\mathrm{c}(R,\lambda_\mathrm{F})}{P_\mathrm{c}(R,\lambda_\mathrm{Ra})} \beta_\mathrm{m,Ra} \exp\left\{-\int_0^R [\alpha(z,\lambda_\mathrm{Ra}) + \alpha(z,\lambda_\mathrm{F})]\mathrm{d}z\right\}, \quad (2.48)$$

where η^* is a calibration coefficient determined by the overall detection efficiency of the fluorescence and Raman channels. Moreover, the particle fluorescence capacity is defined as (Veselovskii et al., 2020)

$$G_\mathrm{F} = \frac{\beta_\mathrm{F}}{\beta_\mathrm{a}}, \quad (2.49)$$

which is a useful parameter for aerosol characterization. For example, smoke and pollen aerosols have large G_F, while dust and urban aerosols have small G_F (Veselovskii et al., 2021, 2022).

2.2.3.4 Depolarization measurements

Depolarization is a process that when a linear polarized incident laser beam is scattered by a non-spherical particle, the polarization plane of the backscattered light rotates with respect to the incident polarization plane. The rotation can result from multiple internal reflections from geometric optical point of view (Liou and Lahore, 1974). The linear depolarization ratio (LDR) is defined as the ratio of the backscattering coefficient perpendicular to the incident polarization plane to that parallel to the incident polarization plane:

$$\begin{aligned} \delta_\mathrm{v} &= \frac{\beta_\perp}{\beta_\parallel} = \frac{\beta_{\mathrm{a},\perp} + \beta_{\mathrm{m},\perp}}{\beta_{\mathrm{a},\parallel} + \beta_{\mathrm{m},\parallel}}, \\ \delta_\mathrm{a} &= \frac{\beta_{\mathrm{a},\perp}}{\beta_{\mathrm{a},\parallel}} = \frac{(1+\delta_\mathrm{m})(\beta_\mathrm{a}+\beta_\mathrm{m})\delta_\mathrm{v} - (1+\delta_\mathrm{v})\delta_\mathrm{m}\beta_\mathrm{m}}{(1+\delta_\mathrm{m})(\beta_\mathrm{a}+\beta_\mathrm{m}) - (1+\delta_\mathrm{v})\beta_\mathrm{m}}, \end{aligned} \quad (2.50)$$

where $\delta_\mathrm{v}, \delta_\mathrm{a}, \delta_\mathrm{m}$ stand for volume, particle and molecule LDRs, respectively. The volume linear depolarization ratio can be accurately measured by means of polarizing beamsplitter cube (PBC) (Freudenthaler et al., 2009), while the particle linear depolarization ratio (PLDR) needs the retrieval of β_a in advance using the Klett or Raman methods. Isotropic spherical particles do not produce depolarization, which can be explained by geometric optics and proved by the following relation (Sassen, 2000)

$$\delta_\mathrm{a} = \frac{P_{11}(\pi) - P_{22}(\pi)}{P_{11}(\pi) + P_{22}(\pi)}, \quad (2.51)$$

where $P_{11}(\pi) = P_{22}(\pi)$ for an isotropic spherical particle and $P_{11}(\pi) \neq P_{22}(\pi)$ for a non-spherical particle. Thus, PLDR is a good indicator of particle sphericity.

2.2.4 LILAS system

LILAS (LIlle Lidar Atmospheric Study) is a multi-wavelength Mie-Raman polarization lidar developed by LOA. Figure 2.10 shows photos of LILAS system under testing and working. It exploits the Nd:YAG crystal and frequency doubling and tripling to emit laser pulses at 355, 532 and 1064 nm with a repetition rate of 20 Hz and energy of 90, 100, 100 mJ, respectively. The wavelengths of backscattering signals are separated by dichroic mirrors. In addition to the corresponding elastic receiving channels, it once had three Raman channels at 387, 530 and 408 nm. The first two were used for deriving $\beta_{\mathrm{a}}, \alpha_{\mathrm{a}}$ at 355 and 532 nm, while the last for water vapor mixing ratio retrieval. Recently, the water vapor channel was replaced with a fluorescence channel to measure the fluorescence signals within the band 444-487 nm (Veselovskii et al., 2020). In the elastic channels (355, 532 and 1064 nm), PBCs are mounted after half-wave plates to achieve depolarization measurements. The detectors used in 1064 nm cross- and parallel- polarized channels are the avalanche photodiodes while the photomultiplier tubes (PMTs) are used in other channels. In order to have a large dynamic range, a gluing method combining analog and photon-counting signals is used, except the 1064 channels where only the analog signals are recorded (Hu, 2018).

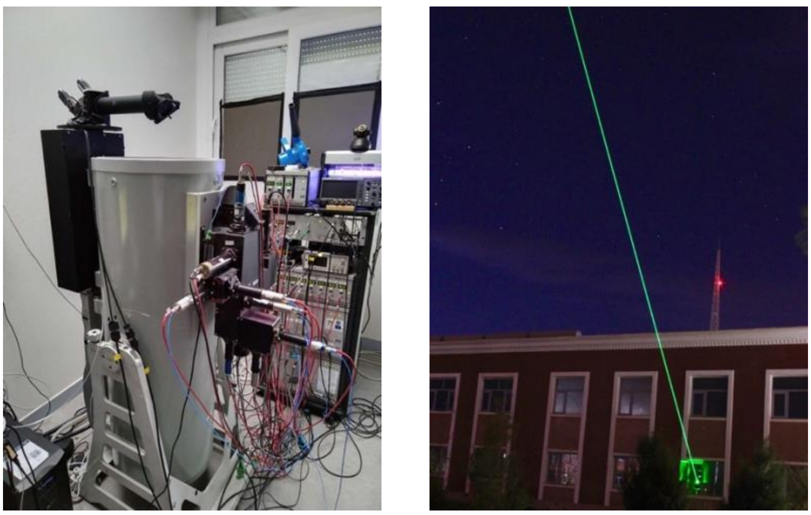


Figure 2.10. LILAS (LIlle Lidar Atmospheric Study) system under testing at the ATOLL observatory (left) and working during the Dust Aerosol Observation (DAO) campaign.

The performance of the lidar system is checked according to some EARLINET checking procedures (Freudenthaler et al., 2018), including Rayleigh fit, telecover test and $\Delta 90^0$ calibration. Before the calculation of aerosol optical properties, data pre-processing such as trigger delay correction, dead-time correction and electronic noise subtraction are also applied. After that, the $3\beta + 2\alpha + 3\delta + 1\beta_F$ profiles of aerosols, that is, backscattering coefficients at 355, 532 and 1064 nm, extinction coefficients at 355 and 532 nm, PLDRs at 355, 532 and 1064 nm, and fluorescence backscattering coefficient centered at 466 nm are acquired using the methods mentioned in Sect. 2.2.3. A description of error analysis can be found in Hu et al. (Hu et al., 2019).

LILAS is currently operated at the ATOLL (ATmospheric Observation in LiLle) platform, in the frame of SNO (Services Nationaux d'Observation), PHOTONS and ACTRIS-CARS (Centre for Aerosol Remote Sensing). The raw data acquired by LILAS (level 0 data) is transmitted to the ACTRIS-DC (Data Centre) and processed there using the Single Calculus Chain (SCC). The level 1 products include profiles of aerosol extinction coefficient, backscattering coefficient and depolarization ratio, together with the resulting profiles of lidar ratio, Angstrom exponent and backscatter-related Angstrom exponent, provided by SCC at near real time. The level 2 products are the same with level 1 products but after fully quality control (QC) and provided every 3 months (Wandinger et al., 2018). In parallel, raw LILAS measurements are also processed by the laboratory self-developed processing server – AUSTRAL (AUtomated Server for the TReatment of Atmospheric Lidars) for more flexible scientific research by LOA. Apart from the standard profiles like provided by SCC, advanced profiles, e.g., fluorescence, water vapor mixing ratio and relative humidity are also generated from AUSTRAL (Ducos et al., 2022). Further, comparisons of extinction, backscattering and PLDR profiles produced by ASUTRAL and SCC (Appendix A) show good consistency.

3 BOREAL algorithm

In this chapter, a complete description of the BOREAL algorithm is presented. It starts with the definition of the forward and inverse problem (Sect. 3.1), followed by a part describing the algorithm theoretical basis, including the forward modeling (Sect. 3.2) and numerical inversion (Sect. 3.3). Next, we provide detailed configurations on BOREAL to concretely implement the retrieval process, including configurations to derive an individual solution (Sect. 3.4) and criteria to identify the final solution space from a set of individual solutions (Sect. 3.5). A flowchart describing the algorithm is presented at the end of this chapter.

3.1 General description of the inverse problem

A physical system under study can be characterized by a set of state variables which could be scalars, vectors or continuous functions. If the state variables are not directly measurable, indirect measurements on observable variables have to be taken instead. In this case, there should be physical laws associating the measurements with the underlying state, which is referred to as forward models. For example, a mathematical expression of a process of forward modeling could be

$$Fx(\tau) = y(t), \tag{3.1}$$

where the state variable $x(\tau)$ is a continuous function defined on $\tau \in [a, b]$, the measurement $y(t)$ is a continuous function defined on $t \in [c, d]$, and the operator F represents the forward model. An inverse problem is, therefore, to solve the operator equation (3.1). In reality, $y(t)$ are obtained on discrete points (e.g., solar radiance measured at discrete wavelengths) so that $x(\tau)$ should be properly discretized in order to make Eq. (3.1) solvable. In addition, there are always errors between real measured quantities and those predicted by the forward model due to both measurement uncertainty and errors in forward modeling. Thus, in practice, Eq. (3.1) is expressed as

$$\mathbf{y} = \mathbf{f}(\mathbf{x}) + \boldsymbol{\varepsilon}, \tag{3.2}$$

where the state vector $\mathbf{x}$, measurement vector $\mathbf{y}$ and error vector $\boldsymbol{\varepsilon}$ have finite dimension; $\mathbf{f}$ consists of scaler functions f_i which are operated on $\mathbf{x}$ to derive the corresponding y_i.

Correspondingly, the inverse problem is to retrieve the state vector **x** from the measurement vector **y** under the presence of the error vector **ε**.

3.2 Forward modeling

3.2.1 Lidar measurements

The backscattering coefficient, β, and extinction coefficient, α, at the wavelength λ can be derived from Raman technique using Eq. (2.44) and Eq. (2.45). According to Eq. (2.28) and Eq. (2.32), they are associated with particle microphysical properties through

$$\beta_\lambda(m_\mathrm{R}, m_\mathrm{I}) = \int_{\ln r} k^M_{\beta,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) v(r) \mathrm{d}\ln\mathrm{r}, \tag{3.3a}$$

$$\alpha_\lambda(m_\mathrm{R}, m_\mathrm{I}) = \int_{\ln r} k^M_{\alpha,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) v(r) \mathrm{d}\ln\mathrm{r}. \tag{3.3b}$$

where the backscattering kernel function, $k^M_{\beta,\lambda}$, and extinction kernel function, $k^M_{\alpha,\lambda}$, have the forms:

$$k^M_{\beta,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) = \frac{3}{16\pi^2 r^3} C^M_{\mathrm{sca},\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) P^M_{11,\lambda,\pi}(m_\mathrm{R}, m_\mathrm{I}, r),$$

$$k^M_{\alpha,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) = \frac{3}{4\pi r^3} C^M_{\mathrm{ext},\lambda}(m_\mathrm{R}, m_\mathrm{I}, r)$$

with M = Sph, Sphd and IH for spherical, spheroid and irregular-hexahedral (IH) particles, respectively. Unlike the spherical kernel functions, the kernel functions for non-spherical particles are also functions of particle shape distribution. The operational inversion algorithm of AERONET assumes *a priori* known shape distribution rather than retrieves it (Dubovik et al., 2006). Given that lidar measurements contain less information than sun-sky photometer measurements (lack of angular measurements), we do not retrieve shape distribution, either. Specifically, for spheroids, the kernel functions are in advance integrated with a laboratory-measured ARD of a Feldspar sample, $n_0(\zeta)$, like the strategy taken by the AERONET algorithm:

$$\begin{aligned} k^{\mathrm{Sphd}}_{\beta,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r) &= \int_{\ln\zeta} \frac{3}{16\pi^2 r^3} C^{\mathrm{Sphd}}_{\mathrm{sca},\lambda}(m_\mathrm{R}, m_\mathrm{I}, r, \zeta) P^{\mathrm{Sphd}}_{11,\lambda,\pi}(m_\mathrm{R}, m_\mathrm{I}, r, \zeta) n_0(\zeta) \mathrm{d}\ln\zeta \\ &= \int_{\ln\zeta} k^{\mathrm{Sphd}}_{\beta,\lambda}(m_\mathrm{R}, m_\mathrm{I}, r, \zeta) n_0(\zeta) \mathrm{d}\ln\zeta, \end{aligned} \tag{3.4a}$$

$$k_{\alpha,\lambda}^{\mathrm{Sphd}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) = \int_{\ln\zeta} \frac{3}{4\pi r^3} C_{\mathrm{ext},\lambda}^{\mathrm{Sphd}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \zeta)\, n_0(\zeta) \mathrm{dln}\zeta$$
$$= \int_{\ln\zeta} k_{\alpha,\lambda}^{\mathrm{Sphd}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \zeta) n_0(\zeta) \mathrm{dln}\zeta\,. \quad (3.4\mathrm{b})$$

The scattering properties provided by the IH model are already integrated with the shape distribution, characterized by their ensemble-weighted degree of sphericity, Ψ (Chap. 2.1.3.3). The model calculated a range of Ψ from 0.695 to 0.785 (Table 2.1). During the sensitivity study, we found lidar-related optical properties are not sensitive to the change of Ψ. Thus, we fix Ψ to 0.71, a value adopted by Saito et al. (Saito et al., 2021) to reproduce most of the laboratory- and field- measurements of dust particles. Accordingly, for the IH kernel functions:

$$k_{\beta,\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) = \frac{3}{16\pi^2 r^3} C_{\mathrm{sca},\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \Psi = 0.71) P_{11,\lambda,\pi}^{\mathrm{Sphd}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \Psi = 0.71)$$
$$= k_{\beta,\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \Psi = 0.71), \quad (3.5\mathrm{a})$$

$$k_{\alpha,\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) = \frac{3}{4\pi r^3} C_{\mathrm{ext},\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \Psi = 0.71) = k_{\alpha,\lambda}^{\mathrm{IH}}(m_{\mathrm{R}}, m_{\mathrm{I}}, r, \Psi = 0.71). \quad (3.5\mathrm{b})$$

In addition, for non-spherical particles, the PLDR, δ (defined by Eq. (2.46), and for simplicity, hereafter without ambiguity, subscript “a” is omitted), is also associated with particle microphysical properties through Eq. (2.47):

$$\delta_\lambda(m_{\mathrm{R}}, m_{\mathrm{I}}) = \frac{\langle P_{11,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\rangle - \langle P_{22,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\rangle}{\langle P_{11,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\rangle + \langle P_{22,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\rangle} \quad (3.6)$$

with

$$\langle P_{ii,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\rangle = \frac{1}{\langle C_{\mathrm{sca},\lambda}^{M}\rangle} \int_{\ln r} \frac{3}{4\pi r^3} C_{\mathrm{sca},\lambda}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) P_{ii,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) v(r) \mathrm{dlnr}$$
$$= \frac{1}{\langle C_{\mathrm{sca},\lambda}^{M}\rangle} \int_{\ln r} k_{ii,\lambda,\pi}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) v(r) \mathrm{dlnr}\,.$$

The phase matrix kernel functions, $k_{ii,\lambda,\pi}^{M}$, have been integrated with the shape distribution, in a same way as Eq. (3.4) for spheroids and Eq. (3.5) for IH particles. It can be seen that $k_{11,\lambda,\pi}^{M} = 4\pi k_{\beta,\lambda}^{M}$.

After integrated with the shape distribution, optical properties at λ, derived from Eq. (3.3) to Eq. (3.6), are functions of $m_{\mathrm{R}}, m_{\mathrm{I}}$ and $v(r)$. To further reduce the number of retrieved variables so as to ameliorate the underdetermination of the inverse system, we overlook the spectral dependence of m_{R} and m_{I}. Laboratory measurements show that m_{R} and m_{I} do not present great variabilities in the UV-VIS region for most aerosol types (D’Almeida et al., 1991). However, it is not the case for dust aerosol of which the spectral dependence of m_{I} is considered in BOREAL (a detailed discussion will be given in Sect. 3.4.2). In addition, $v(r)$

is a smooth continuous function of r, while the measurements are made at finite discrete wavelengths. Accordingly, discretization of $v(r)$ and quadrature of the integrals are needed. Here we follow Twomey (1977) by approximating $v(r)$ to a piecewise linear function with knots $v_1 = v(r_1), v_2 = v(r_2), \dots, v_n = v(r_n)$. For example, the backscattering coefficient between two grid points r_j and r_{j+1} can be expressed as

$$\int_{\ln r_j}^{\ln r_{j+1}} k_{\beta,\lambda}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) v(r) \mathrm{d}\ln r = \int_{\ln r_j}^{\ln r_{j+1}} \frac{\ln r_{j+1} - \ln r}{\ln r_{j+1} - \ln r_j} k_{\beta,\lambda}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) \mathrm{d}\ln r \cdot v_j + \int_{\ln r_j}^{\ln r_{j+1}} \frac{\ln r - \ln r_j}{\ln r_{j+1} - \ln r_j} k_{\beta,\lambda}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}, r) \mathrm{d}\ln r \cdot v_{j+1}. \quad (3.7)$$

Correspondingly, over the whole integral interval $[r_{\min}, r_{\max}]$ (also called the inversion window, see Sect. 3.5), the backscattering coefficient can be expressed as

$$\beta_\lambda(m_{\mathrm{R}}, m_{\mathrm{I}}) = \sum_{j=1}^{n} K_{\beta,\lambda,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\, v_j \quad (3.8a)$$

with

$$K_{\beta,\lambda,j}^{M}(\dots) = \int_{\ln r_{j-1}}^{\ln r_j} \frac{\ln r - \ln r_{j-1}}{\ln r_j - \ln r_{j-1}} k_{\beta,\lambda}^{M}(\dots, r) \mathrm{d}\ln r + \int_{\ln r_j}^{\ln r_{j+1}} \frac{\ln r_j - \ln r}{\ln r_{j+1} - \ln r_j} k_{\beta,\lambda}^{M}(\dots, r) \mathrm{d}\ln r,$$

where the omitted parameters "..." are $m_{\mathrm{R}}, m_{\mathrm{I}}$. The grid points do not include $r_{\min}$ and $r_{\max}$ and $K(r_{\min})$ and $K(r_{\max})$ are excluded from the calculation by prescribing $v(r_{\min}) = v(r_{\max}) = 0$. Likewise, the integrals for calculating α and δ are converted to summation as

$$\alpha_\lambda(m_{\mathrm{R}}, m_{\mathrm{I}}) = \sum_{j=1}^{n} K_{\alpha,\lambda,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}})\, v_j, \quad (3.8b)$$

$$\delta_\lambda(m_{\mathrm{R}}, m_{\mathrm{I}}) = \frac{\sum_j K_{11,\lambda,\pi,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}) v_j - \sum_j K_{22,\lambda,\pi,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}) v_j}{\sum_j K_{11,\lambda,\pi,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}) v_j + \sum_j K_{22,\lambda,\pi,j}^{M}(m_{\mathrm{R}}, m_{\mathrm{I}}) v_j} \quad (3.8c)$$

with the phase matrix element kernels $K_{ii,\lambda,\pi}^{M}$ and extinction kernels $K_{\alpha,\lambda}^{M}$ calculated in the same way as $K_{\beta,\lambda}^{M}$.

The above process of quadrature is, in fact, equivalent to use a set of B-spline functions of the 1st order as base functions to represent $v(r)$ as

$$v(r) \approx \mathbf{v}^T \mathbf{B} = \sum_{j=1}^{n} v_j B_j(r),$$
$$B_j(r) = \begin{cases} 0, & \ln r \le \ln r_{j-1} \\ \dfrac{\ln r - \ln r_{j-1}}{\ln r_j - \ln r_{j-1}}, & \ln r_{j-1} < \ln r \le \ln r_j \\ \dfrac{\ln r_{j+1} - \ln r}{\ln r_{j+1} - \ln r_j}, & \ln r_j < \ln r \le \ln r_{j+1} \\ 0, & \ln r > \ln r_{j+1}. \end{cases} \quad (3.9)$$

The number and order of the used B-spline functions as well as the positions of the knots are not independent of each other and have great impact on retrieval results in the TSVD method (Böckmann, 2001), but with regard to the method used here, 8 B-spline functions of the 1st order with logarithmic equidistant knots are a well-recognized configuration to stabilize the retrieval (Müller et al., 1999; Veselovskii et al., 2002; Müller et al., 2019). Figure 3.1 shows an example of 8 log-equidistant-distributed base functions defined between $r_{\min}$ and $r_{\max}$ and how they discretize a lognormal $v(r)$.

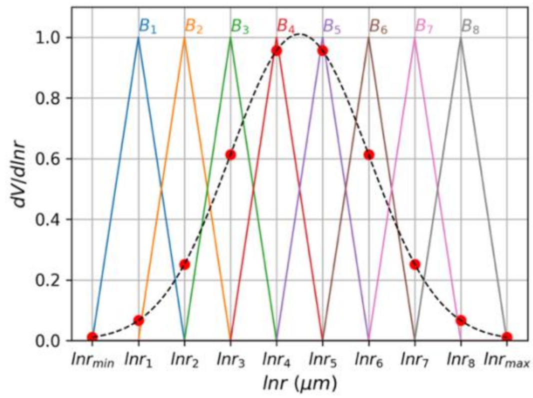


Figure 3.1. Illustration of B-spline functions of the 1st order ($B_1, B_2, \dots, B_8$) defined between an inversion window $[r_{min}, r_{max}]$. A lognormal volume size distribution $v(r)$ as an example is shown in the dashed line, of which the values corresponding to the peaks of the B-spline functions (red points) are selected as a discretization of $v(r)$.

From the discussion above, the 1st forward model related to lidar measurements can be expressed as

$$\mathbf{y}_1 = \mathbf{f}_1(\mathbf{x}) + \boldsymbol{\varepsilon}_1, \tag{3.10}$$

where the state vector is constructed as $\mathbf{x} = [v_1, v_2, \dots, v_8, m_\mathrm{R}, m_\mathrm{I}]^T$**;** the 1st measurement vector consists of lidar-measured optical properties and is constructed as

- for spherical particles: $\mathbf{y}_1 = [\beta_{355}, \beta_{532}, \beta_{1064}, \alpha_{355}, \alpha_{532}]^T$;
- for non-spherical particles: $\mathbf{y}_1 = [\beta_{355}, \beta_{532}, \beta_{1064}, \alpha_{355}, \alpha_{532}, \delta_{355}, \delta_{532}, \delta_{1064}]^T$;

furthermore, $\boldsymbol{\varepsilon}_1$ is the error vector of $\mathbf{y}_1$.

3.2.2 Constraints

So far, the model system is underdetermined (or under-constrained) since the dimension of $\mathbf{x}$ (denoted $\dim(\mathbf{x})$) is larger than that of $\mathbf{y}_1$ (denoted $\dim(\mathbf{y}_1)$). Even if $\dim(\mathbf{x}) \ll \dim(\mathbf{y}_1)$, Eq. (3.7) is the Fredholm integral equation of the 1st kind, a well-known ill-posed problem (Twomey, 1977; Yagola, 2010) which only has stable solution with the help of extra constraints.

Here we consider two types of constraints: the smoothing constraint on $v(r)$ and *a priori* constraint on CRI.

The smoothing constraint stems from the contradiction below: on the one hand, realistic particle size distributions should be smooth and positive; on the other hand, due to the mathematical instability of the ill-posed problem, small perturbation in the measurement vector caused by measurement error will lead to serious oscillation and even negative values in the state vector. To suppress the oscillation, we prescribe the 2nd order derivative of the retrieved $v(r)$ is close to zero. Accordingly, the 2nd forward model related to the smoothing constraint is expressed as

$$\begin{gathered}\mathbf{y}_2 = \mathbf{f}_2(\mathbf{x}) + \boldsymbol{\varepsilon}_2,\\ \mathbf{y}_2 \equiv \mathbf{0},\ \ \mathbf{f}_2(\mathbf{x}) \equiv \mathbf{Hv},\ \ \boldsymbol{\varepsilon}_2 \equiv \boldsymbol{\varepsilon}_\mathrm{S},\end{gathered} \tag{3.11}$$

where $\mathbf{H}$ is the 2nd difference matrix with the form

$$\mathbf{H} = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}$$

to calculate the 2nd order difference of $\mathbf{v}$, which is allowed to vary in a range $[-\boldsymbol{\varepsilon}_\mathrm{S}, \boldsymbol{\varepsilon}_\mathrm{S}]$.

The *a priori* constraint influences the retrieval of CRI by introducing *a priori* statistical results from climatologic data. Such constraint is applied in order to restrict the range of retrieved m_R and m_I, prescribing that the retrieved values should not deviate from the *a priori* values beyond the statistical limit. Thus, the 3rd forward model related to the *a priori* constraint is

$$\begin{gathered}\mathbf{y}_3 = \mathbf{f}_3(\mathbf{x}) + \boldsymbol{\varepsilon}_3,\\ \mathbf{y}_3 \equiv \left[m_{\mathrm{R,a}}, m_{\mathrm{I,a}}\right]^T, \mathbf{f}_3(\mathbf{x}) \equiv [m_\mathrm{R}, m_\mathrm{I}]^T, \boldsymbol{\varepsilon}_3 \equiv \left[\varepsilon_{m_{\mathrm{R,a}}}, \varepsilon_{m_{\mathrm{I,a}}}\right]^T.\end{gathered} \tag{3.12}$$

The mean of the statistics is taken as the *a priori* value $m_{\mathrm{R,a}}$ $\left(m_{\mathrm{I,a}}\right)$; the limit of the statistics as the *a priori* error $\varepsilon_{m_{\mathrm{R,a}}}$ $\left(\varepsilon_{m_{\mathrm{I,a}}}\right)$.

3.3 Numerical inversion

The inversion strategy is based on statistical optimization (Tarantola, 2005; Rodgers, 2000; Dubovik and King, 2000). If error in the forward model is negligible, the measured value is supposed to be a random variable with the forward-modeled value as the mean and measurement uncertainty as the standard deviation. Since measurements on optical properties are non-negative in reality, it is reasonable to assume the measurements conform to multivariate lognormal distribution (Eadie et al., 1971; Tarantola, 2005):

$$\mathbf{Y}_l = \ln \mathbf{y}_l \sim N\big(\mathbf{F}_l(\mathbf{x}), \mathbf{C}_{\mathbf{Y}_l}\big), \quad (l = 1,2,3) \tag{3.13}$$

where $\mathbf{F}_l(\mathbf{x}) = \ln \mathbf{f}_l(\mathbf{x})$ represents the mean vector and $\mathbf{C}_{\mathbf{Y}_l}$ the covariance matrix. By assuming each set of measurements is independent, one is able to construct the likelihood function (Fisher and Russell, 1922) as

$$\prod_l p_l(\mathbf{Y}_l, \mathbf{x}) = \prod_l \frac{1}{\sqrt{(2\pi)^{m_l}\left|\mathbf{C}_{\mathbf{Y}_l}\right|}} \exp\left\{-\frac{1}{2}[\mathbf{Y}_l - \mathbf{F}_l(\mathbf{x})]^T \mathbf{C}_{\mathbf{Y}_l}^{-1}[\mathbf{Y}_l - \mathbf{F}_l(\mathbf{x})]\right\}, \tag{3.14}$$

where p_l is the probability density function (PDF) of the set l, $m_l = \dim(\mathbf{Y}_l)$ and $\left|\mathbf{C}_{\mathbf{Y}_l}\right|$ is the determinant of $\mathbf{C}_{\mathbf{Y}_l}$. According to maximum likelihood estimation (MLE), the $\hat{\mathbf{x}}$ maximizing the likelihood function Eq. (3.14) is taken as the best estimate of $\mathbf{x}$. The maximization of Eq. (3.14) is equivalent to minimize the multi-term quadratic form:

$$h(\mathbf{x}) = \sum_{l=1}^{3} [\mathbf{Y}_l - \mathbf{F}_l(\mathbf{x})]^T \mathbf{C}_{\mathbf{Y}_l}^{-1}[\mathbf{Y}_l - \mathbf{F}_l(\mathbf{x})]\,. \tag{3.15}$$

It should be stressed that by definition in Eq. (3.11), $\mathbf{Y}_2$ is meaningless. However, we keep this notation for having a compact form. In fact, the smoothing constraint is converted to

$$\mathbf{0} = \mathbf{H} \ln \mathbf{v} + \boldsymbol{\varepsilon}_2. \tag{3.16}$$

In practice, instead of the absolute value of $\mathbf{x}$, we retrieve its logarithm $\mathbf{X} = \ln \mathbf{x}$ to ensure the positivity of the retrieved parameters. As a result, the cost function to minimize is

$$h(\mathbf{X}) = \sum_{l=1}^{3} \Phi_l = \sum_{l=1}^{3} [\mathbf{Y}_l - \mathbf{F}_l(\exp \mathbf{X})]^T \mathbf{C}_{\mathbf{Y}_l}^{-1}[\mathbf{Y}_l - \mathbf{F}_l(\exp \mathbf{X})]\,. \tag{3.17}$$

By the statistical estimation theory, the estimate of $\mathbf{X}$ is normally distributed. That is, $\hat{\mathbf{x}}$ also conforms to lognormal distribution. The study by Dubovik and King (2000) proves the connection between the MLE after logarithmic transformation and Chahine-like iteration.

Since Eq. (3.17) is non-linear, the minimization procedure has to be performed numerically. The Newton method is efficient to search a local minimum not too far from the initial point (Chong and Żak, 2013). The basic idea is, at an iteration point $\mathbf{X}^{(k)}$, to approximate the cost function by a quadratic function with the same first and second order derivatives as the cost function. The minimal point of the quadratic function is selected as a new point for the next iteration. The basic iteration formula for the Newton method is

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} - \nabla^2 h\big(\mathbf{X}^{(k)}\big)^{-1} \nabla h\big(\mathbf{X}^{(k)}\big), \tag{3.18}$$

where ∇h and $\nabla^2 h$ are the first and second order derivatives of h, respectively. Given the form of Eq. (3.17), we exploit the Gauss-Newton iteration as a good approximation to Eq. (3.18)

$$\mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \Delta\mathbf{X}^{(k)},$$
$$\left[\sum_{l=1}^{3} \mathbf{J}_{\mathbf{F}_l}^T\left(\mathbf{X}^{(k)}\right)\mathbf{C}_{\mathbf{Y}_l}^{-1}\mathbf{J}_{\mathbf{F}_l}\left(\mathbf{X}^{(k)}\right)\right]\Delta\mathbf{X}^{(k)} = \left\{\sum_{l=1}^{3} \mathbf{J}_{\mathbf{F}_l}^T\left(\mathbf{X}^{(k)}\right)\mathbf{C}_{\mathbf{Y}_l}^{-1}\left[\mathbf{Y}_l - \mathbf{F}_l\left(\mathbf{X}^{(k)}\right)\right]\right\}. \tag{3.19}$$

where $\mathbf{J}_{\mathbf{F}_l}\left(\mathbf{X}^{(k)}\right)$ is the Jacobi matrix of $\mathbf{F}_l$ at $\mathbf{X}^{(k)}$, with the form

$$\mathbf{J}_{\mathbf{F}_1} = \begin{bmatrix} \frac{x_1}{\{\mathbf{f}_1\}_1}\frac{\partial\{\mathbf{f}_1\}_1}{\partial x_1} & \frac{x_2}{\{\mathbf{f}_1\}_1}\frac{\partial\{\mathbf{f}_1\}_1}{\partial x_2} & \cdots & \frac{x_\mathrm{n}}{\{\mathbf{f}_1\}_1}\frac{\partial\{\mathbf{f}_1\}_1}{\partial x_\mathrm{n}} \\ \frac{x_1}{\{\mathbf{f}_1\}_2}\frac{\partial\{\mathbf{f}_1\}_2}{\partial x_1} & \frac{x_2}{\{\mathbf{f}_1\}_2}\frac{\partial\{\mathbf{f}_1\}_2}{\partial x_2} & \cdots & \frac{x_\mathrm{n}}{\{\mathbf{f}_1\}_2}\frac{\partial\{\mathbf{f}_1\}_2}{\partial x_\mathrm{n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{x_1}{\{\mathbf{f}_1\}_\mathrm{m}}\frac{\partial\{\mathbf{f}_1\}_\mathrm{m}}{\partial x_1} & \frac{x_2}{\{\mathbf{f}_1\}_\mathrm{m}}\frac{\partial\{\mathbf{f}_1\}_\mathrm{m}}{\partial x_2} & \cdots & \frac{x_\mathrm{n}}{\{\mathbf{f}_1\}_\mathrm{m}}\frac{\partial\{\mathbf{f}_1\}_\mathrm{m}}{\partial x_\mathrm{n}} \end{bmatrix}_{\mathbf{x}=\exp(\mathbf{X}^{(k)})},$$

$$\mathbf{J}_{\mathbf{F}_2} = \left[\mathbf{H}, \frac{\partial \ln \mathbf{f}_2}{\partial m_\mathrm{R}}, \frac{\partial \ln \mathbf{f}_2}{\partial m_\mathrm{I}}\right] = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 & 0 & 0 \end{bmatrix},$$

$$\mathbf{J}_{\mathbf{F}_3} = \left[\frac{\partial \ln \mathbf{f}_3}{\partial \mathbf{v}}, \frac{\partial \ln \mathbf{f}_3}{\partial m_\mathrm{R}}, \frac{\partial \ln \mathbf{f}_3}{\partial m_\mathrm{I}}\right] = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

When the coefficient matrix before $\Delta\mathbf{X}^{(k)}$ in Eq. (3.19) is not positively definite, it is not invertible and the Gauss-Newton method does not show descendent behavior. To deal with this problem, we introduce the Levenberg-Marquardt (LM) modification to Eq. (3.19):

$$\left[\sum_{l=1}^{3} \mathbf{J}_{\mathbf{F}_l}^T\left(\mathbf{X}^{(k)}\right)\mathbf{C}_{\mathbf{Y}_l}^{-1}\mathbf{J}_{\mathbf{F}_l}\left(\mathbf{X}^{(k)}\right) + \gamma^{(k)}\mathbf{D}\right]\Delta\mathbf{X}^{(k)} = \left\{\sum_{l=1}^{3} \mathbf{J}_{\mathbf{F}_l}^T\left(\mathbf{X}^{(k)}\right)\mathbf{C}_{\mathbf{Y}_l}^{-1}\left[\mathbf{Y}_l - \mathbf{F}_l\left(\mathbf{X}^{(k)}\right)\right]\right\}, \tag{3.20}$$

where the term $\gamma^{(k)}\mathbf{D}$ consists of the scalar damping factor γ and scaling matrix $\mathbf{D}$ to make the coefficient matrix positively definite. When $\gamma \to 0$, Eq. (3.20) is equal to the Gauss-Newton method; when $\gamma \to \infty$, Eq. (3.20) is asymptotic to the gradient method with infinitesimal iteration step. Thus, $\gamma^{(k)}$ should be adjusted with iteration. At beginning of the iteration, a large $\gamma^{(k)}$ ensures reduction of the cost function in case the iteration point is far from the optimal point. As the iteration point approaches to convergence, a small $\gamma^{(k)}$ accelerates the convergence rate. The scaling matrix, $\mathbf{D}$, is diagonal to scale the influence of γ on different state variables since their magnitudes can be quite different (e.g., m_{R} and m_{I}).

3.4 Implementation of the inversion procedure

3.4.1 Inclusion of the scattering models as forward models

Optical properties modeled by $\mathbf{f}_1(\mathbf{x})$ are directly taken from the particle scattering models described in Sect. 2.1.3. Using these pre-calculated optical properties facilitates computational efficiency. For calculating spherical and spheroidal particles, we use kernels $K_{11,\lambda}^{\mathrm{Sphd}}, K_{22,\lambda}^{\mathrm{Sphd}}, K_{\alpha,\lambda}^{\mathrm{Sphd}}$ provided by the Spheroid-Package database. They were derived by integrating the corresponding kernel functions with the base functions over double intervals $\Delta \ln r$, $\Delta \ln \zeta$ using quadrature methods similar to Eq. (3.7) and (3.8). Note that the kernels are also calculated for discrete grids of $m_{\mathrm{R}}, m_{\mathrm{I}}$ and $\lambda_0 = 340$ nm without implementing the quadrature (refer to Table 2.1 for ranges of the grids). Kernels out of the grid points are derived by linear interpolation. Kernels at other wavelengths are derived according to the scale invariance rule (Mishchenko et al., 2002) as

$$K_{\dots,\lambda}^{\mathrm{Sphd}}(\dots, r, \zeta) = 10^3 \frac{\lambda_0}{\lambda} K_{\dots,\lambda_0}^{\mathrm{Sphd}}\left(\dots, \frac{\lambda_0}{\lambda} r, \zeta\right). \tag{3.21}$$

With kernels, Eq. (3.4) is converted from integral to summation and the kernels for spherical particles and for the fixed Feldspar ARD are calculated by

$$K_{\dots,\lambda,r_j}^{\mathrm{Sph}}(\dots) = \sum\nolimits_p K_{\dots,\lambda,r_j}^{\mathrm{Sphd}}(\dots, \zeta_p) n_{\mathrm{Sph}}(\zeta_p), \quad n_{\mathrm{Sph}}(\zeta) = \begin{cases} 1, & \zeta = 1 \\ 0, & \zeta \neq 1 \end{cases}, \tag{3.22}$$

$$K_{\dots,\lambda,r_j}^{\mathrm{Sphd}}(\dots) = \sum\nolimits_p k_{\dots,\lambda,r_j}^{\mathrm{Sphd}}(\dots, \zeta_p) n_0(\zeta_p). \tag{3.23}$$

As a reminder, $n_0(\zeta)$ is the ARD of the Feldspar sample. Unlike the Sphere and Spheroid models, the IH model provides scattering properties (kernel functions) calculated for discrete grids of $(\lambda, m_{\mathrm{R}}, m_{\mathrm{I}}, D, \Psi)$ without implementing any quadrature calculation (refer to Eq. (2.38) and (2.39)). Thus, we apply the quadrature method described by Eq. (3.7) and (3.8) to derive kernels from the corresponding kernel functions for IH particles. Since the IH model calculates the kernel functions with respect to D (the diameter of the circumscribed sphere of the particle), they are firstly converted to the form with respect to the volume-equivalent radius r:

$$k^{\mathrm{IH}}_{\dots,\lambda}(\dots,D) = k^{\mathrm{IH}}_{\dots,\lambda}(\dots,u^{-1}(V)) = k^{\mathrm{IH}}_{\dots,\lambda}\left(\dots,u^{-1}(w(r))\right), \tag{3.24}$$

where $V = u(D)$ is the ensemble-weighted volume as a function of D. A complete mapping from D to V is stored in the IH dataset. By definition, r can be derived from the relation

$$V = w(r) = \frac{4}{3}\pi r^3. \tag{3.25}$$

We extracted the discrete kernel functions for $D \in [0.002, 100]$ µm, which corresponds to $r \in [0.0005, 25]$ µm and at $\lambda = 355, 532, 1064$ nm. Note that the IH model derives the values of kernel functions out of the grid points by linear interpolation. Thus, the IH kernels between the interval $[\ln r_{j-1}, \ln r_{j+1}]$ are calculated by

$$K^{\mathrm{IH}}_{\dots,\lambda,r_j}(\dots) = \begin{bmatrix}\frac{1}{6}, & \frac{1}{3}, & \frac{1}{3}, & \frac{1}{6}\end{bmatrix}\begin{bmatrix}(\ln r_j - \ln r_{j-1})k^{\mathrm{IH}}_{\dots,\lambda}(\dots,r_{j-1})\\ (\ln r_j - \ln r_{j-1})k^{\mathrm{IH}}_{\dots,\lambda}(\dots,r_j)\\ (\ln r_{j+1} - \ln r_j)k^{\mathrm{IH}}_{\dots,\lambda}(\dots,r_j)\\ (\ln r_{j+1} - \ln r_j)k^{\mathrm{IH}}_{\dots,\lambda}(\dots,r_{j+1})\end{bmatrix}. \tag{3.26}$$

There are two sets of radius grids: the model grids refer to the grids at which the scattering models calculate the kernel functions and kernels, while the retrieval grids are those at which the underlying $v(r)$ is discretized and $\mathbf{v}$ is retrieved. In AERONET retrieval strategy, the model grids coincide with the retrieval grids, while in this study, the model grids are finer than the retrieval grids: i.e., 8 retrieval grids versus not less than 14 model grids in an inversion window. Since the quadrature method is based on the piecewise linearization of $v(r)$, $v(r_q)$ at a model grid r_q between the retrieval grids r_j and r_{j+1} can be calculated by

$$v(r_q) = \frac{v_{j+1} - v_j}{\ln r_{j+1} - \ln r_j}\ln r_q + \left(v_j - \frac{v_{j+1} - v_j}{\ln r_{j+1} - \ln r_j}\ln r_j\right). \tag{3.27}$$

In this way, $v(r)$ corresponding to all model grids within the inversion window are able to be acquired, so that the bulk optical properties can be simulated from Eq. (3.8).

3.4.2 Determination of key parameters of the retrieval process

3.4.2.1 Covariance matrices of the measurement vectors

The covariance matrix in each quadratic term appearing in Eq. (3.14) – (3.17) acts as a weight to allocate the contribution of the quadratic term to the entire cost function, although it is not a strict terminology for $\mathbf{C}_{\mathbf{Y}_2}$ because the smoothing constraint is not based on statistical results. To determine $\mathbf{C}_{\mathbf{Y}_1}$, the lidar measurements in $\mathbf{Y}_1$ are assumed to be independent so that $\mathbf{C}_{\mathbf{Y}_1}$ becomes diagonal and the lognormal assumption results in

$$\{\mathbf{C}_{\mathrm{Y}_1}\}_{ij} = \ln\left[\frac{1}{2}\left(1 + \sqrt{1 + \frac{4\sigma_i^2}{\{\mathbf{y_1}\}_i^2}}\right)\right], \quad (i = j) \tag{3.28}$$

where σ_i is the standard deviation of the i-th lidar measurement, $\{\mathbf{y_1}\}_i$. If the relative standard deviation, $\sigma_i/\{\mathbf{y_1}\}_i$ is far less than 1, the $\{\mathbf{C}_{\mathrm{Y}_1}\}_{ij}$ can be approximated as $(\sigma_i/\{\mathbf{y_1}\}_i)^2$ (Dubovik and King, 2000). In practice, we set $\sigma_i/\{\mathbf{y_1}\}_i$ to the one-third of the maximum relative measurement error. An analysis of the maximum relative measurement error in each channel for LILAS can be found in Hu et al. (2019).

We set $\mathbf{C}_{\mathrm{Y}_2}$ to a scalar c_{Y_2} so that it serves as a scaling factor for the 2nd quadratic term. The inverse of c_{Y_2} plays a role similar to the Lagrange multiplier in the regularization method (Müller et al., 1999; Veselovskii et al., 2002). That is, the larger ${c_{\mathrm{Y}_2}}^{-1}$, the smoother the retrieved VSD but the worse the retrieval fits the lidar measurements. To the contrary, the smaller ${c_{\mathrm{Y}_2}}^{-1}$, the stronger oscillation of the retrieved VSD but the better the retrieval fits the lidar measurements. During testing the algorithm, we found compared to $\mathbf{C}_{\mathrm{Y}_1}$, retrieval results are not very sensitive to c_{Y_2} and $[1, 10]$ is a proper range for c_{Y_2}.

The m_{R} and m_{I} can be treated as independent of each other in the UV-VIS-NIR region, especially for the case where the spectral dependence is neglected. Accordingly, $\mathbf{C}_{\mathrm{Y}_3}$ is a 2×2 diagonal matrix with *a priori* variances of m_{R} and m_{I} as the diagonal elements:

$$\{\mathbf{C}_{\mathrm{Y}_3}\}_{11} = \sigma_{m_{\mathrm{R,a}}}^2, \quad \{\mathbf{C}_{\mathrm{Y}_3}\}_{22} = \sigma_{m_{\mathrm{I,a}}}^2.$$

The *a priori* information on particle CRI is important to reduce the underdetermination and better constrain the solution space because plenty of studies have shown the backward lidar measurements might have very limited sensitivity for accurately retrieving CRI, especially the imaginary part (Veselovskii et al., 2002, 2004, 2010; Müller et al., 2013, 2016, 2019). Because CRI is closely related to chemical composition of the aerosol but independent of particle concentration and size distribution, reasonable estimates of $(m_{\mathrm{R,a}}, \sigma_{m_{\mathrm{R,a}}})$ and $(m_{\mathrm{I,a}}, \sigma_{m_{\mathrm{I,a}}})$ can be obtained as long as the type of the aerosol to be retrieved is known in advance. The aerosol type can be determined by either species identification methods using lidar measurements (Burton et al., 2012; Nicolae et al., 2018; Veselovskii et al., 2022) or tracking the source of the aerosol event using back-trajectory analysis (Stein et al., 2015).

At the current stage, BOREAL provides three sets of *a priori* constraints on CRI, corresponding to: absorbing, non-absorbing and dust aerosols, respectively, with values listed in Table 3.1. The "absorbing" and "non-absorbing" types assume spectrally independent m_{I}

for spherical particles (Chang et al., 2022), while the "dust" type is specially used for dust retrieval, considering the spectral variation of dust m_I. The *a priori* constraint on m_R is the same for all aerosol types because we found during the sensitivity test, compared to the imaginary part, lidar measurements have higher sensitivity to the real part, and it is not necessary to designate different values for different aerosol types.

Table 3.1. *A priori* constraints on CRI for different aerosol types.

	$m_{\mathrm{I,a}}$	$\sigma_{m_{\mathrm{I,a}}}$	$m_{\mathrm{R,a}}$	$\sigma_{m_{\mathrm{R,a}}}$
Absorbing	0.015	0.01	1.5	0.1
Non-absorbing	0.005	0.005	1.5	0.1
Dust	0.005 (355 nm)	0.005 (355 nm)	1.5	0.1

The absorbing aerosols are characterized by their relatively higher $m_{\mathrm{I,a}}$ possibly because they contain black carbon generated through combustion processes. Most of biomass burning aerosols (BBAs) and anthropogenic soot from fossil fuel combustion are of this kind. The non-absorbing type mainly includes sea-salt particles from maritime regions, water-soluble aerosols related to gas-particle conversion and sulfate aerosol generated from both industrial activities and natural processes (volcanic eruption), characterized by relatively lower $m_{\mathrm{I,a}}$. Dubovik et al. (2002) found that geographical locations also have a great impact on particle absorption. For example, the urban-type aerosol in Mexico City is found to be more absorbing than that in Goddard Space Flight Center (GSFC) probably due to higher amount of anthropogenic combustion of fossil fuel; BBAs generated from different vegetation types (forest vs. savanna) and combustion processes (flaming vs. smoldering) with various ambient temperature, relative humidity as well as aging process can present quite distinct absorbing properties. Thus, knowledge on aerosol source is of great importance to properly set the *a priori* constraint on CRI.

Dust aerosol is considered separately due to the spectral variation of m_I. Remote sensing (Dubovik et al., 2002), in-situ (Granados-Muñoz et al., 2016) and laboratory measurements (Di Biagio et al., 2019) show dust presents the highest absorption in the UV and much less absorption towards the NIR region. A study by Veselovskii et al. (2010) indicates that ignoring the spectral variation of dust m_I will lead to retrieval error in particle volume concentration of 17-25%, as well as increases of the uncertainties of other retrieval parameters. In this study, we account for spectrally dependent dust m_I in a simple way: only retrieve the m_I at 355 nm; the values at 532 and 1064 nm are derived from regression analysis to laboratory measurements made by Di Biagio et al. (2019). Figure 3.2 shows the relation of the imaginary parts of CRI of

19 global distributed dust samples measured at 355 ($m_{\mathrm{I},355}$), 532 ($m_{\mathrm{I},532}$) and 1064 ($m_{\mathrm{I},1064}$) nm, as well as the results of linear regression. One can see a strong correlation (0.98) with a larger slope (0.52) between $m_{\mathrm{I},355}$ and $m_{\mathrm{I},532}$ and a relatively weak correlation (0.68) with a smaller slope (0.14) between $m_{\mathrm{I},355}$ and $m_{\mathrm{I},532}$. In addition, given the fact that lidar measurements are not sensitive to $m_{\mathrm{I},1064}$ especially when $m_{\mathrm{I},1064}$~0.001 (a detailed discussion is given in Sect. 4.2), we fix it to 0.001 throughout retrieval.

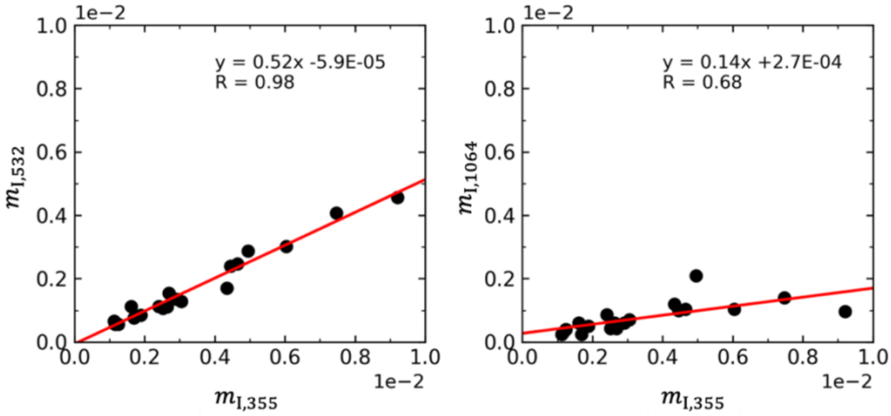


Figure 3.2. Relation of laboratory-measured (Di Biagio et al., 2019) imaginary parts of CRI (black points) at 355 ($m_{\mathrm{I},355}$), 532 ($m_{\mathrm{I},532}$) and 1064 ($m_{\mathrm{I},1064}$) nm and linear fitting results (red line). The measured values are from 19 global distributed dust samples, interpolated or extrapolated to the lidar wavelengths.

3.4.2.2 Damping factor and scaling matrix

As mentioned above, $\gamma^{(k)}$ is set to ensure the reduction of the cost function in every iteration step. However, too large $\gamma^{(k)}$ can significantly reduce the convergence rate. Basically, there are several strategies to update the damping factor. One is to find $\gamma^{(k)}$ causing the greatest reduction of the cost function in each iteration step. However, the computational burden is massive (Rodgers, 2000). Another strategy is to use a large $\gamma^{(k)}$ right from the start of iteration to ensure the reduction in the first step. After an iteration step, decrease $\gamma^{(k)}$ if the cost function reduces; otherwise, increase $\gamma^{(k)}$, do not update $\mathbf{X}^{(k)}$ and try again (Press et al., 2002). In this study, we adopt a method similar to the latter strategy where the reduction of $\gamma^{(k)}$ is related to the magnitude of the cost function:

If $h\left(\mathbf{X}^{(k)} + \Delta\mathbf{X}^{(k)}\right) > h\left(\mathbf{X}^{(k)}\right)$ then

$$\gamma^{(k)} \coloneqq 2\gamma^{(k)}, \tag{3.29a}$$

and try again, else:

$$\gamma^{(k+1)} = \max\left\{\frac{2h(\mathbf{X}^{(k)})}{E_h}, \frac{\gamma^{(k)}}{3}\right\}, \quad \mathbf{X}^{(k+1)} = \mathbf{X}^{(k)} + \Delta\mathbf{X}^{(k)}, \tag{3.29b}$$

with $E_h = \sum_{l=1}^{3}\dim(\mathbf{Y}_l) - \dim(\mathbf{X})$. Eq. (3.29) indicates that at beginning of iteration the cost function usually has large magnitude since the initial point $\mathbf{X}^{(0)}$ is far from the optimal point and as a result, $\gamma^{(0)}$ is also large. As the iteration proceeds, $h(\mathbf{X}^{(k)})$ approaches to its local minimum and correspondingly, $\gamma^{(k)}$ decreases to accelerate the convergence. Other more sophisticated updating strategies can be found in Fletcher (1971), Moré (1978), etc.

The scaling matrix **D** was firstly proposed by Marquardt (1963) to account for difference in magnitude of the elements of the state vector. There is no united form of **D** but a diagonal matrix is preferable. Here we follow the suggestion by Dubovik and King (2000): namely, its diagonal elements are inversely proportional to the square of the maximal ranges of the corresponding state variables:

$$\mathbf{D} = \begin{bmatrix} 1/\Delta_{\mathbf{v}}^2 & \mathbf{0} & 0 \\ \mathbf{0} & 1/\Delta_{m_\mathrm{R}}^2 & 0 \\ 0 & 0 & 1/\Delta_{m_\mathrm{I}}^2 \end{bmatrix}, \quad \begin{cases} \Delta_{\mathbf{v}} = \frac{1}{2}(\ln v_{\max} - \ln v_{\min})\mathbf{E} = 2.54\mathbf{E} \\ \Delta_{m_\mathrm{r}} = \frac{1}{2}\left(\ln m_{\mathrm{R,max}} - \ln m_{\mathrm{R,min}}\right) = 0.07, \\ \Delta_{m_\mathrm{i}} = \frac{1}{2}\left(\ln m_{\mathrm{I,max}} - \ln m_{\mathrm{I,min}}\right) = 2.3 \end{cases} \tag{3.30}$$

where **E** represents the unit matrix.

3.4.3 Convergence criteria and stopping conditions

There have been lots of widely-used convergence criteria in optimization problems. For example, the iteration could stop if one of the following conditions is satisfied (Chong and Żak, 2013):

(1) $\frac{|h(\mathbf{X}^{(k+1)}) - h(\mathbf{X}^{(k)})|}{\max\{1, |h(\mathbf{X}^{(k)})|\}} < \varepsilon_T$;

(2) $\frac{\|\mathbf{X}^{(k+1)} - \mathbf{X}^{(k)}\|}{\max\{1, \|\mathbf{X}^{(k)}\|\}} < \varepsilon_T$;

(3) $\|\nabla h(\mathbf{X}^{(k)})\| < \varepsilon_T$.

The threshold ε_T is usually set to 10^{-4} for condition (3) or 10^{-5} for conditions (1) and (2). However, applying these rules in our cases often makes the iteration end up with an unrealistic solution: the retrieved VSD has extremely large values for $r > 2\mu\mathrm{m}$. At the same time, the fitting residual in each term of the cost function is quite small, meaning a complete convergence is reached. It is because the extinction and backscattering Kernels decrease rapidly and tend to be identical for $r > 2\mu\mathrm{m}$ (Veselovskii et al., 2004). Checks on iteration sequences found that in the beginning, the state vector is able to evolve towards the right direction; however, after certain steps, elements of $\mathbf{v}$ corresponding to large r start to increase while the other elements almost do not change. Meanwhile, the quadratic term in the cost function representing the smoothing constraint keep decreasing while the lidar measurement term nearly stagnates, until the above convergence criteria are met. Another problem is that when $\mathbf{X}^{(k)}$ is far from the optimal point, $h(\mathbf{X}^{(k)})$ could be high, which in turns makes $\gamma^{(k)}$ high and results in a small iteration step. On the other hand, ∇h in this region could also be small (h is flat) so that the above criteria are easily satisfied after a few iterations. Therefore, to ensure correct convergence, we need: (1) a metric that serves as a measure of the fitting quality and is able to identify the iteration step where the state vector starts to deteriorate; (2) in addition to the criteria on the relative variation between two steps, criteria related to the absolute value of the cost function. Note that by looking at h as a function of $\mathbf{Y}_l$, it conforms to the chi-square distribution with an expectation of E_h (see Eq. (3.29)). Thus, in this study, the iteration will stop if one of the two following conditions is satisfied:

(1) $h(\mathbf{X}^{(k)}) < E_h$ and for $\forall a \in (\mathbf{Y}_1 - \ln \mathbf{f}_1(\mathbf{X}) - \boldsymbol{\varepsilon}_1), |a| < 0$;

(2) the maximum iteration number of 30 is reached.

The first condition makes sure the overall fitting residual of the quadratic form is comparable to the overall measurement uncertainty and in particular, the fitting residual for every lidar measurement does not exceed its maximum measurement error.

3.4.4 Initial guess

An initial guess of the state vector $\mathbf{X}^{(0)}$ is needed to initialize the optimization procedure. An initial guess as near the solution as possible helps in realizing correct convergence and accelerating convergence rate. To the contrary, arbitrarily selected $\mathbf{X}^{(0)}$ could trap the state vector in a region where $h(\mathbf{X})$ is large while $\nabla h(\mathbf{X})$ is small (a plateau of the cost function) or lead it to another local minimum. The latter is more common for the ill-posed lidar-aerosol retrieval problem where the non-uniqueness of the solution has been extensively discussed by Chemyakin et al. (2016). Considering that $m_{\mathrm{R,a}}, m_{\mathrm{I,a}}$ should not be far from the actual values, we set $m_{\mathrm{R}}^{(0)} = m_{\mathrm{R,a}}$ and $m_{\mathrm{I}}^{(0)} = m_{\mathrm{I,a}}$. Next, the elements of $\mathbf{v}^{(0)}$ is set to an identical value $v^{(0)}$ derived by fitting the measured α at 532 nm:

$$v^{(0)} = \frac{\alpha_{532}}{\sum_{j=1}^{n} K_{\alpha,532,j}^{M}\left(m_{\mathrm{R}}^{(0)}, m_{\mathrm{I}}^{(0)}\right)}. \tag{3.31}$$

3.4.5 Calculation of other microphysical and optical properties

After $v(r), m_{\mathrm{R}}$ and m_{I} are retrieved by the statistical optimal inversion method described above, particle volume concentration V_{t}, effective radius r_{eff} and bulk single scattering albedo $\langle\varpi\rangle$ are calculated with Eq. (2.23), Eq. (2.25) and Eq. (2.29), respectively.

3.5 Selection of individual solutions

The above mentioned forward modeling and inversion processes need a specific interval of radius, $[r_{\mathrm{min}}, r_{\mathrm{max}}]$, to construct retrieval grids and account for model grids. We call such interval an inversion window and the state vector retrieved within the inversion window is referred to as the "individual solution" for the inversion window. The equivalent radius of most aerosol particles varies from 0.05 to 10 μm (Lenoble et al., 2013). In the AERONET retrieval algorithm, a fixed inversion window, from $r_{\mathrm{min}} = 0.05$ μm to $r_{\mathrm{max}} = 15$ μm, is set, within which $v(r)$ is retrieved at 22 log-equidistant retrieval grids that coincide with the model grids. However, compared with AERONET measurements, the lidar measurements employed for aerosol retrieval contain much less information content so that $v(r)$ is only retrieved at 8 retrieval grids. If we just fix the inversion window to [0.05 μm, 15 μm], the same as the AERONET algorithm, the error resulting from the quadrature will be too large. Figure 3.3 shows the resampling errors as functions of the mode radius, r_{v}, of a lognormal VSD with $\ln\sigma_{\mathrm{g}} = 0.5$ (see Eq. (2.26) for its definition). The resampling error is defined as the relative difference between the optical properties calculated from $v(r)$ represented by fine radius grids (denoted as $\mathbf{v}_{\mathrm{fine}}$) and those calculated from $v(r)$ represented by reduced grids (denoted as $\mathbf{v}_{\mathrm{redu}}$):

$$\varepsilon_{\mathrm{rs}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{\{\mathbf{f}_1(\mathbf{v}_{\mathrm{fine}}, m_{\mathrm{R}}, m_{\mathrm{I}}) - \mathbf{f}_1(\mathbf{v}_{\mathrm{redu}}, m_{\mathrm{R}}, m_{\mathrm{I}})\}_i}{\{\mathbf{f}_1(\mathbf{v}_{\mathrm{fine}}, m_{\mathrm{R}}, m_{\mathrm{I}})\}_i}\right)^2}. \tag{3.32}$$

It can be seen that the larger $\varepsilon_{\mathrm{rs}}$, the larger errors in the forward model caused by the quadrature process. Figure 3.3 indicates that for an inversion window [0.05 μm, 15 μm], $\varepsilon_{\mathrm{rs}}$ increases with the increase of r_{v}, reaching its maximum at r_{v}~3 μm. According to the used scattering model and combination of the optical properties, the maximum of $\varepsilon_{\mathrm{rs}}$ ranges from 6% to 17%.

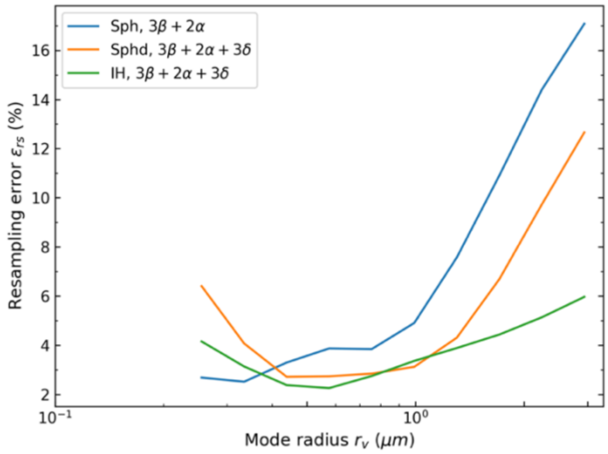


Figure 3.3. Resampling error versus mode radius of the lognormal VSD with a standard deviation of 0.5 calculated for the inversion window [0.05 μm, 15 μm]. The numbers of fine and reduced radius grids are 22 and 8, respectively. Different scattering models and combinations of optical properties are accounted for and the corresponding results are shown for different colors.

Undoubtedly, there is an optimal inversion window with length and position that best represent the underlying $v(r)$. Nevertheless, it is hard to be found because the location and spread of the $v(r)$ to retrieve is unknown. To this end, a set of inversion windows with changing positions and lengths are defined. A solution set consisting of individual solutions for every pre-defined inversion window is then obtained, from which "qualified" individual solutions are selected to constitute the so-called solution space. Although such strategy has been adopted by many other lidar-aerosol retrieval algorithms based on linear inversion (Veselovskii et al., 2002, 2012; Müller et al., 2016, 2019), one evident contrast here, however, is that the size of the solution set is greatly reduced because VSD and CRI are simultaneously retrieved in one inversion window. In addition, the logarithmic transformation ensures the positivity of all retrieved VSDs, which simplifies the selection process. The following selection criteria are based on the consideration that the individual solutions for properly set inversion windows should well fit the measurements and meanwhile, be physically "reasonable" (i.e., VSD is of a reasonable shape and CRI lies in a reasonable range).

The selection process consists of two steps. In the first step, the *VSD shape* criteria select individual solutions of which the VSDs have lognormal-like shape, which form the *good-shaped* solution set. The complementary of the *good-shaped* set constitutes the *bad-shaped* solution set, as shown in Figure 3.4. The assumption of lognormally distributed VSD (either mono-mode or multi-mode) should be reasonable since such size distribution of aerosols has been widely proved by both remote sensing and in-situ measurements (Davies, 1974; Ott, 1990; Dubovik et al., 2002; Reid et al., 2008). However, instead of constraining the mode radius or standard deviation (Eq. (2.25)) of the retrieved $v(r)$, the *shape* criteria only constrain the relationship between $v(r_j)$ where r_j is near $r_{\max}$ or $r_{\min}$ and the maximum of $v(r)$, denoted as $v_{\max}$. Specifically, the left and right sides of the retrieved VSD should satisfy

- for the left side:

$$\begin{cases} v_1 < v_2 \\ v_1 < 0.5v_{\max} \end{cases} \text{or} \quad \begin{cases} v_1 \geq v_2 \\ v_1 < 0.05v_{\max} \end{cases}; \tag{3.33}$$

- for the right side:

$$\begin{cases} v_8 < v_7 \\ v_8 < 0.5v_{\max} \end{cases} \text{or} \quad \begin{cases} v_8 \geq v_7 \\ v_8 < 0.05v_{\max} \end{cases}; \tag{3.34}$$

Furthermore, the number of modes of the retrieved VSD should not exceed 2.

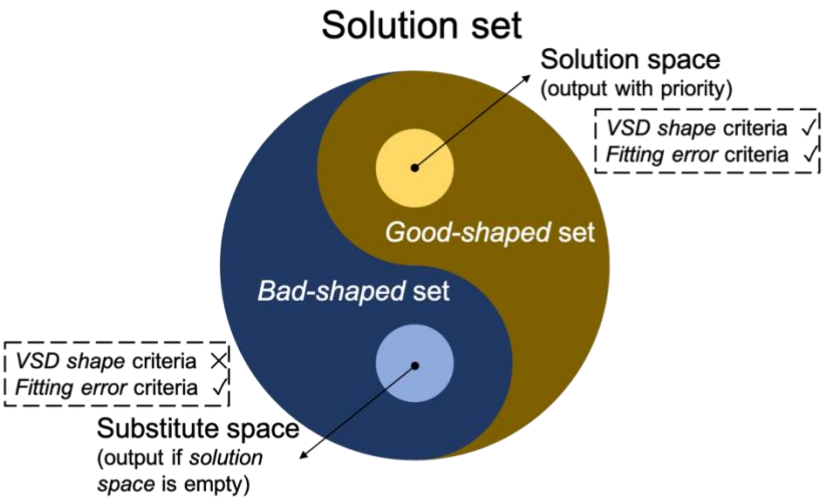


Figure 3.4. Relationship of the solution set, *good-shaped* solution set, *bad-shaped* solution set, solution space and substitute space.

After the two sub sets of the solution set – the *good-shaped* and *bad-shaped* solution sets are identified, in the second step, the *fitting error* criteria select the individual solutions with small fitting error from the two sub sets, respectively. The fitting error measures the distance between the real lidar measurements and the measurements reproduced by the retrieved parameters, which is defined as

$$\varepsilon_{\text{fit}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\{\mathbf{y}_1 - \mathbf{f}_1(\hat{\mathbf{x}})\}_i}{\{\mathbf{y}_1\}_i}\right)^2}, \tag{3.35}$$

where $\hat{\mathbf{x}}$ is the retrieved state vector. Accordingly, the *fitting error* criteria process the *good-shaped* solution set in the following steps:

(1) sort the individual solutions as the ascending order of ε_{fit};

(2) select the first 20% of the sorted individual solutions and, in the remaining part, the individual solutions with $\varepsilon_{\mathrm{fit}} < \bar{\sigma}_{\mathrm{m}}$. $\bar{\sigma}_{\mathrm{m}}$ is the root-mean square of relative measurement standard deviation:

$$\bar{\sigma}_{\mathrm{m}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\sigma_i}{\{\mathbf{y}_1\}_i}\right)^2}. \tag{3.36}$$

The selected solutions are output as the solution space and the final solution is the mean of the solution space. When the *good-shaped* solution set is empty, the *fitting error* criteria process the *bad-shaped* solution set in the same way as described above and as a result, the substitute space is output.

3.6 Chapter summary

This chapter describes the principle and implementation of the BOREAL algorithm. BOREAL is constructed by configuration, forward, inversion and post-process modules. The configuration module receives input from users, initializes configuring arguments and passes them to the other modules. The forward module includes models describing physical processes of particle bulk extinction, backscattering and depolarization, and constraints on particle VSD and CRI. As the core of the algorithm, the inversion module receives input measurements, performs numeric inversion based on the maximum likelihood estimation method and retrieves particle VSD, m_{R} and m_{I} for a selected inversion window. It calls the forward module in each iteration procedure. The post-process module calculates V_{t}, r_{eff} and bulk SSA from the output of the inversion module and determines the solution space. Figure 3.5 shows a flowchart illustrating the algorithmic logic.

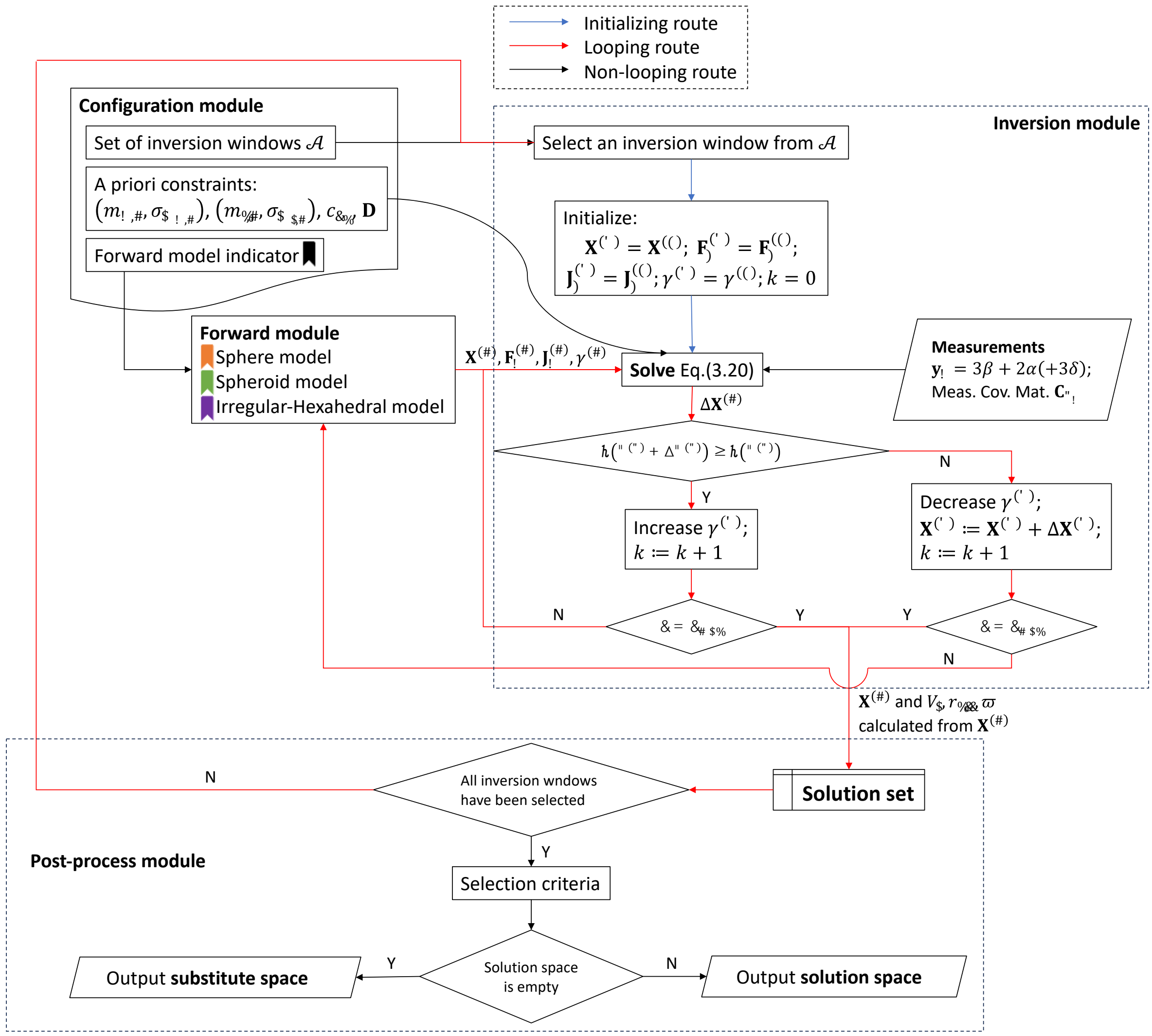


Figure 3.5. Flow chart of BOREAL algorithm

In the next chapter, the performance of BOREAL, retrieval accuracy and measurement sensitivity will be evaluated by ways of simulation.

4 Assessment of BOREAL performance by simulation and sensitivity study

In this chapter, we assess measurement sensitivity and retrieval accuracy of BOREAL by retrieving pre-defined aerosol models from synthetic data for both error-free and error-contaminated conditions. The aerosol models, noise model and strategy used for the simulation are described in Sect. 4.1. The simulation is conducted in two branches aiming at retrieving spherical and non-spherical particles, respectively, of which the results are separately presented in two sections (Sect. 4.2, Sect. 4.3). Sect. 4.4 summarizes the whole chapter.

4.1 Description of aerosol and noise models

4.1.1 Aerosol model

The assessment of algorithmic performance relies on synthetic aerosol optical properties. So, firstly, it is necessary to parameterize aerosol microphysical properties from which the synthetic optical properties can be calculated. As mentioned in previous chapters, by particle radius, atmospheric aerosol can be classified as Aitken nuclei ($r < 0.1$ μm), large nuclei (0.1 μm $\leq r \leq 1$ μm) and giant nuclei ($r > 1$ μm). There have been representative mathematical expressions to describe aerosol size distribution, such as negative power distribution (Junge, 1955), Γ distribution (Deirmendjian, 1969), modified-Γ distribution (Hansen and Travis, 1974). However, based on large numbers of in-situ measurements, Whitby (1978) proposed a summation of lognormal functions to describe aerosol size distribution in the whole size range. That study reveals that general aerosol ensembles are usually composed of three modes: a nucleation mode corresponding to the Aitken range, an accumulation mode corresponding to the large nuclei range and a coarse model corresponding to the giant nuclei range. These additive modes can be expressed by lognormal functions to form the total size distribution:

$$q(r) = \sum_i \frac{Q_i}{\sqrt{2\pi}(\ln \sigma_{g,i})} \exp\left[-\frac{(\ln r - \ln r_{q,i})^2}{2\ln^2 \sigma_{g,i}}\right], \tag{4.1}$$

where the subscript i represents the specific mode and the meaning of other parameters have been explained in Eq. (2.26). The measure of the distribution, q, could be number ($q = n$), surface area ($q = s$) and volume ($q = v$), whereas Q represents the concentration of the corresponding measure. As shown in Figure 4.1, the nucleation, accumulation and coarse modes respectively predominate in the number, surface area and volume (or mass) distributions.

Given that lidar measurements have little sensitivity to the nucleation mode (Burton et al., 2016) and the volume size distribution (VSD) is used throughout this study, we define lognormally distributed VSD models consisting of up to 2 modes: the accumulation (also called "fine") mode and coarse mode. This assumption is also consistent with large numbers of retrievals of typical aerosol types with sun-sky photometers (Dubovik et al., 2002). For spherical aerosols, we set 4 VSD types composed of a mono-fine (MF) mode, a mono-coarse (MC) mode, bi-modes with the fine part dominating (BF) and bi-modes with the coarse part dominating (BC), respectively. Detailed parameterization of these VSD types is given in Table 4.1.

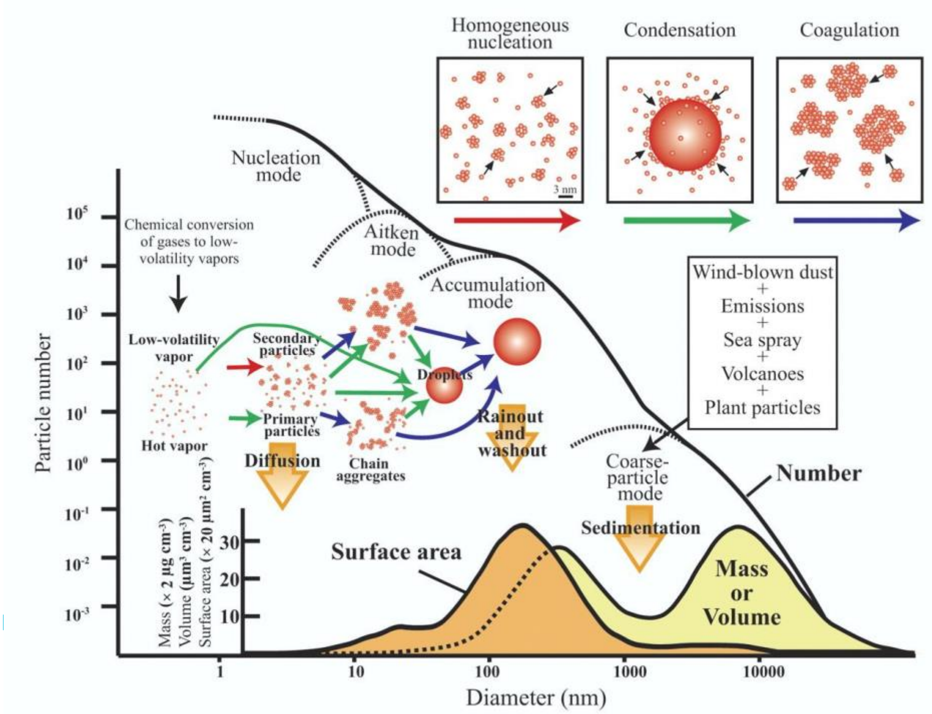


Figure 4.1. Schematic of number, surface area, volume or mass size distributions predominated by atmospheric aerosols with different modes. (cited from Buseck and Adachi (2008)).

For aerosols composed of non-spherical particles, we focus on the retrieval of dust aerosol. Due to the annual emission and residence time, mineral dust is the most abundant aerosol species which is mostly generated by natural processes such as wind blowing and can endure long-range transport (Ginoux et al., 2012). Typical size of dust particles ranges from hundreds of nanometers to tens of micrometers. Freshly emitted dust particles usually contain loose soil aggregates with diameter larger than 20 μm which, however, cannot stably exist due to processes of wind erosion and sandblasting as well as the high settling rate (Shao, 2008; Reid et al., 2008). Instead, a coarse mode with diameter between 1 and 20 μm predominates (Gomes et al., 1990, p.199; Alfaro et al., 1997). Reid et al. (2008) reported a volume median diameter of 3-5 μm for different dust sources and conclude that it is determined by regional soil properties rather than external factors such as wind speed. On the other hand, a fine mode with diameter smaller than 1 μm is sometimes observed (Dubovik et al., 2002; Osborne et al., 2008; Weinzierl et al., 2009; Mamouri and Ansmann, 2014). Theories like the scale-invariant fragmentation (Kok, 2011) and sandblasting (Gomes et al., 1990) can explain the injection of the fine mode. In addition, during transport processes, a shift to smaller sizes can happen due to the deposition of the coarse mode which, however, is not evident (Reid et al., 2008; Hu, 2018). Maring et al. (2003) report a 12% reduction of the measured VMD from 4.1 to 3.6 μm after probably 5-6 days transport, most of which took place in the first 2 days. Accordingly, we define three VSD types to mimic the size distribution of typical dust aerosols. Specifically, two monomodal VSDs differing in mode radius and a bimodal VSD with the coarse part predominating, represent freshly-emitted dust (FD), transported dust (TD) and bimodal dust (BD), respectively. Detailed parameterization can be found in Table 4.1.

Table 4.1 also shows the values of CRI for both spherical and non-spherical particles. For spherical particles, 5 spectrally independent m_{R} from 1.4 to 1.6, and m_{I} from 0.001 to 0.02 are defined. For non-spherical particles, since the aim is dust aerosol, the spectral dependence of m_{I} is considered. Specifically, three spectrally independent m_{R} and three $m_{\mathrm{I},355}$ corresponding to the lower extremes, means and upper extremes of the statistical results provided by Di Biagio et al. (2019) are set. The values of $m_{\mathrm{I},532}$ are determined by the regression relation shown in Figure 3.2. Finally, for $m_{\mathrm{I},1064}$, a single value of 0.001 is considered.

The *a priori* constraints on CRI are consistent with Table 3.1. For spherical particles, the *absorbing* type is assigned if the true $m_{\mathrm{I}} > 0.01$; otherwise, the *non-absorbing* type is assigned.

Table 4.1. Parameterization of the spherical and non-spherical aerosol models used in the simulation. The meaning of the parameters used to construct lognormal VSD can be found in Eq. (2.26).

	Spherical				**Non-spherical (dust)**		
VSD type	**MF**	**MC**	**BF**	**BC**	**FD**	**TD**	**BD**
V_f ($\mu\mathrm{m}^3/\mathrm{cm}^3$)	1	-	0.67	0.17	-	-	0.1
$r_{\mathrm{v,f}}$ (μm)	0.2	-	0.2	0.2	-	-	0.13
$\ln\sigma_{\mathrm{g,f}}$	1.49	-	1.49	1.49	-	-	1.49
V_c ($\mu\mathrm{m}^3/\mathrm{cm}^3$)	-	1	0.33	0.83	1	1	0.9
$r_{\mathrm{v,c}}$ μm	-	1.2	2	2	2	1	2
$\ln\sigma_{\mathrm{g,c}}$	-	1.82	1.82	1.82	1.82	1.82	1.82
r_{eff} ($\mu\mathrm{m}^3$)	0.18	0.99	0.26	0.7	1.67	0.84	0.69
m_R	1.4, 1.45, 1.5, 1.55, 1.6				1.4, 1.5, 1.6		
$m_{\mathrm{I},355}$	0.001, 0.005, 0.01, 0.015, 0.02				0.001, 0.005, 0.009		
$m_{\mathrm{I},532}$	$= m_{\mathrm{I},355}$				$= 0.52 m_{\mathrm{I},355}$		
$m_{\mathrm{I},1064}$	$= m_{\mathrm{I},355}$				0.001		

It is worth pointing out that in realistic situations, dust particles can be mixed with other more spherical aerosol particles so that the observed PLDR is lower than the pure case (Tesche et al., 2009). Many previous studies where the spheroid model was utilized to retrieve dust aerosol from lidar measurements tried to characterize the mixing state by assuming that the mixed ensemble contains a certain fraction of spherical particles that have the same VSD and CRI as the spheroidal particles. Correspondingly, an additional state variable, spherical volume fraction (SVF), is introduced and retrieved (Veselovskii et al., 2010; Müller et al., 2013; Tesche et al., 2019). Nevertheless, in this study, we decide not to retrieve SVF and only consider pure dust cases. The main reason is that the more unknown variables to retrieve, the higher the underdetermination of the inversion system. For instance, PLDR is sensitive to not only particle shape but also particle size (Gasteiger and Freudenthaler, 2014). Transported dust with smaller mode radius or bimodal dust containing submicron mode also presents lower PLDR. As will be seen in following sections, there are potential cross-talks between the state variables due to the underdetermination of the retrieval system. Retrieving SVF in addition will undoubtedly deepen the cross-talks. Indeed, Müller et al. (2013) and Tesche et al. (2019) retrieved unrealistic SVF if only inverting $3\beta + 2\alpha$ data. Although Tesche et al. (2019) claim that incorporating spectral PLDR improves the retrieval accuracy of SVF, we are careful about this parameter before comprehensive sensitivity study is conducted. Another technical difficulty precluding us from retrieving the mixing state is the lack of the a priori information on the CRI of the mixed aerosol. In this regard, new retrieval strategies utilizing proper particle mixing rules should be planned in the future.

4.1.2 Measurement noise model and simulation setup

For lidar measurements, random error (i.e., noise) in each channel is represented in the relative form. The influence of random error on retrieval accuracy is evaluated by assuming normally-distributed relative noise. Further, we assume the errors are independent of each other with a standard deviation equal to one third of the maximum measurement error (same as the corresponding standard deviation set in BOREAL, see Eq. (3.28)). Table 4.2 shows the relationship between the maximum measurement error and standard deviation set in BOREAL and used in the noise model.

Table 4.2. Maximum measurement errors for state-of-the-art lidar systems (Hu et al., 2019, 2020) and standard deviations set in BOREAL and used in the noise model which are derived from the assumption of Gaussian relative measurement error.

	Maximum measurement error	**Standard deviation set in BOREAL**	**Standard deviation used in the noise model**
α_{355}	10%	3.3%	3.3%
α_{532}	10%	3.3%	3.3%
β_{355}	10%	3.3%	3.3%
β_{532}	10%	3.3%	3.3%
β_{1064}	20%	6.7%	6.7%
δ_{355}	15%	5%	5%
δ_{532}	15%	5%	5%
δ_{1064}	15%	5%	5%

Different simulation strategies are employed to evaluate BOREAL performance of retrieving spherical and non-spherical particles. For spherical particles, the sphere model is used to generate synthetic $3\beta + 2\alpha$ data from the defined aerosol models in Table 4.1 and, at the same time, serves as the forward model of BOREAL to invert the generated $3\beta + 2\alpha$ data. For non-spherical particles, firstly, two sets of $3\beta + 2\alpha + 3\delta$ optical data are respectively generated by the spheroid and IH models from the defined aerosol models in Table 4.1. Then, within each set of optical data, the following configurations are applied in turns:

(1) ($3\beta + 2\alpha$, sphere) – i.e., inverting the $3\beta + 2\alpha$ data with the sphere model;

(2) ($3\beta + 2\alpha$, non-sphere) – i.e., inverting the $3\beta + 2\alpha$ data with the corresponding non-spherical model used for generating the optical data;

(3) ($3\beta + 2\alpha + 3\delta$, non-sphere) – i.e., inverting the $3\beta + 2\alpha + 3\delta$ data with the corresponding non-spherical model used for generating the optical data.

The simulation strategy for non-spherical particles permits evaluating not only the retrieval performance of the spheroid and IH models when the full ($3\beta + 2\alpha + 3\delta$) lidar measurements are inverted, but also assessing the influences of using the reduced measurements ($3\beta + 2\alpha$) and the sphere model on retrieval accuracy. Of course, for both spherical and non-spherical particle retrievals, apart from inverting the error-free optical data, additional 100 inversions of the simulated data perturbed by the noise model (error-contaminated data) are preformed to assess the influence of measurement error. The corresponding results are called the error-contaminated retrievals, of which the statistical properties are compared with the modeled values. Figure 4.2 displays the flowchart of the whole simulation.

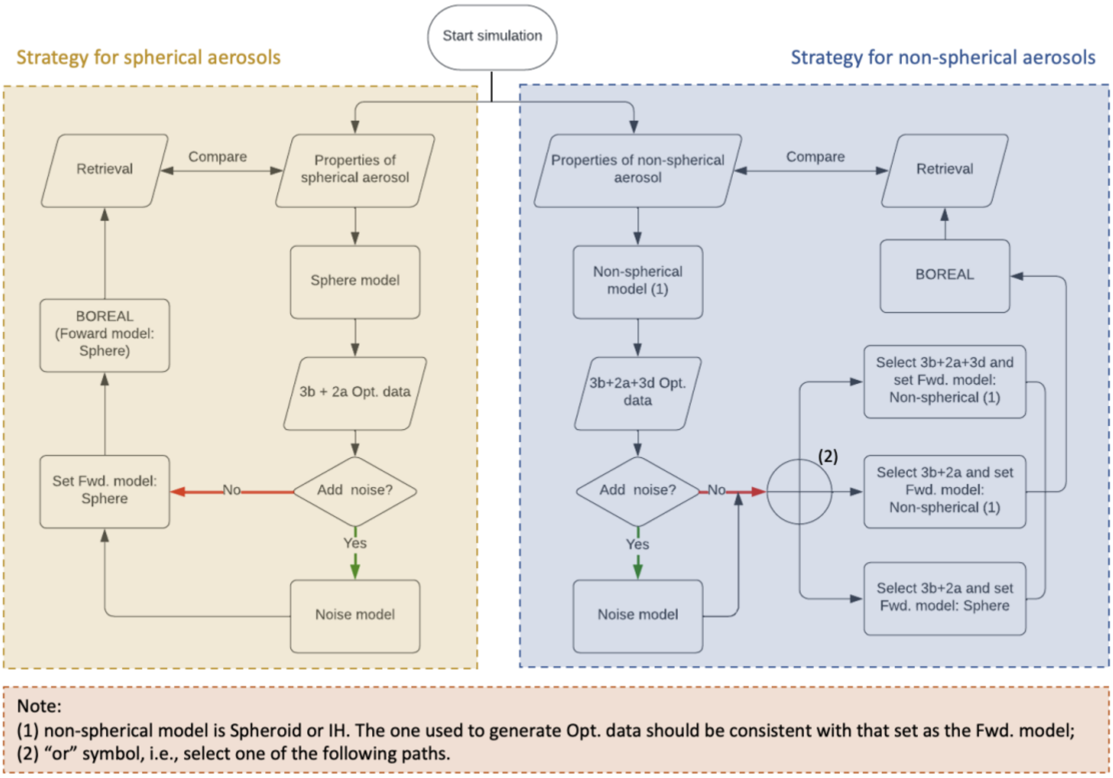


Figure 4.2. Flowchart of the simulation.

In the following sections, results of spherical particle and non-spherical particle retrievals will be presented separately. We will focus on the accuracy of retrieved VSD, CRI, SSA, as well as the size integral parameters, namely the total volume concentration, V_{t}, and effective radius, r_{eff}. In particular, we present measurement fitting error as an indicator of retrieval quality. The absolute and relative error of a retrieved quantity, x are defined as

$$\varepsilon(x) = \hat{x} - x^*,$$
$$\varepsilon(x) = \frac{\hat{x} - x^*}{x^*} \times 100\%, \tag{4.2}$$

respectively, where $\hat{x}$ represents the retrieved value and x^* indicates the true value. Eq. (4.2) is also used to describe fitting error when $\hat{x}$ represents the optical quantity recalculated from retrieved microphysical properties and x^* indicates the real measured optical quantity. We also utilize the root-mean-square error (RMSE) of a quantity that is expressed as a vector. For example, the RMSEs of spectral SSA and measurement fitting can be expressed as

$$\begin{aligned} \varepsilon(\varpi)_{\mathrm{RMS}} &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\widehat{\varpi_{\lambda_i}} - {\varpi_{\lambda_i}}^*\right)^2} \\ \varepsilon(\mathbf{y}_1)_{\mathrm{RMS}} = \varepsilon_{\mathrm{fit}} &= \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{\{\hat{\mathbf{y}}_1\}_i - \{\mathbf{y}_1^*\}_i}{\{\mathbf{y}_1^*\}_i}\right)^2} \end{aligned}. \tag{4.3}$$

4.2 Retrieval of spherical particles

4.2.1 Results under the error-free condition

The results presented in this section are all derived under the error-free condition. Figure 4.3 displays boxplots of $\varepsilon(V_{\mathrm{t}}), \varepsilon(r_{\mathrm{eff}}), \varepsilon(m_{\mathrm{R}}),$ and $\varepsilon(m_{\mathrm{I}})$ categorized by VSD type. The lower and upper whiskers correspond to the 5 and 95 percentiles of the datum, respectively. Considering both the deviation and dispersion of the retrieval errors, the retrieval of MF particles has the best accuracy, whereas that of BC particles has the worst. BOREAL performs better in the cases where fine-mode particles predominate than in the cases where coarse-mode particles predominate. Plus, a clear positive correlation between $\varepsilon(V_{\mathrm{t}})$ and $\varepsilon(r_{\mathrm{eff}})$ is observable, consistent with the analysis by Kolgotin et al. (2016). It can be seen that in not less than 90% of the cases V_{t} and r_{eff} can be retrieved with accuracies better than 30% and 35%; in not less than 70% of the cases better than 25% and 15%. Furthermore, in more than 90% of the MC cases and 75% of the BF cases, V_{t} and r_{eff} are underestimated. For m_{R}, it can be retrieved with an accuracy better than 0.05 in not less than 90% of the cases. For m_{I}, in not less than 70% of the BF cases, and not less than 90% of the rest cases, it can be retrieved with an accuracy better than 0.005.

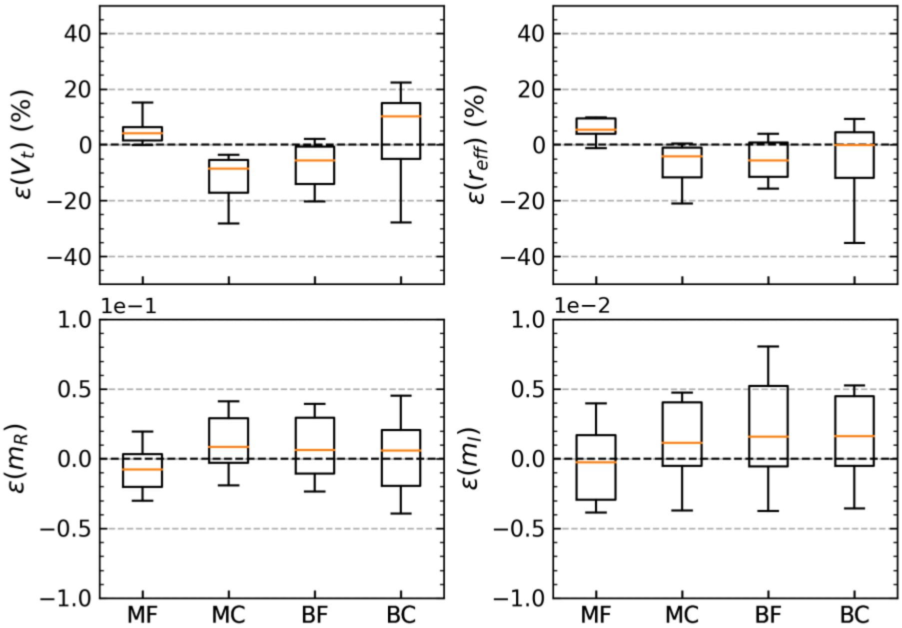


Figure 4.3. Boxplots of retrieval errors: $\varepsilon(V_\mathrm{t}), \varepsilon(r_\mathrm{eff}), \varepsilon(m_\mathrm{R})$, $\varepsilon(m_\mathrm{I})$ of the spherical cases in Table 4.1, categorized by the VSD type and under the error-free condition. The lower and upper whiskers of a box correspond to 5 and 95 percentiles of the datum, while the lower and upper box edges and the yellow inner line represent 25, 75 and 50 percentiles, respectively.

The upper panels of Figure 4.4 show boxplots of retrieval error of SSA with respect to the wavelength. Except for MC particles, SSA can be retrieved with accuracy better than 0.05 at all wavelengths in more than 90% of the cases. The SSAs of the bimodal types (BF and BC) are underestimated more often, especially for the BC type. The lower panels show the fitting error of each measurement. The fitting error in each channel is smaller than the corresponding measurement uncertainty set in BOREAL (Sect. 4.1.2), which indicates the convergence criteria (Sect. 3.4.3) work well for each retrieval. However, since these cases are free from measurement error (error-free), it is foreseen that the fitting error will evidently increase after introducing measurement error (error-contaminated) or when applying the algorithm to real lidar measurements. Next, we will discuss the retrievals of VSD and CRI in detail for each VSD type.

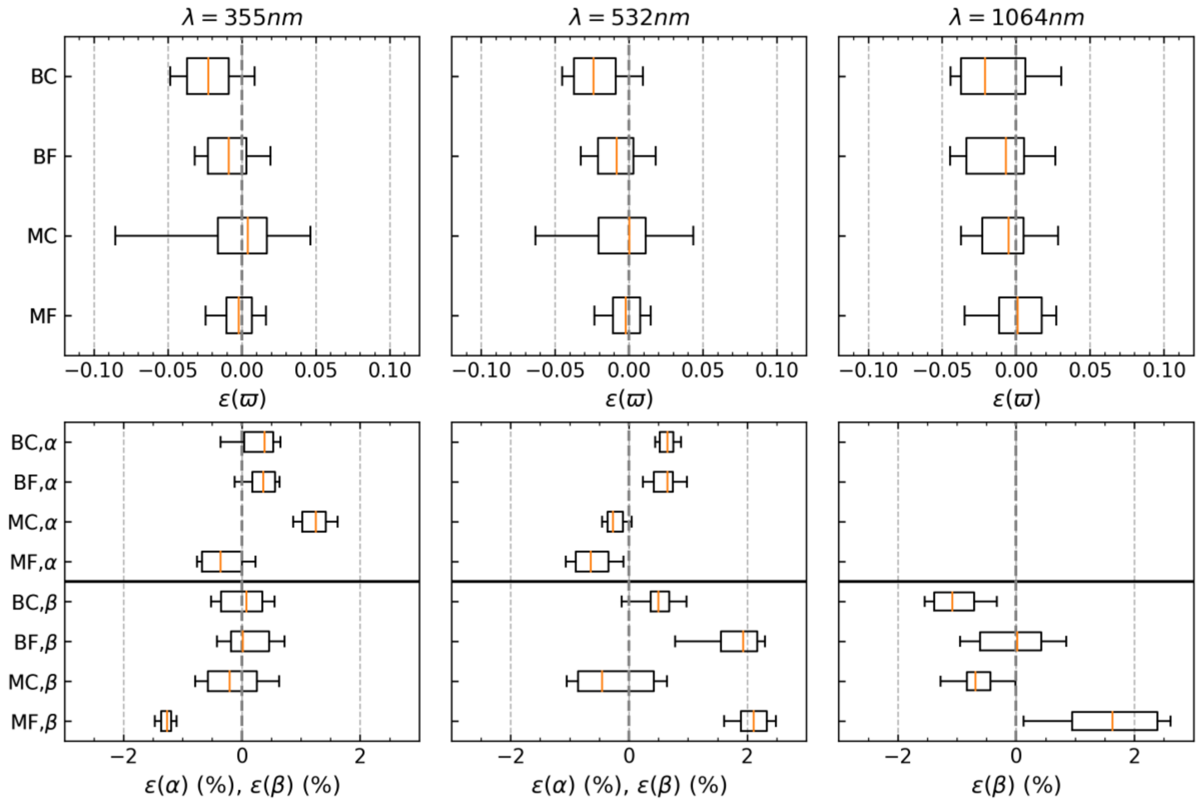


Figure 4.4. Same as Figure 4.3, but shows retrieval error of SSA ($\varepsilon(\varpi)$) and fitting error of each measurement with respect to wavelengths.

Figure 4.5 shows retrieved VSDs against variations of true m_{R} (denoted as m_{R}^*) and true m_{I} (denoted as m_{I}^*). As we will see by comparing with other VSD types, retrieval of MF particles is of the best quality and stability. As m_{R}^* and m_{I}^* change, r_{v}, and S are quite stable and precise, with slight perturbation on $v_{\max}$. Figure 4.6 shows the retrieval of CRI. One can clearly see the influence of the *a priori* value, $m_{\mathrm{I,a}}$, on the retrieval of m_{I} – the results are categorized into two clusters corresponding to the used *a priori* constraints. In particular, the imaginary parts are retrieved quite close to the *a priori* value of 0.005 for all the cases where the true value, m_{I}^*, is 0.001. This suggests the lack of sensitivity of the $3\beta + 2\alpha$ measurements to the m_{I} below 0.005, which can be further demonstrated by small variation rate of the measurement to the change of m_{I}. The variation rate of the measurement is defined as the maximum relative variation within an interval of a state variable:

$$\rho(y) = \frac{2(\max\{y(x)\} - \min\{y(x)\})}{\max\{y(x)\} + \min\{y(x)\}} \times 100\%, \quad x \in [a, b]. \tag{4.4}$$

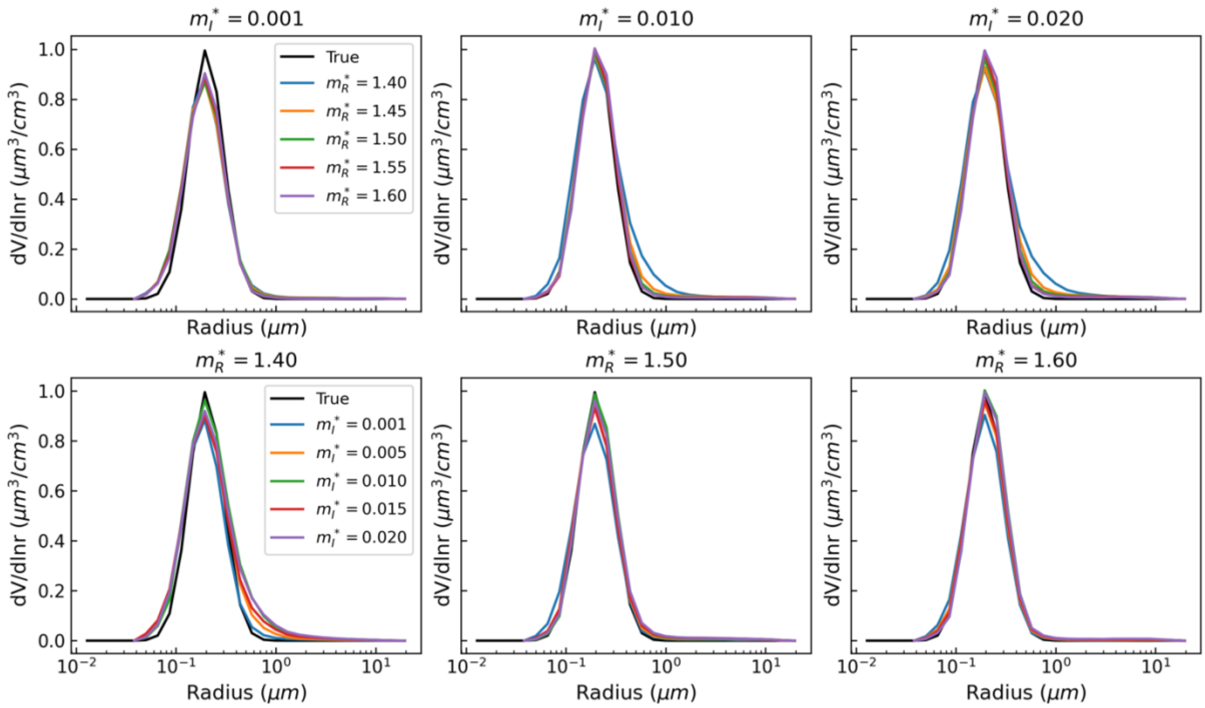


Figure 4.5. Results of VSD retrieval for the MF particle ensemble under the error-free condition, shown as slices of the true imaginary part (m_{I}^*) (upper panels) and true real part (m_{R}^*) (lower panels).

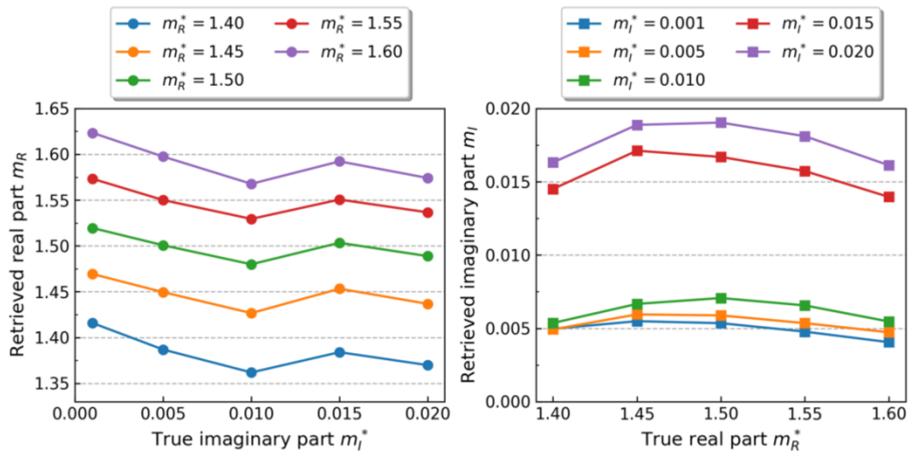


Figure 4.6. Results of CRI retrieval for the MF particle ensemble under the error-free condition, shown of functions of the true CRI.

Figure 4.7 displays the variation rate of the $3\beta + 2\alpha$ data to $m_{\mathrm{I}} \in [0.001, 0.005]$. It can be seen that compared with backscattering coefficients, extinction coefficients always have low variation rates less than 2% for all values of m_{R}. The variation rates of backscattering coefficients increase with the decrease of wavelength and increase of m_{R}, but most of the time do not exceed the prescribed measurement error. This indicates that in a real retrieval of fine-mode particles with m_{I}^* less than 0.005, changes in $3\beta + 2\alpha$ measurements caused by the variation of m_{I} and noise are difficult to be distinguished. In addition, simultaneously variations of the state variables can lead to compensating changes of the measurements, which is referred to as the "cross-talk" between the state variables by Burton et al. (2016). The cross-talk stems from the underdetermination of the inversion system and results in non-unique solutions. A study on covariance and correlation matrices of the state variables by Burton et al. (2016) shows that retrieved m_{R} and m_{I} are highly positive correlated, which is consistent with the results here: as shown in Figure 4.6, overestimation of m_{I} is accompanied by overestimation of m_{R}, and vice versa.

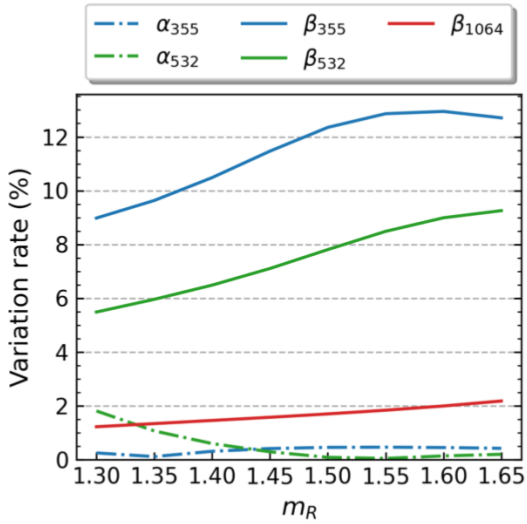


Figure 4.7. Variation rate (see Eq. (4.4) for definition) of $3\beta + 2\alpha$ optical data to m_{I} within the range from 0.001 to 0.005, with respect to the variation of m_{R}. The VSD of the MF ensemble is used for the calculation.

Figure 4.8 and Figure 4.9 show the retrieval of VSD and CRI for the MC ensemble. Although the VSD retrieval accuracy is worse than that for the MF ensemble, the mono-coarse-mode is still able to be identified. Compared with the MF ensemble, m_{R} of the MC is more overestimated in the cases where m_{R}^* and m_{I}^* are both low. For the cases where $m_{\mathrm{R}}^* > 1.4$ and $m_{\mathrm{I}}^* > 0.001$, the retrieval accuracy of m_{R} is comparable with the corresponding MF cases. On the other hand, except for $m_{\mathrm{I}}^* = 0.001, 0.005$ which measurements are probably not sensitive to, the retrieval of the imaginary part shows a dependence on both $m_{\mathrm{I,a}}$ and m_{R}^* (the retrieved m_{I} decreases with the increase of m_{R}^*), while it seems to be less relevant to the true value, m_{I}^*.

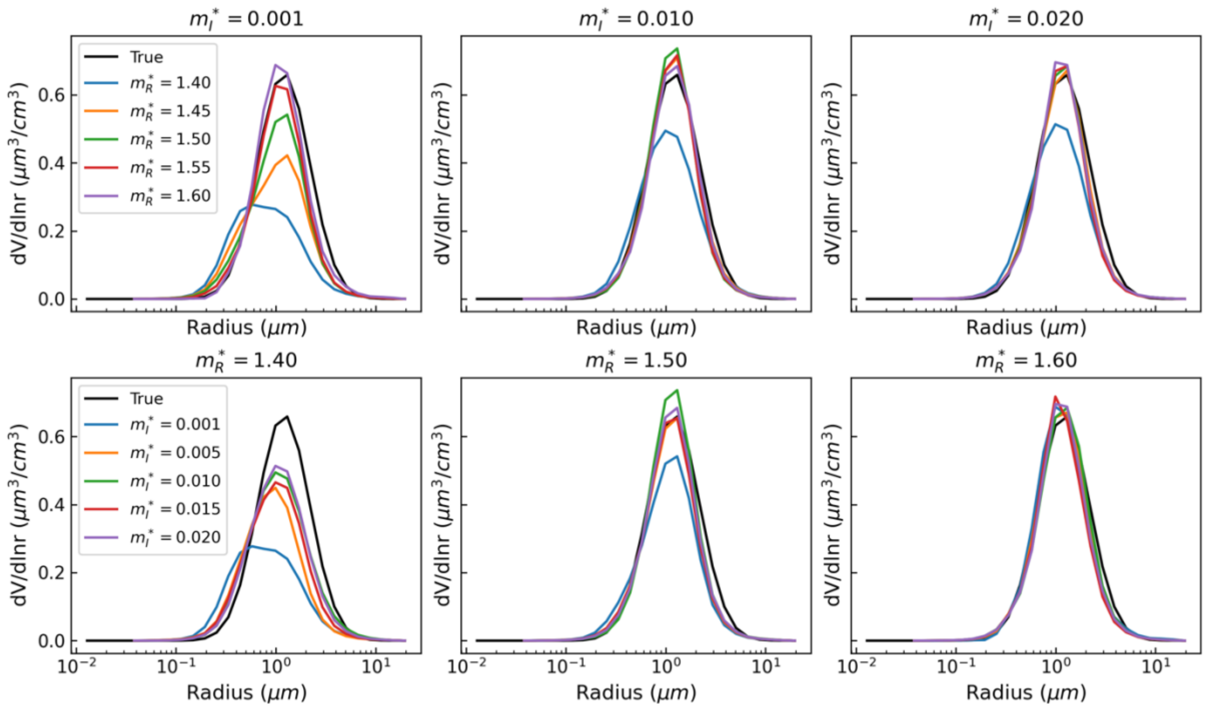


Figure 4.8. Same as Figure 4.5 but for the MC ensemble.

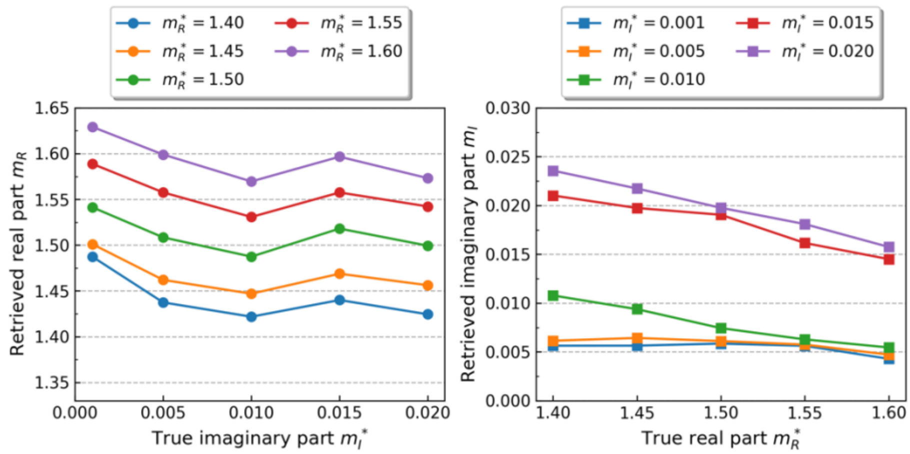


Figure 4.9. Same as Figure 4.6 but for the MC ensemble

The distribution of the CRI retrieval suggests a combining impact of the *a priori* value and cross-talk on retrieval accuracy. To further demonstrate this, we select 4 cases where the true and retrieved values of CRI are listed in Table 4.3. Several common features can be extracted from these cases: (1) m_{R} and m_{I} are simultaneously over- or under- estimated; (2) the retrieved imaginary parts cluster around their used *a priori* values; (3) from Case 1 to 3, VSDs are retrieved with good quality and (4) in Case 4, the retrieved real part is also close to the *a priori* value and the VSD is highly underestimated.

Table 4.3. True CRI (m_R^*, m_I^*) and retrieved CRI $(\hat{m}_R, \hat{m}_I)$ of the selected cases to demonstrate the cross-talk between the real and imaginary parts of the CRI.

	Case 1	Case 2	Case 3	Case 4
(m_R^*, m_I^*)	(1.6, 0.001)	(1.6, 0.01)	(1.6, 0.02)	(1.4, 0.001)
$(\hat{m}_R, \hat{m}_I)$	(1.63, 0.004)	(1.57, 0.005)	(1.57, 0.016)	(1.49, 0.006)

Figure 4.10 shows the contour plots of β_{355}, β_{532} and β_{1064} against m_R and m_I. The backscattering coefficients were calculated from the MC ensemble and normalized to the maximum. The variation rates of β_{355}, β_{532}, β_{1064} between two adjacent contour levels are 10%, 10% and 20%, respectively, corresponding to the maximum measurement errors used in BOREAL. The positive cross-talk between m_R and m_I is clearly shown by the inclined patterns. The figure also displays the coordinates of the true (solid markers) and retrieved (hollow markers) states of the four selected cases. One can see that for the first three cases, the distributions of the true and retrieved states are almost perpendicular to the gradient of the state variables, indicating little sensitivity of the measurements along these directions. Whereas, in the last case, the distance between the true and retrieved states crosses several levels. Table 4.4 lists the relative changes in $3\beta + 2\alpha$ due to the variation of CRI from the true to retrieved state while the VSD keeps unchanged. Compared with the first three cases, considerable variations of the optical properties in Case 4 are found.

The analysis of the selected cases reveals how the *a priori* constraint on CRI and cross-talk impact retrieval accuracy together. For intermediate or high real parts (i.e., $m_R^* \geq 1.5$), $\hat{m}_I$ is close to the *a priori* value, $m_{I,a}$, while $\hat{m}_R$ is more determined by the cross-talk causing little variation in optical properties; as a result, VSD can be retrieved with reliable accuracy in these cases. To the contrary, for low real parts (i.e., $m_R^* \sim 1.4$), both $\hat{m}_R$ and $\hat{m}_I$ are close to the *a priori* values, $m_{R,a}$ and $m_{I,a}$, and the cross-talk happens between VSD and CRI; as a result, (1) $\hat{m}_R$ is usually overestimated since $m_{R,a} = 1.5$ in this study and (2) V_t is underestimated. We stress that the intrinsic reasons for cross-talk and "*a-priori*-constraint effect" are underdetermination and lack of sensitivity of the lidar inversion system. Thus, it is of significant importance to acquire as accurate as possible *a priori* constraints on CRI, especially on m_I.

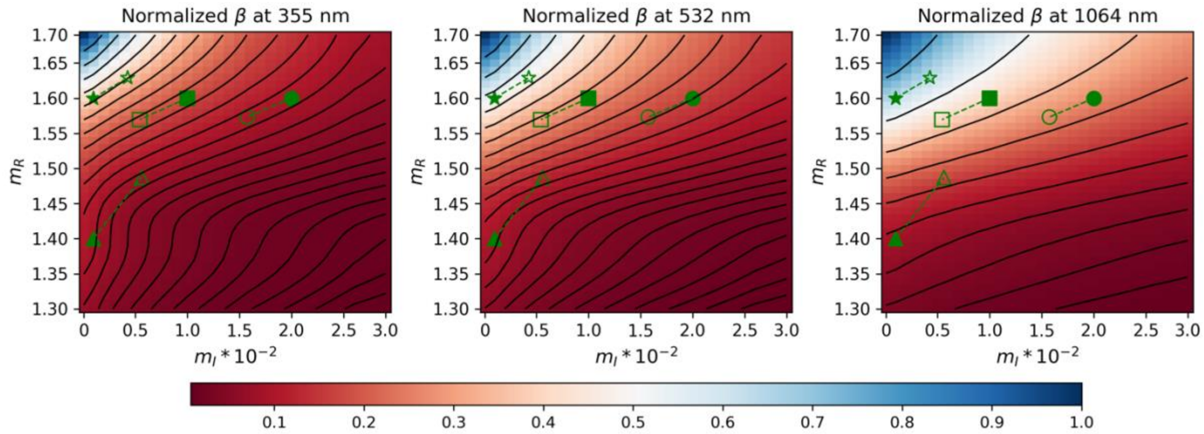


Figure 4.10. Contour plots of β_{355}, β_{532} and β_{1064} against m_{R} and m_{I} and normalized to the maximum. The VSD of the MC ensemble is used for the calculation. The variation rates of β_{355}, β_{532}, β_{1064} between two adjacent contour levels are 10%, 10% and 20%, respectively. There are also shown the coordinates of the true CRI (solid markers) and retrieved CRI (hollow markers) for Case 1 (star markers), Case 2 (square markers), Case 3 (circle markers) and Case 4 (triangle markers).

Table 4.4. Relative changes in the 3β + 2α measurements when m_{I} varies from the true to the retrieved value.

	α_{355}	α_{532}	β_{355}	β_{532}	β_{1064}
Case 1	-0.2%	-0.5%	-4.0%	-3.6%	-0.4%
Case 2	0.2%	0.5%	5.6%	3.8%	-0.8%
Case 3	0.2%	0.5%	4.4%	3.2%	-1.0%
Case 4	-1.5%	-2%	-10.3%	31.8%	91.9%

Figure 4.11 and Figure 4.12 show the retrieval of VSD and CRI for the BF ensemble. The fine-mode part of VSD can be retrieved with good accuracy in all cases, whereas the accuracy of the coarse-mode part depends on m_{R}^* and m_{I}^*. In general, the coarse-mode part of VSD is retrieved with higher accuracy when both the true real and imaginary parts are high ($m_{\mathrm{R}}^* \geq 1.5$ and $m_{\mathrm{I}}^* \geq 0.01$). To the contrary, the coarse-mode radius, $r_{\mathrm{v,c}}$, or the coarse-mode volume concentration, V_{c}, is more underestimated when either the true real part or imaginary part is low ($m_{\mathrm{R}}^* < 1.5$ or $m_{\mathrm{I}}^* < 0.01$), which is the cause of negative $\varepsilon(V_{\mathrm{t}})$ and $\varepsilon(r_{\mathrm{eff}})$ in Figure 4.3. The worst retrieval of the coarse-mode VSD corresponds to $m_{\mathrm{R}}^* = 1.4$ and $m_{\mathrm{I}}^* = 0.001$. Note that it is in these cases that the largest retrieval error occurs for MC particles, as mentioned above. However, the situation is more complicated here because apart from the cross-talks between the real and imaginary parts of CRI, the CRI and a single mode, there is also cross-talk between the fine and coarse modes, particularly given that the optical kernels are of different magnitudes in different ranges of the size parameter. The pattern of $\widehat{m}_{\mathrm{R}}$ is between that for the MF particles and that for the MC particles. Whereas the pattern of $\widehat{m}_{\mathrm{I}}$ is more similar to that for the MC particles, with values that gradually decrease with the increase of m_{R}^*.

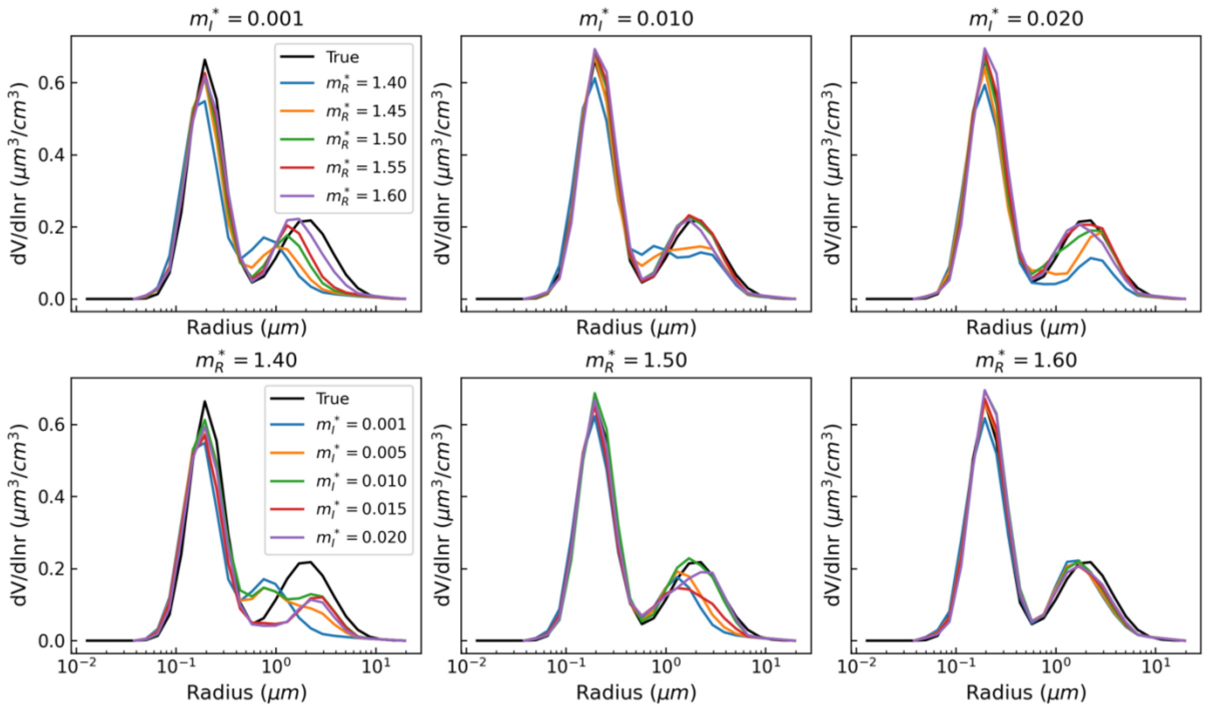


Figure 4.11. Same as Figure 4.5 but for the BF ensemble.

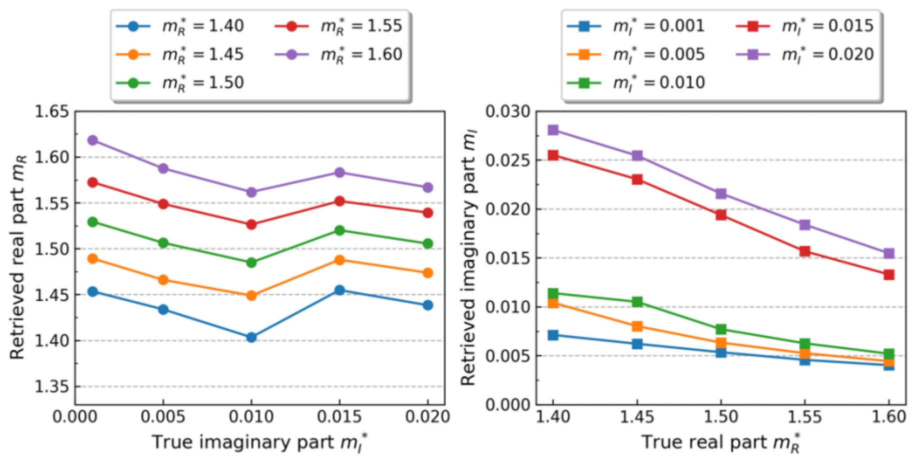


Figure 4.12. Same as Figure 4.6 but for the BF ensemble

Figure 4.13 and Figure 4.14 show the retrieval of VSD and CRI for the BC ensemble. One can see that retrieval of the fine-mode VSD is again quite stable. Nevertheless, the fine-mode geometric standard deviation, $\sigma_{\mathrm{g,f}}$, and the maximum value, $v_{\mathrm{max,f}}$ are over- and under-estimated, respectively. By contrast, the coarse-mode parameters are less stable and can be over- or under- estimated depending on the true CRI. In general, $V_{\mathrm{t,c}}$ keeps growing as m_{R}^* and m_{I}^* grow, resulting in underestimates at first and overestimates afterwards. On the other hand, with increases of m_{R}^* and m_{I}^*, $r_{\mathrm{v,c}}$ and $\sigma_{\mathrm{g,c}}$ are retrieved with increasing accuracy. The pattern of the CRI retrieval is quite similar to that for the BF ensemble, except for the cases where

$m_R^* \leq 1.45$ and $m_I^* = 0.001$. In these cases, the influence of the coarse-mode particles predominates, leading to a pattern more similar to that for the MC ensemble: namely, the highly overestimated m_R and m_I accompanied by the severe underestimated $V_{t,c}$.

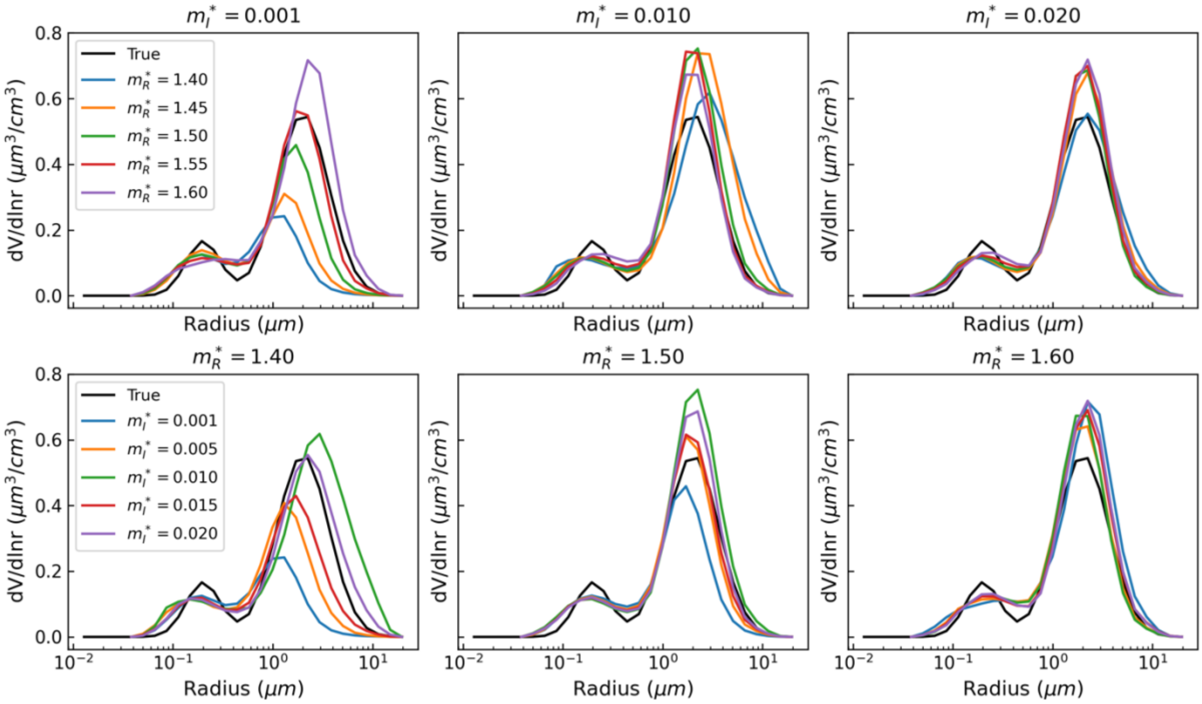


Figure 4.13. Same as Figure 4.5 but for the BC ensemble

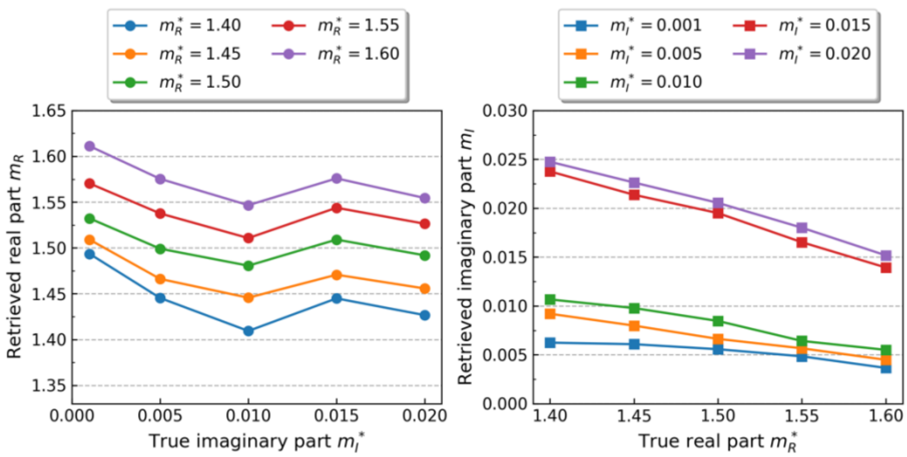


Figure 4.14. Same as Figure 4.6 but for the BC ensemble

The results and discussion above show the microphysical properties of the mono-fine-mode (MF) particles can be retrieved with the highest accuracy and stability by BOREAL. For bimodal distributed particles, the fine-mode part is still able to be retrieved with enough accuracy and stability. Whereas the retrieval quality of coarse-mode particles depends on particle CRI: in general, the largest retrieval error for coarse-mode particles emerges at the lowest m_{R}^{*} and m_{I}^{*}, and the retrieval error decreases with the increase of both parts of the CRI. We conclude that without considering measurement noise and error in the forward model, the main issues that retrieving spherical particles from the $3\beta + 2\alpha$ data faces are non-unique solutions and cross-talk between state variables which stem from both lack of sensitivity and underdetermination of the retrieval system. Accordingly, it is important to increase the accuracy of the *a priori* constraint on CRI, particularly on the imaginary part, and increase the information content by incorporating more measurements. The former could be realized with the help of more comprehensive climatology data; the latter is proved by incorporating the three depolarization measurements (3δ) at 355, 532 and 1064 nm in non-spherical particle retrieval, which will be discussed in detail in Sect. 4.3.

4.2.2 Influence of measurement noise

The $3\beta + 2\alpha$ optical dataset are perturbed by the noise model described in Sect. 4.1.2 for 100 times to form the error-contaminated optical data. Statistics of the retrievals for all error-contaminated data reveals the influence of measurement noise. Figure 4.15 and Figure 4.16 show boxplots of retrieval error as well as fitting error when inverting the error-contaminated data, just like the error-free cases shown in Figure 4.3 and Figure 4.4. Compared with the results from the error-free optical data, dispersion of the major parts (~65%) of the retrieval errors does not change much, indicating the robustness of retrieval. It suggests that although the incorporated constraints (both smoothing and *a priori* constraints) may bias the retrieval from the correct value to some extent, they are important to suppress the influence of measurement error. What is essential is to improve the accuracy of the constraints, particular the *a priori* constraint on m_{I}. As for the measurement fitting, errors in spectral extinction coefficients apparently enlarge while those in backscattering almost keep unchanged.

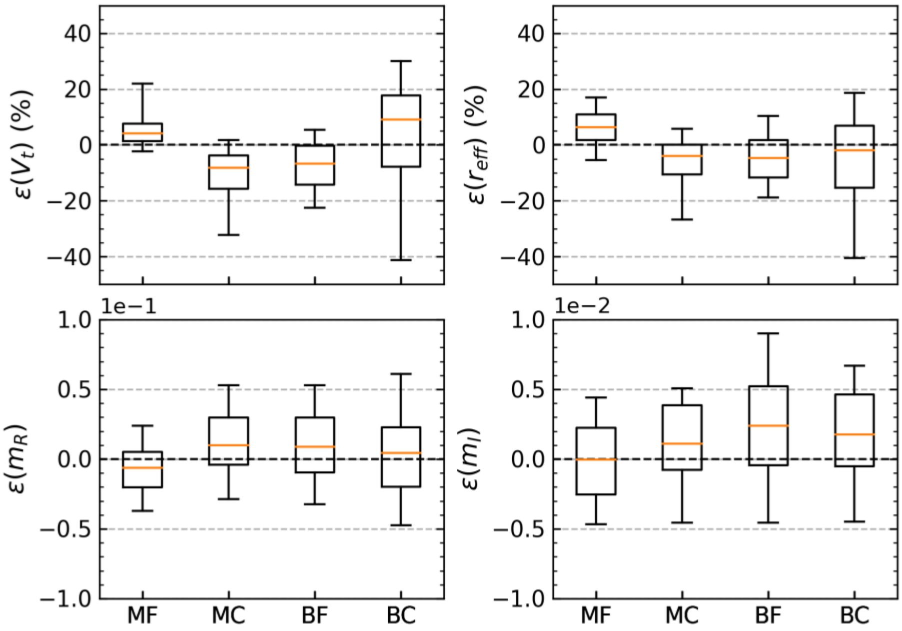


Figure 4.15. Same as Figure 4.3 but under the error-contaminated condition.

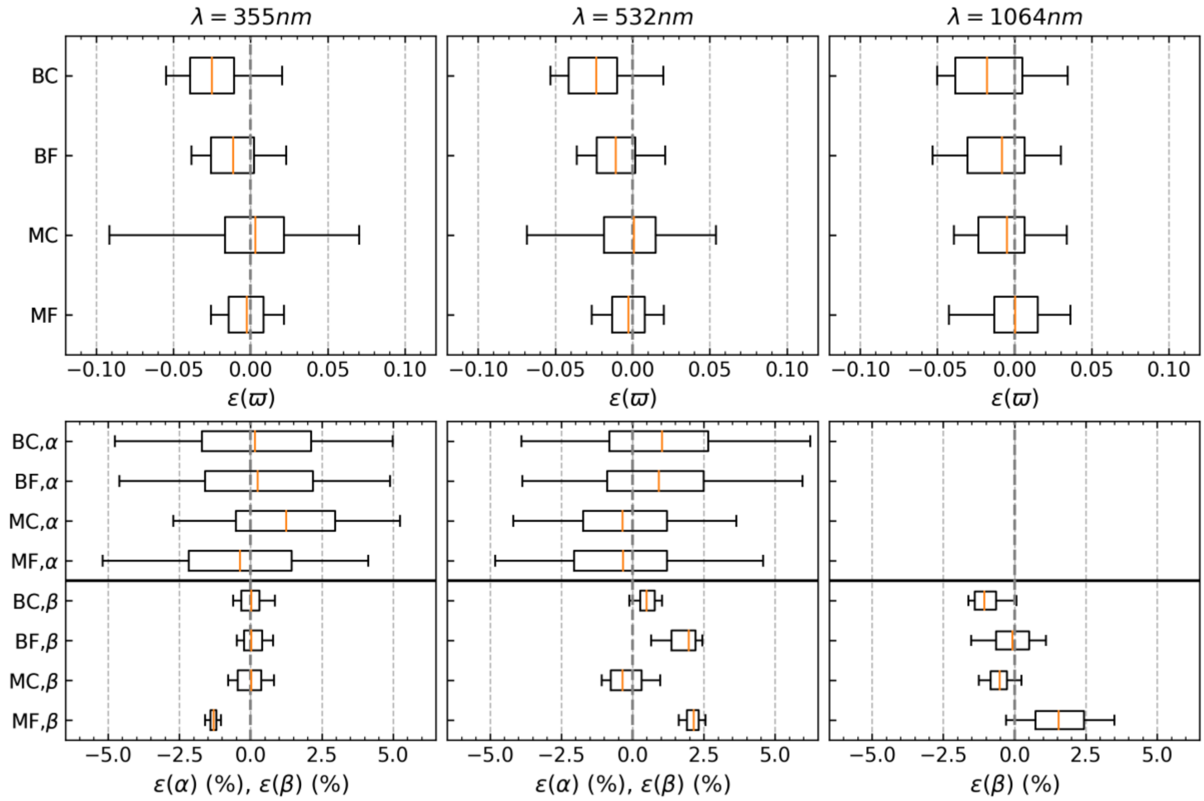


Figure 4.16. Same as Figure 4.4 but under the error-contaminated condition.

To further compare retrieval accuracies under error-free and error-contaminated conditions, Table 4.5 shows the magnitudes of the means and standard deviations of the errors in retrieving $V_t, r_{eff}, m_R, m_I, \varpi$, together with the mean and standard deviation of the fitting error, ε_{fit} are calculated. The mean and standard deviation of a retrieval error for a VSD type are derived by performing statistical calculation to all results corresponding to this VSD type. The summation of the two quantities is taken as a measure of the total retrieval accuracy for that retrieval variable. It once again demonstrates the noise level for $3\beta + 2\alpha$ measurements from state-of-the-art lidar systems does not has a strong influence on the retrieval stability of BOREAL. The extent of accuracy reduction is comparable or less than the standard deviation of the noise.

Table 4.5. Comparison of the retrieval accuracy derived under the error-free (EF) condition and error-contaminated (EC) condition. The magnitudes of the mean and standard deviation are respectively the first and second terms inside the paratheses, and the total retrieval accuracy is the number outside the paratheses.

	VSD Type	$\varepsilon(V_t)(\%)$	$\varepsilon(r_{eff})(\%)$	$\varepsilon(m_R) \times 10^{-2}$	$\varepsilon(m_I) \times 10^{-3}$	$\varepsilon(\varpi)_{RMS} \times 10^{-2}$	$\varepsilon_{fit}(\%)$
EF	**MF**	12 (6+6)	10 (6+4)	3 (1+2)	3 (0+3)	2 (1+1)	2 (2+0)
	MC	22 (12+10)	16 (7+9)	4 (1+3)	4 (1+3)	5 (3+2)	1 (1+0)
	BF	16 (8+8)	12 (5+7)	4 (1+3)	6 (2+4)	3 (2+1)	1 (1+0)
	BC	22 (2+20)	23 (7+16)	4 (1+3)	5 (2+3)	4 (3+1)	1 (1+0)
EC	**MF**	13 (6+7)	13 (6+7)	3 (1+2)	3 (0+3)	2 (1+1)	2 (2+0)
	MC	22 (11+11)	16 (6+10)	4 (1+3)	4 (1+3)	5 (3+2)	1 (1+0)
	BF	16 (7+9)	14 (5+9)	4 (1+3)	6 (2+4)	3 (2+1)	1 (1+0)
	BC	24 (3+21)	23 (5+18)	3 (0+3)	5 (2+3)	4 (3+1)	1 (1+0)

4.2.3 Retrieval accuracy for typical aerosol types

In previous sections, retrieval of the aerosol models covering comprehensive ranges of state variables helps in establishing a general analysis on BOREAL's performance. However, some of the aerosol models are not very common in nature. For instance, the MC or BC ensemble of spherical particles is more likely to be non-absorbing sea salt aerosol of which both real and imaginary parts of refractive index are unlikely to be very high; if the MF ensemble is composed of biomass-burning aerosol with a high content of black carbon, it is less possible to have a low imaginary part. For practical usage, it is worth evaluating the retrieval accuracy for several typical aerosol types based on the above retrieval results. Therefore, based on the study by Dubovik et al. (2002), a subset of the previously defined aerosol models is extracted. The elements of the subset represent biomass-burning, urban and oceanic aerosols. Moreover, according to the type of burnt vegetation, the biomass-burning aerosol is further divided into the grass-burning aerosol (GBA) and forest-burning aerosol (FBA). Plus, by the difference in chemical composition, the urban aerosol rich in water-soluble particles from industrial activities is designated as the water-soluble urban aerosol (WUA), while that generated by fossil fuel combustion is designated as the combustion urban aerosol (CUA). The microphysical properties of these aerosol types are listed in Table 4.6.

Table 4.6. Microphysical properties of typical aerosol models. Note that they are extracted from Table 4.1 and form a subset of the predefined aerosol models.

Aerosol type	m_R	m_I	**VSD type**
Grass-burning aerosol (GBA)	1.5	0.015	MF, BF
Forest-burning aerosol (FBA)	1.5	0.005	
Water-soluble-urban aerosol (WUA)	1.4	0.005	
Combustion-urban aerosol (CUA)	1.45	0.01	
Oceanic aerosol (OA)	1.4	0.001	MC, BC

Table 4.7 lists magnitudes of retrieval errors for each aerosol type when the optical data is error-free. It is no doubt that the largest error occurs in the OA retrieval since compared to others, this type is characterized by small m_R, m_I and large r_v. Retrieval errors in V_t, r_{eff} and m_R for GBA, FBA, WUA and CUA are all not greater than 15%, 15% and 0.03, respectively. In all cases, spectral SSA can be retrieved with an accuracy better than 0.04. In general, retrieval accuracy for monomodal types is superior to that for bimodal types, except for CUA. Table 4.8 lists the retrieval errors under the error-contaminated condition. The statistics of each 100-runing set is represented by the mean and standard deviation (in the paratheses). It can be seen that the means do not have systematical bias from the values shown in Table 4.8 and the standard deviations are comparable with or less than the noise level.

Table 4.7. Magnitudes of retrieval errors for each aerosol type under the error-free condition.

Aero. type and sub-type	**Biomass-burning aerosol**				**Urban aerosol**				**Oceanic aerosol (OA)**	
	GBA		**FBA**		**WUA**		**CUA**			
VSD type	**MF**	**BF**	**MF**	**BF**	**MF**	**BF**	**MF**	**BF**	**MC**	**BC**
$\varepsilon(V_t)(\%)$	3	10	2	9	12	14	10	5	44	49
$\varepsilon(r_{eff})(\%)$	6	7	5	6	3	13	11	2	37	47
$\varepsilon(m_R) \times 10^{-2}$	0.3	2	0.1	1	1	3	2	0.1	9	9
$\varepsilon(m_I) \times 10^{-3}$	2	4	0.9	1	0.1	5	3.3	1	5	5
$\varepsilon(\varpi)_{RMS} \times 10^{-2}$	1	2	1	1	0.2	3	2	0.3	4	4

Table 4.8. Same as Table 4.7, but under the error-contaminated condition. Mean and standard deviation (inside the paratheses) of 100 retrievals are shown.

Aero. type and sub-type	Biomass-burning aerosol				Urban aerosol				Oceanic aerosol (OA)	
	GBA		FBA		WUA		CUA			
VSD type	MF	BF	MF	BF	MF	BF	MF	BF	MC	BC
$\varepsilon(V_{\mathrm{t}})(\%)$	3 (2)	11 (4)	3 (2)	8 (3)	11 (7)	15 (4)	10 (4)	4 (2)	43 (3)	49 (2)
$\varepsilon(r_{\mathrm{eff}})(\%)$	6 (5)	8 (3)	5 (6)	5 (6)	5 (6)	12 (4)	12 (5)	0.1 (6)	36 (4)	46 (5)
$\varepsilon(m_{\mathrm{R}}) \times 10^{-2}$	0.4 (0.4)	2 (0.2)	0 (0.4)	0.7 (0.4)	1 (1)	3 (0.5)	2 (0.8)	0.2 (0.2)	9 (0.5)	9 (0.4)
$\varepsilon(m_{\mathrm{I}}) \times 10^{-3}$	2 (1)	5 (1)	1 (0.6)	2 (0.5)	0 (0.4)	5 (0.4)	3 (1)	0.3 (0.4)	5 (0.2)	5 (0.1)
$\varepsilon(\varpi)_{\mathrm{RMS}} \times 10^{-2}$	1 (1)	2 (1)	0.7 (0.3)	0.8 (0.3)	0.3 (0.2)	0.3 (0.2)	2 (0.4)	0.4 (0.2)	4 (0.1)	4 (0.2)

4.3 Retrieval of non-spherical particles (mineral dust aerosol)

4.3.1 Scattering properties computed with different scattering models

Since this section involves the use of different scattering models, namely the Sphere, Spheroid and IH models, before analyzing retrieval processes, it is necessary to investigate the comparability of the relevant scattering properties computed with the different scattering models by forward simulation.

Figure 4.17 shows bulk phase matrix elements $\langle P_{11}\rangle$, $\langle P_{12}\rangle$ and $\langle P_{22}\rangle$ simulated with the IH, Spheroid and Sphere models at 532 nm. The phase function, $\langle P_{11}\rangle$, is normalized to the value at a scattering angle of 30°, while $\langle P_{12}\rangle$ and $\langle P_{22}\rangle$ are normalized to $\langle P_{11}\rangle$. They are calculated from a particle ensemble that has a monomodal lognormal distribution with $r_{\mathrm{eff}} = 1.5\mu\mathrm{m}$, $m_{\mathrm{R}} = 1.5$ and $m_{\mathrm{I}} = 0.0015$. These values of the microphysical properties of the ensemble are close to typical dust aerosol according to Dubovik et al. (2002) and Di Biagio et al. (2019). It can be seen that there is difference in $\langle P_{11}\rangle$ between spherical and non-spherical particles for scattering angle greater than 80°, while significant differences in $\langle P_{12}\rangle$ and $\langle P_{22}\rangle$ are observable across the whole scattering angle range. On the other hand, the angular variations of the phase matrix elements simulated with the Spheroid and IH models are in general consistent. In particular, the quite similar $\langle P_{11}\rangle$ produced by the two non-spherical models agree with the conclusion drawn by Mishchenko et al. (2002) that detailed shape and orientation of a single particle is not important to the bulk scattering intensity due to the averaging effect. The most noticeable difference between the two non-spherical models emerges for $\langle P_{22} \rangle$ – the simulated angular distributions show a same variation trend but with different magnitudes.

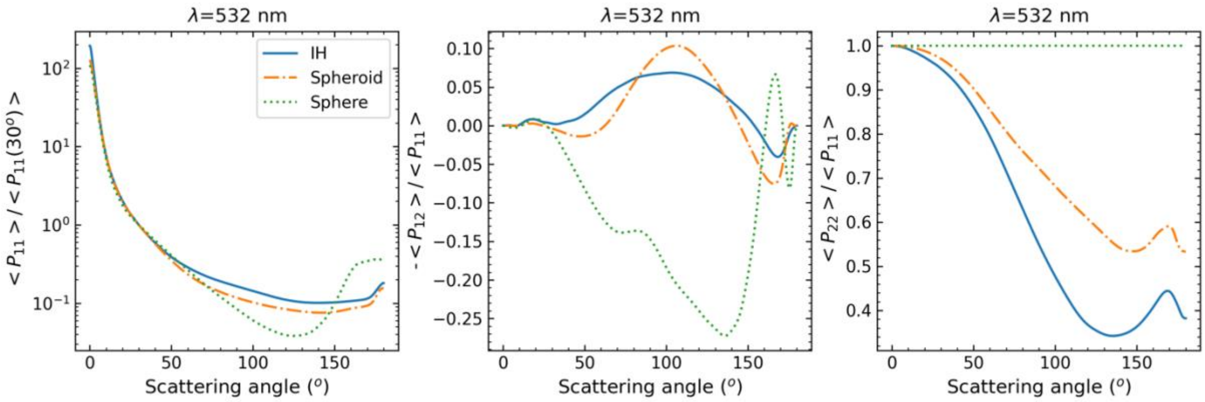


Figure 4.17. Bulk phase matrix elements $\langle P_{11} \rangle$, $\langle P_{12} \rangle$ and $\langle P_{22} \rangle$ simulated with the IH, Spheroid and Sphere models at 532 nm, calculated from a particle ensemble with monomodal lognormal VSD (r_{eff} = 1.5 µm) and CRI=1.5-0.0015i. $\langle P_{11} \rangle$ is normalized to the value at a scattering angle of 30°, while $\langle P_{12} \rangle$ and $\langle P_{22} \rangle$ are normalized to $\langle P_{11} \rangle$.

Figure 4.18 shows bulk SSA and asymmetry factor against the size parameter, for $m_{\mathrm{R}} = 1.5$ but different values of m_{I}, simulated with the IH, Spheroid and Sphere models. The varying size parameter is calculated from $\lambda = 355\mathrm{nm}$ and a series of VSDs having the same σ_{g} with the VSD used in Figure 4.17 but different mode radii. The SSAs computed by different models do not show significant difference for $m_{\mathrm{I}} = 0.001$ or size parameter less than 10. As the size parameter and m_{I} become larger, compared to the non-spherical models which produce quite similar SSA throughout, the Sphere model produces smaller SSA. In particular, the figure indicates that the SSA of an ensemble with $r_{\mathrm{v}} = 2\mu\mathrm{m}$ and $m_{\mathrm{I}} = 0.01$ at 355 nm, which corresponds to a size parameter of 35, will decrease by 0.04 if the spherical rather than non-spherical model is used. For the asymmetry factor, difference between the Sphere and Spheroid models reaches its maximum when the size parameter is around 7 and the difference starts to decrease after this value, while, to the contrary, the difference between the Spheroid and IH models starts to emerge and gradually increase for size parameter greater than 7.

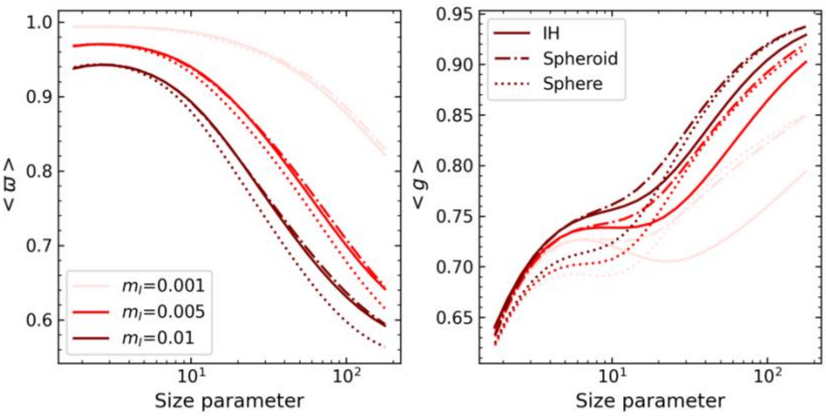


Figure 4.18. Bulk SSA, $\langle\varpi\rangle$, and asymmetry factor $\langle g\rangle$ against the size parameter, for $m_{\mathrm{R}} = 1.5$ but different values of m_{I}, simulated with the IH, Spheroid and Sphere models. A series of VSDs with the same geometric standard deviation as the VSD used in Figure 4.17 but different mode radii are used for the calculation.

Figure 4.19 shows α, β and δ of particle ensembles with different $r_{\mathrm{eff}}, m_{\mathrm{R}}$ and m_{I} at 532 nm simulated by the three scattering models. It can be seen that as r_{eff} grows, α firstly increases and then decreases. At the same time, sensitivity of α to m_{R} and particle shape gradually vanishes. For small particles, α of spheres is the highest probably because for the same volume-equivalent radius, a spherical particle has the largest projected area. On the other hand, the variation of m_{I} does not change α. The variation trend of β is the same as α. However, it shows stronger dependence on particle shape: the β of spherical particles is obviously stronger than that of non-spherical particles. Meanwhile, unlike α, there is obvious sensitivity of β to both m_{R} and m_{I}. The PLDR, δ, of spheroidal and IH particles present distinct and complex patterns. When r_{eff} is small, δ simulated with the Spheroid model and that simulated with the IH model have the same variation trend, but the latter is a little lower in magnitude. However, as r_{eff} grows, the δ curves from the two non-spherical models diverge and present disparate trends. It is worth figuring out that for small r_{eff}, δ shows more sensitivity to particle size than CRI, while the situation reverses as r_{eff} grows.

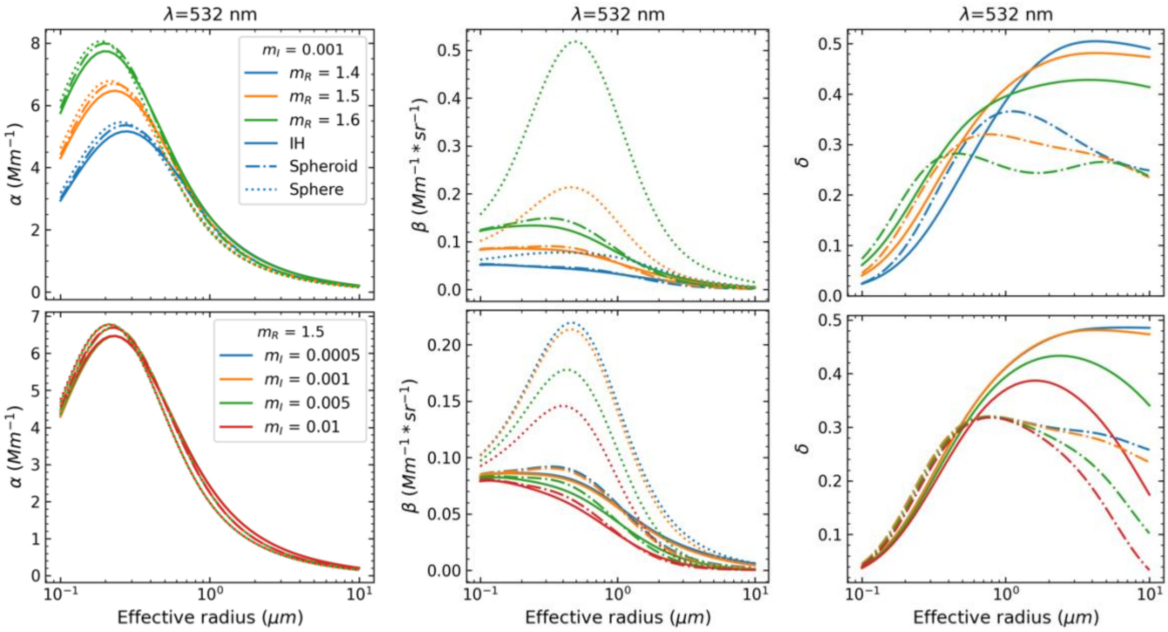


Figure 4.19. α, β and PLDR (δ) at 532 nm simulated with the IH, Spheroid and Sphere models for different m_{R}, m_{I} and r_{eff}. The VSDs used for the calculation are the same as those used in Figure 4.18.

From the above simulations one can see the scattering properties simulated with the IH, Spheroid and Sphere models are within comparable ranges. Compared to angular-integrated properties ($\langle\varpi\rangle, \langle g\rangle$and α), the angular-related properties ($\beta, \delta$and $\langle P_{ij}\rangle$) are more sensitive to the change of particle shape. For lidar-related properties, the main difference between the spherical and two non-spherical models lies in β, while the main difference between the two non-spherical models lies in δ. The sensitivity to CRI can be found in β and partly found in δ depending on the size parameter.

To further compare difference between the two non-spherical models when reproducing lidar-measured optical properties, Figure 4.20 displays the RMS differences in $(3\beta + 2\alpha)$ and in $(3\beta + 2\alpha + 3\delta)$ simulated by the Spheroid and IH models for different CRIs and effective radii. It can be seen that the difference between the two non-spherical models apparently enlarges after the incorporation of 3δ. For both $(3\beta + 2\alpha)$ and $(3\beta + 2\alpha + 3\delta)$ data, the modeling difference grows with the increase of r_{eff}. A sudden increase of the slope occurs at r_{eff}~1 μm for $(3\beta + 2\alpha)$ data. Whereas for $(3\beta + 2\alpha + 3\delta)$ data, the point where the leap of the slope happens and the magnitude depend on m_{R} and m_{I}, respectively. In particular, difference in $(3\beta + 2\alpha + 3\delta)$ of a particle ensemble with r_{eff} of 10 μm and m_{I} of 0.01 can reach as high as 100%.

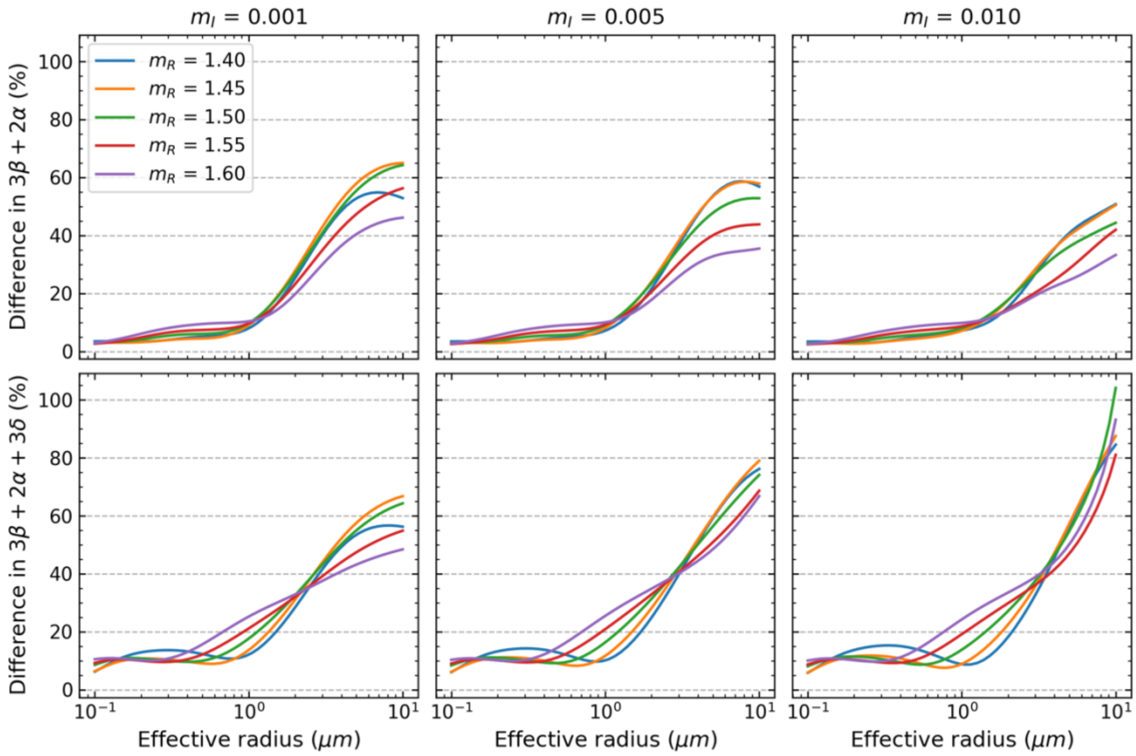


Figure 4.20. RMS differences in $(3\beta + 2\alpha)$ (upper row) and in $(3\beta + 2\alpha + 3\delta)$ (lower row) simulated by the Spheroid and IH models for different CRIs and effective radii. The VSDs used for the calculation are the same as those used in Figure 4.18

4.3.2 Retrieval results

The non-spherical aerosol models defined in Table 4.1 are retrieved using the retrieval configurations specified in Sect. 4.1.2. The following sub sections respectively present the results derived when the non-spherical scattering model is set to the IH model (Sect. 4.3.2.1) and Spheroid model (Sect. 4.3.2.2).

4.3.2.1 IH model

Figure 4.21 shows retrievals of VSD for the three dust VSD types (columns) and at different values of CRI (rows). Both the real and imaginary parts increase from the top row to the bottom row. The optical data are error-free. It can be seen that using the non-spherical model to invert the full $3\beta + 2\alpha + 3\delta$ synthetic measurements receives the best result. The influence of using the Sphere model and inverting $3\beta + 2\alpha$ data on retrieval accuracy differs for different VSD types and CRIs. Except for $(m_{\mathrm{R}}^* = 1.6, m_{\mathrm{I}}^* = 0.009)$, the Sphere model underestimates V_{t}, overestimates σ_{g} and causes apparent bias of r_{v}. From Figure 4.19 we infer that the underestimation of V_t by the Sphere model is a reaction to the enhancement of backscattering due to the spherical assumption. Moreover, in the case where $(m_R^* = 1.4, m_I^* = 0.001)$, severe underestimation of V_t happens no matter which scattering model is used, as long as only $3\beta + 2\alpha$ data are inverted. As seen in the scenario of retrieving spherical particles, this is owed to the underdetermination of the $3\beta + 2\alpha$ system and accompanied by noticeable overestimation of m_R and m_I, indicating severe cross-talks between the state variables. When the IH model is used as the forward model in BOREAL, the mode radius and geometry standard deviation are easier to be underestimated in the absence of 3δ. It is worth mentioning that for the case $(m_R^* = 1.4, m_I^* = 0.001)$, the incorporation of 3δ greatly improves the retrieval accuracy, particularly for the monomodal VSD types (TD and FD). This suggests the sensitivity of depolarization measurements to large particles, which has been already proved from aspects of theoretical calculation (Sassen, 2000) and retrieval application (Dubovik et al., 2006; Gasteiger and Freudenthaler, 2014).

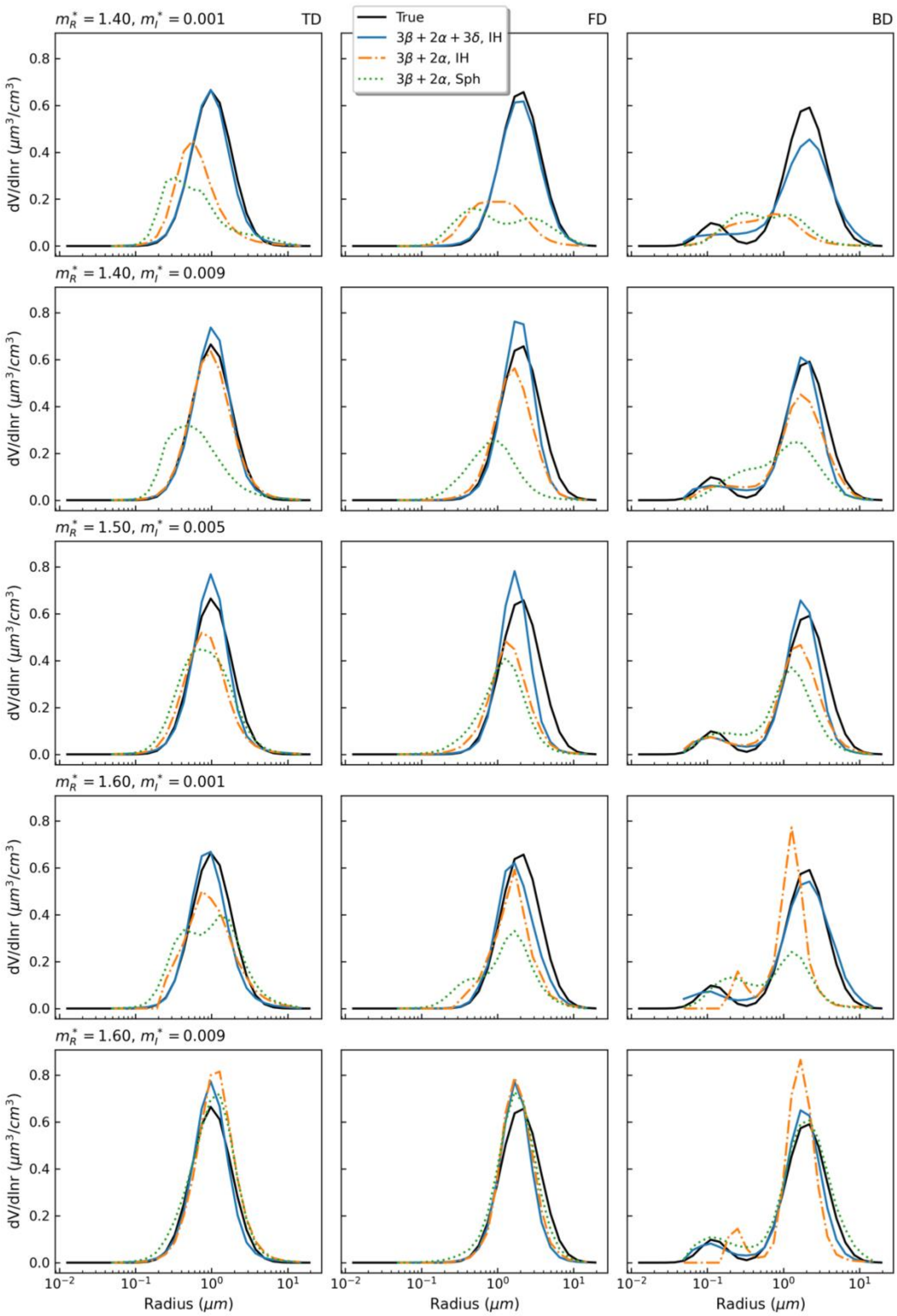


Figure 4.21. True VSDs (black solid lines) of the dust types in Table 4.1 and the retrievals that are derived from the configuration ($3\beta + 2\alpha + 3\delta$, IH) (blue solid lines), ($3\beta + 2\alpha$, IH) (orange dash-dot lines), and ($3\beta + 2\alpha$, Sphere) (green dot lines), under the error-free condition. The rows correspond to different true CRIs and the columns correspond to the three VSD types: transported dust (TD), fresh dust (FD) and bimodal dust (BD).

To further evaluate the retrieval accuracy of non-spherical particles, we examine, under the error-free condition, the retrieval error, $\varepsilon(V_{\mathrm{t}})$, $\varepsilon(r_{\mathrm{eff}})$, $\varepsilon(m_{\mathrm{R}})$, $\varepsilon(m_{\mathrm{I}})$, $\varepsilon(\varpi)_{\mathrm{RMS}}$, and the fitting error, $\varepsilon_{\mathrm{fit}}$. As shown in Figure 4.22, the improvement of retrieval accuracy once the 3δ data are included is clearly seen, while other features are also worth mentioning. Firstly, in the absence of depolarization measurements, both Sphere and IH models underestimate the size ($V_{\mathrm{t}}, r_{\mathrm{eff}}$) and overestimate the CRI of the ensemble when m_{R}^{*} and m_{I}^{*} are low. This trend is consistent with the retrievals of coarse-mode spherical particles presented in Sect. 4.2. Secondly, the use of Sphere model leads to large enhancement of backscattering, as shown in Figure 4.19, which in turn influences retrieval accuracy in two ways. Specifically, to cancel the enhancement, (1) VSD and m_{I} can be retrieved in reasonable ranges while m_{R} is significantly underestimated; or (2) m_{R} can be retrieved in a reasonable range while the particle size ($V_{\mathrm{t}}, r_{\mathrm{eff}}$) is largely underestimated and m_{I} is severely overestimated. The former circumstance happens in the cases where $m_{\mathrm{R}}^{*} > 1.4$ and is also reported by Veselovskii et al. (2010). The latter occurs at $m_{\mathrm{R}}^{*} = 1.4$ and might cause more serious outcomes: it would mislead one into considering the retrieved ensemble is composed of more absorbing particles with lower SSA – the optical property more determined by m_{I} than by m_{R}. Finally, Figure 4.22 shows some abnormal results derived with the retrieval configuration ($3\beta + 2\alpha$, IH) at $m_{\mathrm{R}}^{*} = 1.6$. In these cases, each retrieval can only find one individual solution located in the Substitute space (Sect. 3.5) and as a result, the fitting errors soar over 8% compared to other cases. The retrieved VSDs shown in Figure 4.21 suggest the abnormal overestimates of r_{eff} in these cases could be due to the underestimates of modal geometry standard deviation. We found increasing $c_{\mathbf{Y}_2}$ (see Sect. 3.4.2.1) could improve the results for these cases to some extent.

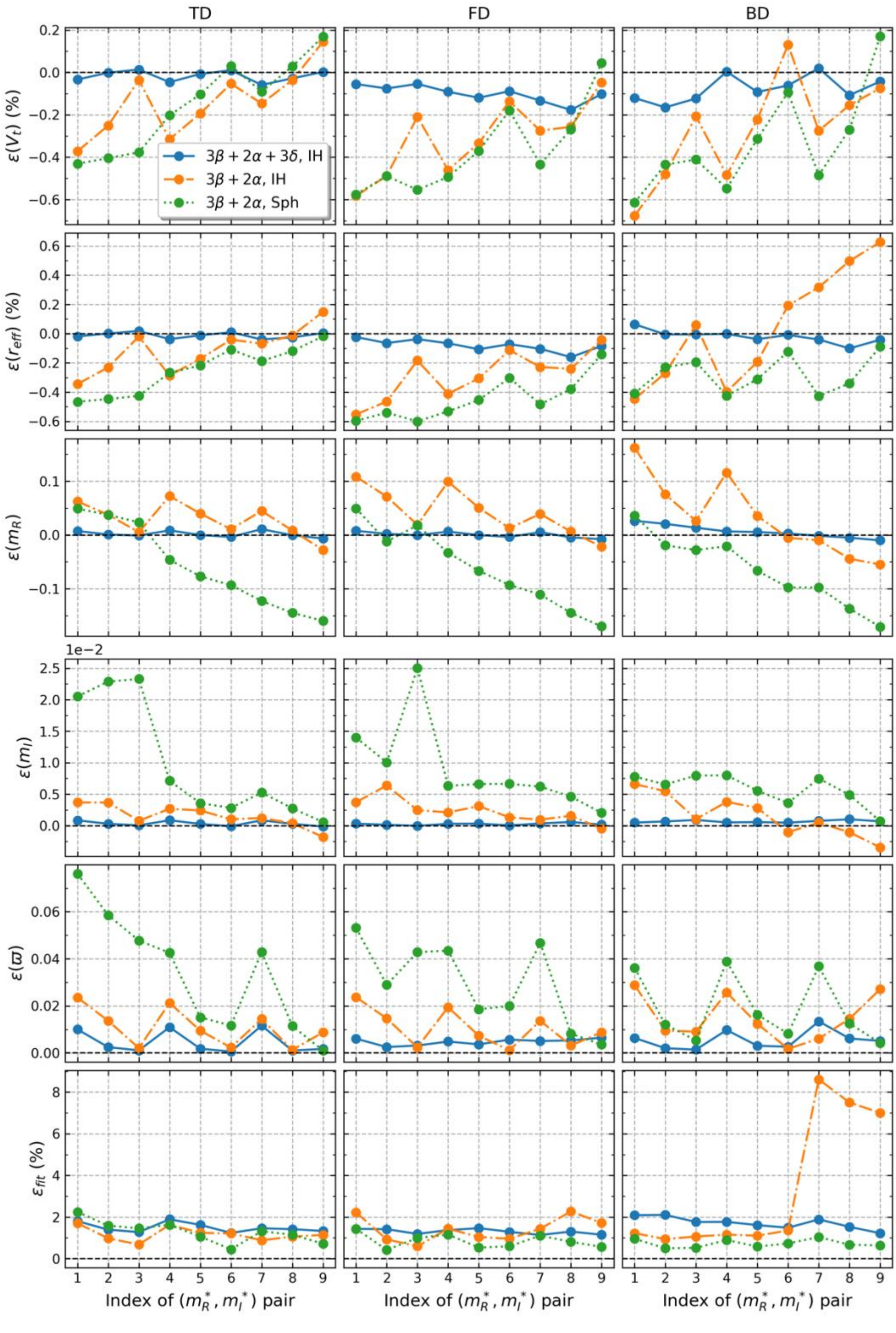


Figure 4.22. Distributions of $\varepsilon(V_\mathrm{t})$, $\varepsilon(r_\mathrm{eff})$, $\varepsilon(m_\mathrm{R})$, $\varepsilon(m_\mathrm{I})$, $\varepsilon(\varpi)_\mathrm{RMS}$, and the fitting error, ε_fit, derived from ($3\beta + 2\alpha + 3\delta$, IH) (blue solid), ($3\beta + 2\alpha$, IH) (orange dash-dot), ($3\beta + 2\alpha$, Sphere) (green dot), under the error-free condition, with respect of the true CRIs. The x-axis in each panel suggests the index of the true CRI values, as listed in Table 4.9.

Table 4.9. Index of the true CRI $(m_{\mathrm{R}}^{*}, m_{\mathrm{I}}^{*})$ pairs as the x-axis of each panel in Figure 4.22.

Index	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
$(m_{\mathrm{R}}^{*}, m_{\mathrm{I}}^{*})$	(1.4, 0.001)	(1.4, 0.005)	(1.4, 0.009)	(1.5, 0.001)	(1.5, 0.005)	(1.5, 0.009)	(1.6, 0.001)	(1.6, 0.005)	(1.6, 0.009)

To further evaluate the performance of different retrieval configurations on retrieving the defined non-spherical particles, we calculated the magnitudes of mean and standard deviation of the retrieval and fitting errors for each VSD Type, and exploit the summation of the two quantities as a measure of type-resolved accuracy. The calculation results for both error-free (EF) and error-contaminated (EC) conditions are listed in Table 4.10. In the paratheses, values on the left represent the mean and those on the right represent the standard deviation; the summation out of the paratheses measures the overall accuracy. For the error-free cases, retrieval accuracy deteriorates with the increases of the mode radius and number of modes. The incorporation of 3δ reduces retrieval errors by a factor of 2 to 14 with regard to different retrieval quantities and VSD types. Compared to inverting $3\beta + 2\alpha$ with the IH model, results derived with the Sphere model have worse accuracy due to errors in forward modeling. On the other hand, regardless retrieval configuration, fitting errors for the error-free cases are much smaller than the set measurement standard deviation (4.2% for $3\beta + 2\alpha$, 4.5% for $3\beta + 2\alpha + 3\delta$), other than the BD type retrieved with the configuration ($3\beta + 2\alpha$, IH) which has been discussed above as a special case.

Comparing the results derived under the error-contaminated condition with those under the error-free condition, one can see measurement noise causes the largest relative reduction of the retrieval accuracy for the ($3\beta + 2\alpha + 3\delta$, IH) configuration. For the relative retrieval errors like $\varepsilon(V_{\mathrm{t}})$ and $\varepsilon(r_{\mathrm{eff}})$, the increase in magnitude is comparable to the standard deviation of the measurement noise. In most cases, the decline of statistical accuracy due to the introduction of measurement noise is mainly reflected by the rise of data dispersion (increase of the standard deviation). Nonetheless, in some cases, the mean values also apparently change under the noisy condition, compensating the increase of standard deviation to some extent. Compared to the error-free cases, the fitting error, $\varepsilon_{\mathrm{fit}}$, rises to a magnitude comparable to the standard deviation of the noise, as well as the standard deviation set in BOREAL.

Table 4.10. Same as Table 4.5, but for non-spherical particles retrieved with different retrieval configurations.

		Rtv. Conf.	$\varepsilon(V_{\mathrm{t}})(\%)$	$\varepsilon(r_{\mathrm{eff}})(\%)$	$\varepsilon(m_{\mathrm{R}}) \times 10^{-3}$	$\varepsilon(m_{\mathrm{I}}) \times 10^{-4}$	$\varepsilon(\varpi)_{\mathrm{RMS}} \times 10^{-3}$	$\varepsilon_{\mathrm{fit}}(\%)$
Error-free	**TD**	**3+2+3**	4 (2+2)	3 (1+2)	8 (2+6)	8 (4+4)	9 (5+4)	1 (1+0)
		3+2	29 (14+15)	26 (11+15)	58 (28+30)	33 (16+17)	19 (11+8)	1 (1+0)
		Sph	35 (15+20)	40 (25+15)	134 (59+75)	188 (99+89)	58 (34+24)	2 (1+1)
	FD	**3+2+3**	14 (10+4)	12 (8+4)	6 (1+5)	5 (3+2)	6 (5+1)	1 (1+0)
		3+2	47 (31+16)	44 (28+16)	84 (43+41)	42 (24+18)	17 (10+7)	2 (1+1)
		Sph	56 (37+19)	59 (45+14)	133 (63+70)	156 (91+65)	46 (29+17)	1 (1+0)
	BD	**3+2+3**	14 (8+6)	6 (2+4)	18 (7+11)	9 (7+2)	10 (6+4)	2 (2+0)
		3+2	50 (27+23)	41 (4+37)	102 (33+69)	47 (16+31)	24 (15+9)	6 (3+3)
		Sph	56 (33+23)	40 (28+12)	129 (67+62)	81 (58+23)	32 (19+13)	1 (1+0)
Error-contaminated	**TD**	**3+2+3**	10 (1+9)	12 (3+9)	19 (2+17)	13 (2+11)	15 (9+6)	4 (3+1)
		3+2	32 (12+20)	29 (9+20)	63 (27+36)	37 (15+22)	22 (13+9)	3 (2+1)
		Sph	37 (16+21)	42 (26+16)	133 (61+72)	177 (92+85)	57 (34+23)	3 (2+1)
	FD	**3+2+3**	19 (3+16)	16 (0+16)	16 (2+14)	9 (0+9)	12 (7+5)	4 (3+1)
		3+2	50 (31+19)	47 (28+19)	92 (46+46)	68 (30+38)	24 (14+10)	3 (2+1)
		Sph	60 (39+21)	62 (46+16)	132 (58+74)	185 (101+84)	53 (34+19)	2 (1+1)
	BD	**3+2+3**	22 (3+19)	27 (6+21)	36 (11+25)	20 (6+14)	20 (12+8)	4 (3+1)
		3+2	59 (34+25)	41 (17+24)	126 (63+63)	66 (36+30)	28 (18+10)	4 (2+2)
		Sph	58 (36+22)	45 (30+15)	124 (61+63)	110 (70+40)	41 (27+14)	1 (1+0)

4.3.2.2 Spheroid model

The results for the Spheroid cases are shown in Figure 4.23 and Figure 4.24 and Table 4.11. These results are similar to those in the IH cases so that similar conclusions can be drawn. In spite of this, it should be noted that compared with the IH model, the accuracy declines when using the Spheroid model to retrieve FD and BD types from the full $3\beta + 2\alpha + 3\delta$ data, which is probably due to the fact that the 3δ calculated by the two models have different sensitivity to the retrieval parameters.

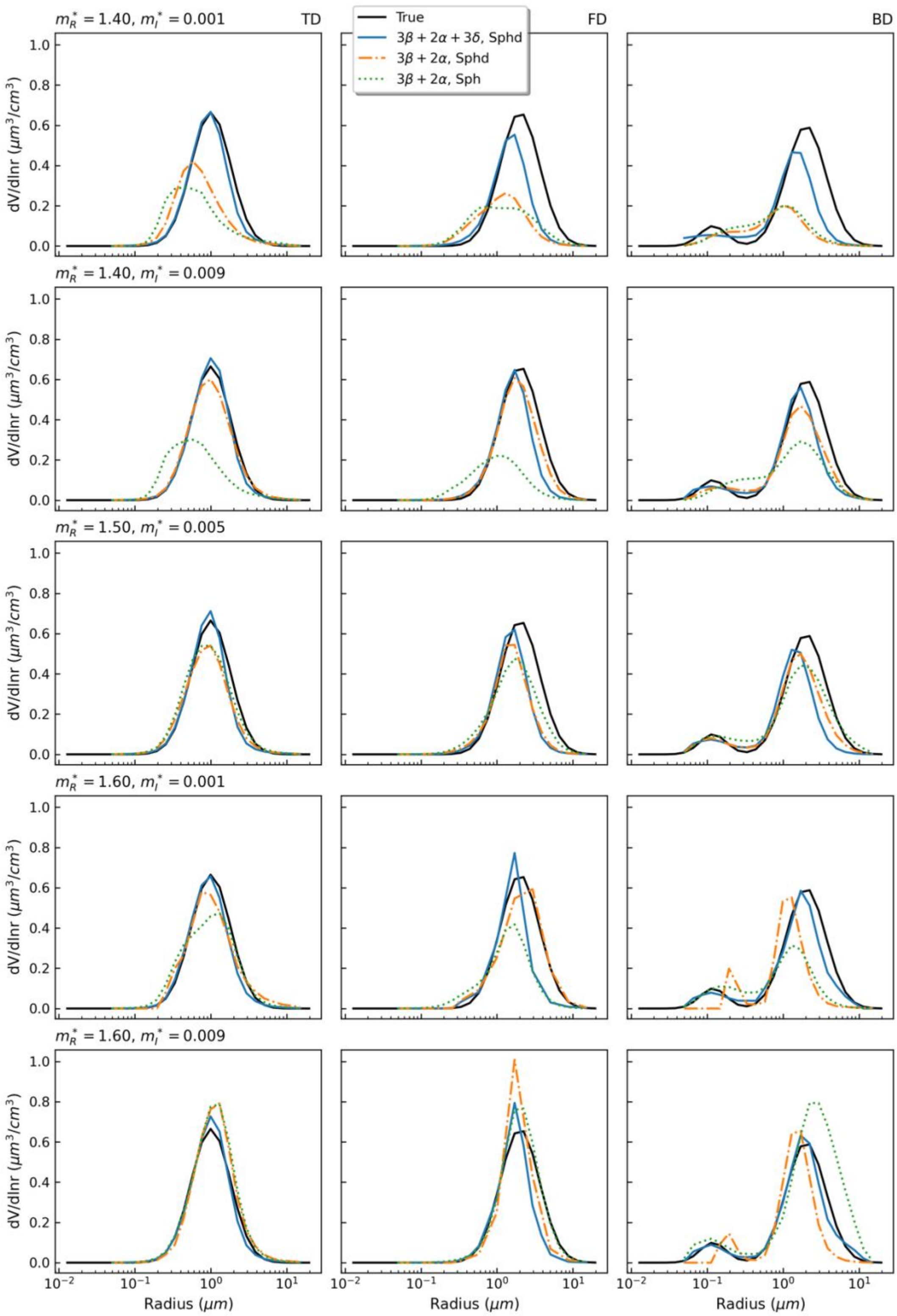


Figure 4.23. Same as Figure 4.21, but the Spheroid model is used to generate the $3\beta + 2\alpha + 3\delta$ data and serves as the forward non-spherical model in the retrieval process.

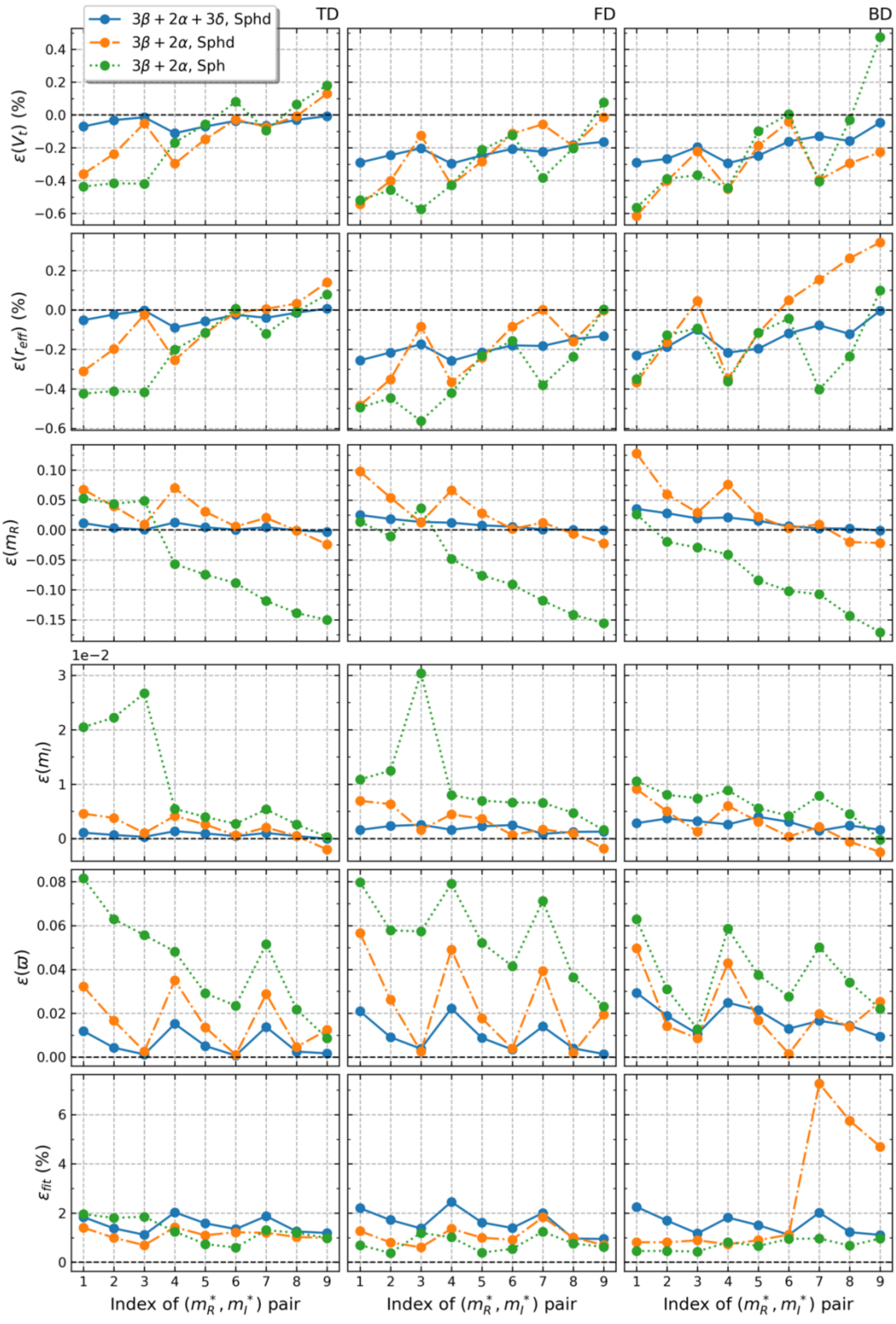


Figure 4.24. Same as Figure 4.22, but the Spheroid model is used to generate the $3\beta + 2\alpha + 3\delta$ data and serves as the forward non-spherical model in the retrieval process.

Table 4.11. Same as Table 4.10, but the Spheroid model is used to generate the $3\beta + 2\alpha + 3\delta$ data and serves as the forward non-spherical model in the retrieval process.

		Rtv. Conf.	$\varepsilon(V_t)(\%)$	$\varepsilon(r_{eff})(\%)$	$\varepsilon(m_R) \times 10^{-3}$	$\varepsilon(m_I) \times 10^{-4}$	$\varepsilon(\varpi)_{RMS} \times 10^{-3}$	$\varepsilon_{fit}(\%)$
Error-free	**TD**	**3+2+3**	8 (5+3)	6 (3+3)	9 (4+5)	11 (7+4)	11 (6+5)	2 (2+0)
		3+2	27 (12+15)	22 (8+14)	53 (24+29)	39 (19+20)	28 (16+12)	1 (1+0)
		Sph	36 (14+22)	37 (18+19)	131 (54+77)	195 (100+95)	65 (43+22)	1 (1+0)
	FD	**3+2+3**	27 (23+4)	24 (20+4)	17 (9+8)	24 (18+6)	17 (10+7)	2 (2+0)
		3+2	41 (24+17)	36 (20+16)	63 (27+36)	54 (27+27)	43 (24+19)	1 (1+0)
		Sph	51 (31+20)	50 (33+17)	131 (66+65)	177 (98+79)	73 (55+18)	1 (1+0)
	BD	**3+2+3**	28 (20+8)	21 (14+7)	26 (14+12)	36 (28+8)	24 (18+6)	2 (2+0)
		3+2	48 (32+16)	26 (2+24)	78 (32+46)	61 (27+34)	36 (21+15)	5 (3+2)
		Sph	50 (20+30)	34 (18+16)	135 (75+60)	93 (63+30)	53 (37+16)	1 (1+0)
Error-contaminated	**TD**	**3+2+3**	12 (1+11)	11 (1+10)	17 (3+14)	18 (4+14)	18 (11+7)	4 (3+1)
		3+2	29 (11+18)	26 (8+18)	59 (26+33)	47 (21+26)	31 (18+13)	3 (2+1)
		Sph	37 (15+22)	38 (19+19)	127 (53+74)	187 (97+90)	66 (43+23)	3 (2+1)
	FD	**3+2+3**	30 (18+12)	26 (14+12)	21 (7+14)	32 (14+18)	23 (13+10)	4 (3+1)
		3+2	44 (26+18)	39 (22+17)	66 (29+37)	61 (32+29)	47 (27+20)	3 (2+1)
		Sph	54 (33+21)	52 (34+18)	130 (62+68)	197 (108+89)	79 (58+21)	2 (1+1)
	BD	**3+2+3**	32 (17+15)	26 (10+16)	37 (16+21)	55 (28+27)	35 (20+15)	6 (3+3)
		3+2	52 (30+22)	34 (13+21)	94 (44+50)	74 (42+32)	40 (23+17)	3 (1+2)
		Sph	52 (27+25)	41 (22+19)	126 (66+60)	124 (79+45)	59 (41+18)	1 (1+0)

4.4 Chapter summary

This chapter evaluates retrieval performance of BOREAL via simulations. Retrieval accuracy is assessed by comparing the inversion results of synthetic measurements with the realistic aerosol models (microphysical properties) that generate the synthetic measurements. When retrieving spherical particles, the Sphere model is used as the forward model and $3\beta + 2\alpha$ measurements are inverted. Simulation results show that the best retrieval accuracy is achieved if the aerosol ensemble is only composed of fine-mode particles. One of main factors that influence retrieval stability turns out to be the cross-talk between state variables which stems from both lack of sensitivity and underdetermination of the retrieval system. Therefore, it is of great importance to improve the accuracy of the *a priori* constraints, particularly on the imaginary part of CRI. On the other hand, the algorithm shows good performance on resisting measurement noise. The reduction of retrieval accuracy caused by measurement noise is less than or comparable with the standard deviation of the introduced noise. An estimate of retrieval accuracy and fitting error for retrieving spherical particles from $3\beta + 2\alpha$ data under both error-free and error-contaminated conditions can be found in Table 4.5.

Next, to evaluate BOREAL's capability of retrieving non-spherical dust particles, results derived from using different forward models (Sphere, Spheroid and IH models) and types of input measurements ($3\beta + 2\alpha$, $3\beta + 2\alpha + 3\delta$) are compared and discussed. It adequately shows that the incorporation of 3δ can thoroughly improve retrieval accuracy. In the absence of 3δ measurements, using the correct non-spherical scattering model as the forward model faces issues similar with those in the spherical particle retrieval, while using the wrong Sphere model usually leads to two outcomes depending on the true CRI: loss of (1) accuracy of the real part (underestimated) or (2) accuracy of both volume concentration (underestimated) and the imaginary part (overestimated). Summaries of retrieval accuracy for non-spherical particles under error-free and error-contaminated conditions can be found in Table 4.10 and Table 4.11.

In the next chapter, we will present results of retrieving real aerosol events detected by LILAS. Compared to the synthetic measurements, a distinct point of inverting real measurements is that the error of the forward model is included. Accordingly, apart from evaluating retrieval accuracy for real aerosols, applying BOREAL with different scattering models included to the real measurements also provides us with a tool to evaluate the capability of the scattering models to reproduce the real lidar data.

5 Application of BOREAL to real lidar data

In this chapter, BOREAL is employed to retrieve aerosol properties of various atmospheric events detected by LILAS, including 4 cases of biomass burning aerosol (BBA), 2 cases of dust aerosol (DA) and 1 case of continental polluted aerosol. The BBA particles were emitted by wildfires in North America and then transported towards western Europe. During the transport complex ageing processes of BBA particles might significantly alter their optical and microphysical properties. The DA cases consist of one fresh dust measured near the source during the Dust Aerosol Observation (DAO) campaign (Hu et al., 2020) and another transported dust event observed at ATOLL. For the DA cases, results derived from different scattering models (i.e., the Sphere, Spheroid and IH models) and measurements (i.e., conventional ($3\beta + 2\alpha$) and ($3\beta + 2\alpha + 3\delta$)) are compared. The continental aerosol case happened in the springtime of Lille when heavy aerosol masses were concentrated in the boundary layer (BL) and the atmospheric condition above the BL is clear.

5.1 Transported biomass burning aerosols

Biomass burning is one of the main contributors to atmospheric pollutants like black carbon (BC) and primary organic aerosol (POA), trace gases and greenhouse gases. Biomass burning aerosols (BBA) have profound influence on regional and global radiation balance, cloud formation and precipitation, as well as human health (Andreae, 2019). Apart from anthropogenic sources such as agricultural burning and indoor biofuel use, open vegetation fire (wildfire) is becoming a more and more important natural source of BBA in recent years, with increases in both frequency and intensity (Schoennagel et al., 2017). Fresh BBA emitted from wildfire are composed of fine aggregates of BC cores coated with organic carbon (OC) condensations and thus are usually more absorbing compared to other aerosol types (China et al., 2013). However, during long-range transport, complicated ageing processes can happen, altering the size, shape, inner composition and as a result, the optical properties of the BBA particles.

5.1.1 Overview of the selected BBA cases

During routine observations at the ATOLL/Lille platform, aged BBA plumes originated from different sources were detected by LILAS. Here we select four aged BBA episodes to retrieve. Among them, Case 1 and Case 2 are two distinct BBA plumes from 2023 Canadian wildfire and simultaneously passing over Lille; Case 3 and Case 4 are BBAs generated by Creek fires and Oregon fires respectively, subject to 2020 Californian Wildfire. Detailed information about the four selected cases is shown in Table 5.1.

Because ATOLL/Lille is also an AERONET station, we present AERONET retrievals closest to the cases for information. To ensure the quality of the selected AERONET retrievals, the following criteria are applied:

(1) if available, adopt level 2 retrievals in priority;

(2) otherwise, adopt level 1.5 retrievals with sky residual < 5%, solar zenith angle > 50° and AOD_{440} > 0.15;

(3) from the retrievals selected by (1) or (2), pick the one closest to the lidar measurement.

The AERONET retrievals determined by the above criteria for each case, together with the columnar AOD_{440}, are shown in Table 5.1 as well.

Table 5.1. Information about the selected BBA cases.

	Case 1	**Case 2**	**Case 3**	**Case 4**
Obser. site	ATOLL/Lille	ATOLL/Lille	ATOLL/Lille	ATOLL/Lille
Time period	20:30-22:30 UTC, 27 May 2023	20:30-22:30 UTC, 27 May 2023	22:00 UTC 12-03:00 UTC 13, Sep. 2020	23:00 UTC 17-03:00 UTC 18, Sep. 2020
Source	West Canada	Northwestern US	California, US	Oregon, US
Ageing	5 days	5 days	4 days	6 days
Layer	3.5-5 km	12-12.3 km	5-6.5 km	7-8 km
Layer AOD_{532}	0.065	0.009	0.157	0.083
AERONET retrieval	17:58 UTC, 27 May 2023		13:55 UTC, 11 Sep. 2023	7:07 UTC, 18 Sep. 2020
Columnar AOD_{440}	0.261		0.853	0.182
Reference	Hu et al. (scientific report, 2023)		Hu et al. (2022)	

About 45 days of smoke plumes ranging from the troposphere to lower stratosphere were detected by LILAS since 14 May 2023. Back trajectory analysis suggests these particles come from fire emission in Canada. Figure 5.1 shows a one-week time series of the backscattered lidar signal at 1064 nm, where an almost continuous layer separated from the BL can be identified. Aerosol classification based on fluorescence capacity, PLDR and relative humidity suggests it has high concentration of BBA (Hu et al., *scientific report*, 2023). Here we focus on the measurement taken between 20:30-22:30 UTC, 27 May 2023. The time-averaged vertical profiles of aerosol optical properties are shown in Figure 5.2, from which two distinct layers, one from 3 to 5.5 km (Case 1) and another from 11.5 to 12.5 km (Case 2), are able to be identified. Figure 5.3 is a combination between the back trajectory of the lower layer (4 km) and the higher layer (12 km) and the UVAI measured by the ozone mapping and profiler suite (OMPS) satellite, from which different transport paths of the aerosol masses can be identified. The UVAI mapping indicates the burning source of the higher layer (i.e., northeastern coast of the US), while cannot trace the burning source of the lower layer.

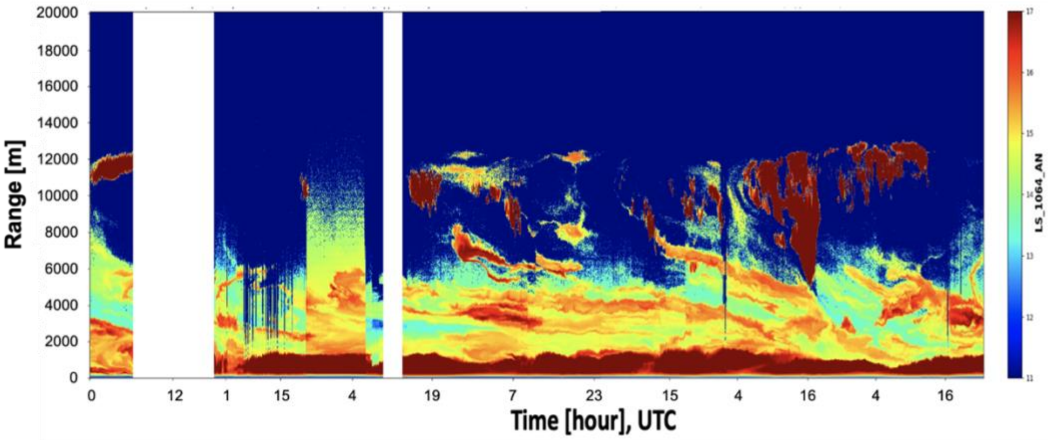


Figure 5.1. One week (from 24 to 30 May 2023) time series of LILAS range-corrected backscattered signal at 1064 nm, at ATOLL/Lille. (cited from Hu et al. (*scientific report, 2023*), AUSTRAL processing)

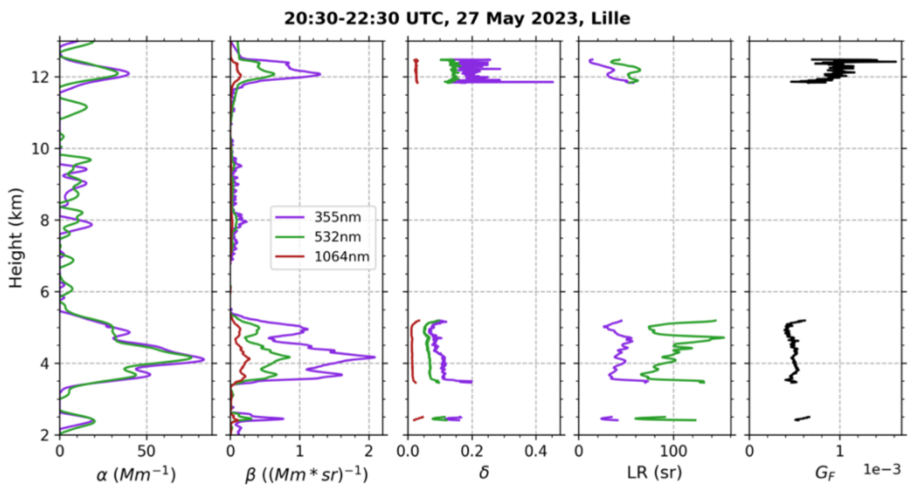


Figure 5.2. Optical profiles of extinction coefficient (α), backscattering coefficient (β), PLDR (δ), lidar ratio (LR) and fluorescence capacity ($G_{\mathrm{F}} = \beta_{\mathrm{F}}/\beta_{532}$), averaged between 20:30 UTC and 23:30 UTC, 27 May 2023, at ATOLL/Lille.

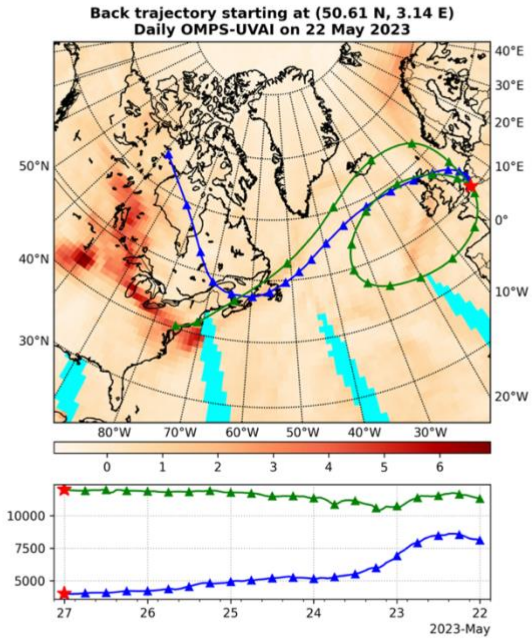


Figure 5.3. 5-day backward trajectory starting on 27 May 2023 at two heights: 4 km represented by the blue curve and 12 km represented by the green curve, over ATOLL/Lille (start marker), together with the UVAI measured by the ozone mapping and profiler suite (OMPS) satellite on 22 May 2023.

The 2020 California Wildfire started from the beginning of May and lasted until the end of that year, making it the most extensive wildfire event in Californian modern history (https://www.fire.ca.gov/incidents/2020/, last access: October 16, 2023). In September, large amounts of aerosol plumes were continuously detected by LILAS (Figure 5.4). We focus on two representative time intervals to retrieve: Case 3 between 22:00 UTC 12 and 03:00 UTC 13, and Case 4 between 23:00 UTC 17 and 03:00 UTC 18. The time-averaged vertical profiles of aerosol optical properties corresponding to the two cases are shown in Figure 5.5 and Figure 5.6, respectively. Back trajectory analysis to the selected layers (5-6.5 km in Case 3 and 7-8 km in Case 4) confirms they are aged BBAs respectively originated from the fires in Creek, Southwestern California and Oregon, the US (Hu et al., 2022).

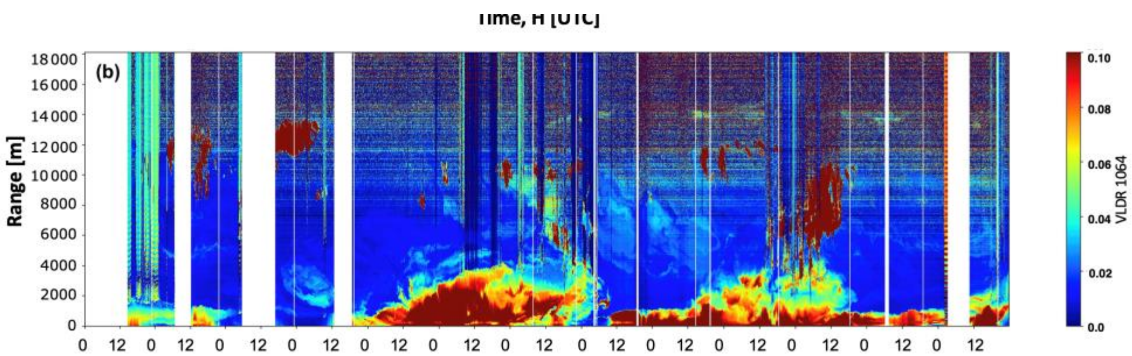


Figure 5.4. LILAS range-corrected backscattered signal at 1064 nm during the smoke episode on 10-22 September 2020, at ATOLL/Lille (cited from Hu et al. (2022), AUSTRAL processing).

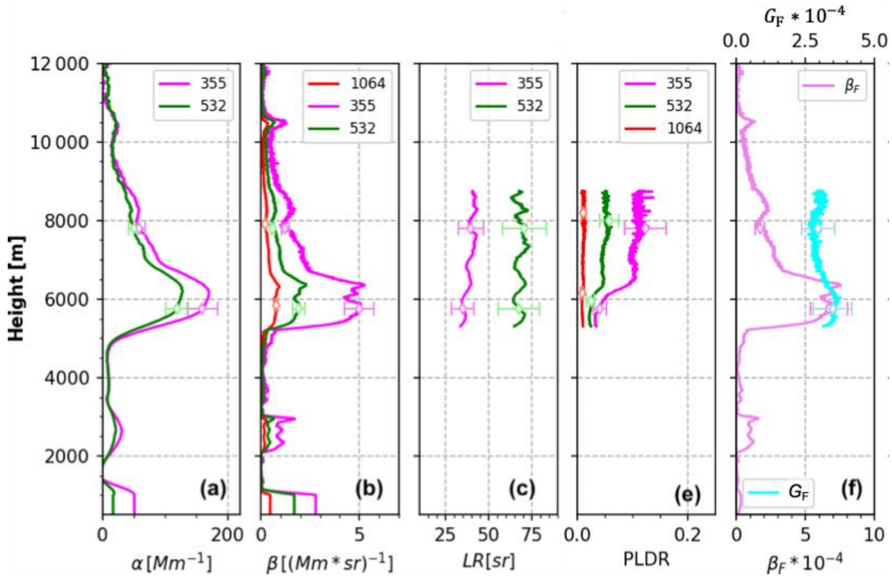


Figure 5.5. Optical profiles of *α, β, LR, β_F* and G_F (β_F/β_{532}), averaged between 22:00 UTC 11 September and 03:00 UTC 12 September, 2020, at ATOLL/Lille (Case 3) (adapted from Hu et al. (2022), AUSTRAL processing).

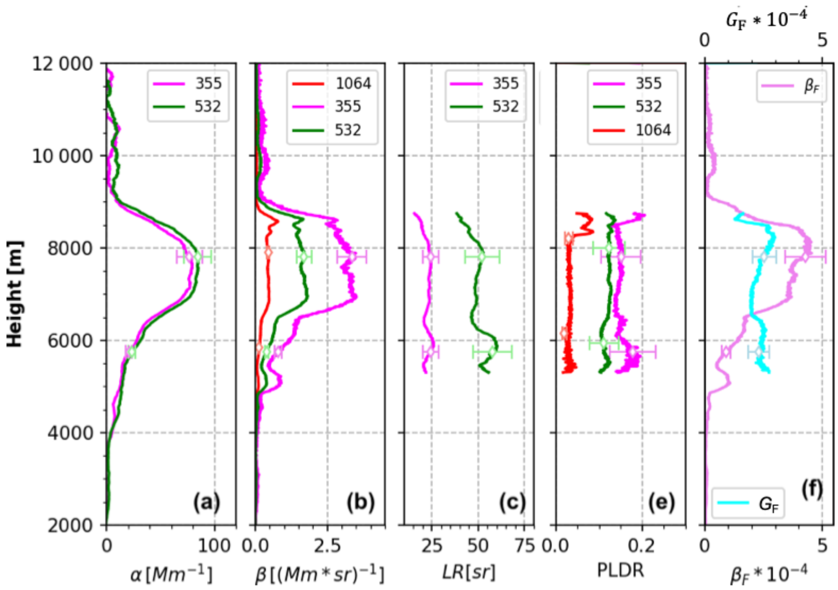


Figure 5.6. Same as Figure 5.5, but the average time is between 22:30 UTC 17 September and 03:00 UTC 18 September, 2020 (Case 4) (adapted from Hu et al. (2022), AUSTRAL processing).

5.1.2 Results and discussion

Figure 5.7 shows the retrieval of VSD and SSA for the selected cases. Other retrieved microphysical properties are listed in Table 5.2. The *a priori* constraint on CRI is set to the "Absorbing" type (Table 3.1) and the Sphere model is used as the forward model of BOREAL. Considering the non-negligible PLDR at 355 and 532 nm for the Californian Wildfire cases, the non-spherical models are also applied to these cases and the relevant results will be presented at the end of this sub section.

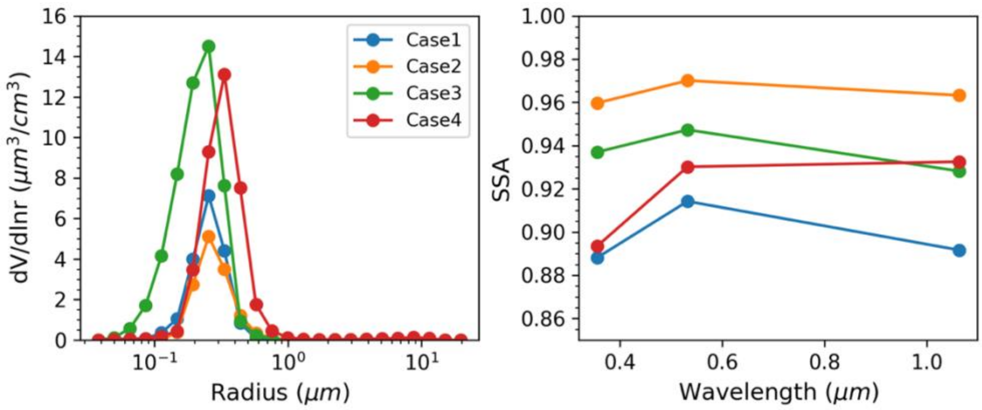


Figure 5.7. Retrieval of VSD (left) and SSA (right) for the selected cases. The *a priori* constraint on CRI is set to the "Absorbing" type (Table 3.1) and the Sphere model is used as the BOREAL forward model. Refer to Table 4.8 for retrieval accuracy.

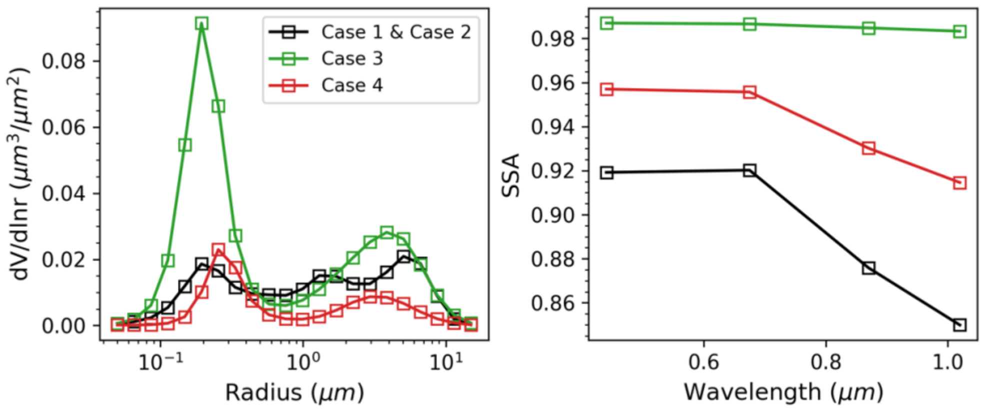


Figure 5.8. AERONET-retrieved VSD (left) and SSA (right) for the measurement times listed in Table 5.1. Also refer to Table 5.1 for the corresponding AOD.

Table 5.2. Fitting error, as well as other microphysical properties retrieved with BOREAL for the selected BBA cases. AERONET-retrieved fine-mode effective radius, $r_{\mathrm{eff,f}}$, $\overline{m_{\mathrm{R}}}$ and $\overline{m_{\mathrm{I}}}$ (spectrally averaged values) are also listed.

	BOREAL					**AERONET**		
	$\varepsilon_{\mathrm{fit}}$(%)	$V_{\mathrm{t}}(\mu\mathrm{m}^3/\mathrm{cm}^3)$	$r_{\mathrm{eff}}(\mu\mathrm{m})$	m_{R}	m_{I}	$r_{\mathrm{eff,f}}$	$\overline{m_{\mathrm{R}}}$	$\overline{m_{\mathrm{I}}}$
Case 1	12	4.89	0.24	1.54	0.019	0.21	1.53	0.013
Case 2	14	3.57	0.27	1.51	0.007			
Case 3	5	13.86	0.19	1.57	0.011	0.19	1.56	0.002
Case 4	8	9.68	0.31	1.53	0.015	0.27	1.54	0.009

The VSD retrieved in each case is composed of a mono-fine mode. The retrieved values of r_{eff} are within the typical range of transported BBA which is generally larger than fresh particles (Reid et al., 2005). The values of r_{eff} indicate a correlation between the particle size and ageing time: the longest transported BBA from the Oregon fire (Case 4) has the largest r_{eff} of 0.31 μm, while the shortest transported BBA from the Creek fire (Case 3) has the smallest r_{eff} of 0.19 μm. The BBAs from the

Canadian fire present intermedium values. It has been well-proved that the size of BBA particles grows with the ageing process. According to the study by Reid et al. (2005) (as well as the references therein), freshly emitted BBA is initially formed by BC aggregates with small size, which serve as condensation nuclei and the main absorbing component in BBA. In general, the size of BBA increases for the first time after a few hours of the generation as the organic carbon (OC), which is less absorbing than BC, condenses upon the BC core. During the ageing process, the size can keep increase as a result of coagulation of OC. Meanwhile, with the accumulation of the OC coating, BBA particles become less absorbing (increase of SSA and decrease of m_{I}) (Abel et al., 2003). The retrieved SSA varies between 0.91 and 0.97 at 532 nm, well consistent with the range for aged BBA and higher than that for fresh BBA. Note that in the Canadian fire case, the lower layer (Case 1) and upper layer (Case 2) show quite different SSAs, resulting from the lower m_{I} retrieved for the upper layer. This might be explained by the difference of the fire source, which is further associated with the vegetation type, combustion process and ambient temperature. It has been demonstrated that the BBA generated from the flaming process which often occurs in grass-type burning and produces higher contents of BC are more absorbing, compared to the BBA generated from the smoldering process which is often the case for forest-type burning with lower temperature and produces higher contents of OC (Reid et al., 2005). Different transport paths and altitudes of the two layers, which are related to particle ambient conditions such as temperature, pressure and relative humidity (RH), could be another factor affecting the chemical composition of the BBA.

The AERONET retrievals (VSD and SSA) in Table 5.1 for the cases are shown in Figure 5.8. The AERONET-retrieved fine-mode effective radius, $r_{\mathrm{eff,f}}$, $\overline{m_{\mathrm{R}}}$ and $\overline{m_{\mathrm{I}}}$ (spectrally averaged values) are listed in Table 5.2. Note that they are shown here for reference only rather than comparison with the BOREAL retrievals. Differences in measurement time and vertical resolution between the sun-sky photometer and lidar make their results less comparable. However, for BOREAL and AERONET retrievals, the variations of r_{eff} $\left(r_{\mathrm{eff,f}}\right)$, m_{R} ($\overline{m_{\mathrm{R}}}$) and m_{I} ($\overline{m_{\mathrm{I}}}$) from Case 1 to 4 have consistent trends. In particular, the BOREAL-retrieved effective radii show good consistency with the AERONTE-retrieved fine-mode effective radii (although the r_{eff} for Case 2 is apparently larger, it has limited impact on the columnar property due to the low concentration).

For further comparison, Table 5.3 lists retrievals of BBA microphysical properties reported by other literatures. It can be seen the ranges of m_{R}, m_{I} and ϖ_{532} derived from different studies are quite variable, related to processes of combustion, ageing and transport. For example, in the measurements of the long-lasting stagnant haze over sites of Mexico and US (Kreidenweis et al., 2001), higher contents of sulfate from local pollution were examined, which lead to more hygroscopic particles with lower m_{R} and higher ϖ_{532}. Potential systematic difference in retrieval results may be also caused by different instruments and retrieval methods used. For example, in spite of the same datasets as Dubovik et al. (2002), Yamasoe et al. (1998) derived systematically higher m_{R} (from 1.53 at 440 nm to 1.58 at 1020 nm, results are not shown here) with a different retrieval strategy.

It is worth mentioning that PLDRs in these cases are not negligible. The phenomenon that BBA in the upper troposphere and lower stratosphere (UTLS) has higher PLDR than that in the troposphere, as shown in Figure 5.2, has been previously observed (Hu et al., 2019; Khaykin et al., 2020). An enhancement of PLDR of the BBA from the Oregon fire compared to the BBA from the Creek fire can also be identified from Figure 5.5 and Figure 5.6. Therefore, for Case 2 and Case 4, we compared the results derived from the non-spherical models (i.e., the Spheroid and IH models) with those derived from the Sphere model (not shown here). When inverting ($3\beta + 2\alpha$), we found, comparing to the Sphere model, the non-spherical models lead to higher SSA due to a reduction in m_{I} and at the same time, higher $\varepsilon_{\mathrm{fit}}$ (e.g., in Case 4, $\varepsilon_{\mathrm{fit}}$ increases to 20% for the Spheroid model and 34% for the IH model, compared to 8% for the Sphere model.) When inverting ($3\beta + 2\alpha + 3\delta$), the usage of non-spherical models leads to an extra mode centered at ~ 0.05 μm, as well as high fitting error (15% for the Spheroid model and 19% for the IH model).

Table 5.3. Microphysical properties of fresh or aged BBA reported by other authors

Reference	Instrument	Source	Ageing	r_{eff}(μm)	m_{R}	m_{I}	ϖ_{532}
Dubovik et al. (2002)	Sun phot.	Amazon forest	Statistic averaged	-	1.47±0.03	0.009±0.003	0.93±0.02
		US, Canada boreal		-	1.5±0.04	0.009±0.003	0.94±0.02
		Brazil cerrado	Statistic averaged	-	1.52±0.01	0.015±0.004	0.89±0.03
		Zambia savanna		-	1.51±0.01	0.021±0.004	0.84±0.02
Kreidenweis et al. (2001)	Sun phot. + In-Situ	US, Mexico	2 days-2 weeks	0.15-0.16	1.41-1.45	-	0.97-0.98
Wandinger et al. (2002)	Lidar	Canada boreal	8 days	0.16-0.27	1.64-1.77	0.043-0.053	0.79-0.83
Müller et al. (2005)	Lidar	Canada boreal	2 weeks	0.24-0.4	1.39-1.56	0.001-0.006	0.89-0.98
Alados-Arboledas et al. (2011)	Lidar + Star phot.	Iberian forest	1 day	0.13-0.17	1.49-1.53	0.02	0.8-0.87
Pereira et al. (2014)	Lidar	Iberian forest	1-2 days	0.17-0.19	1.49-1.61	0.01-0.024	0.89-0.95

5.2 Desert dust aerosols

5.2.1 Overview of the selected desert dust cases

As mentioned previously (Sect. 3.4.2.1, Sect. 4.1.1), the optical and microphysical properties of dust aerosols depend on both properties of the source soil and transport processes. Thus, microphysical properties of fresh dust (Case 1) and transported dust (Case 2) are retrieved and compared with each other. Case 1 records the dust episode observed on 15 April 2019 at a meteorological station in Kashi, China (39.50N, 75.93E), close to the Taklamakan desert. It belongs to the first phase of the DAO campaign where intensive field measurements close enough to the Taklamakan dust source were taken during the dust outbreak season (Hu et al., 2020). Figure 5.9 shows the range-corrected signal at 532 nm starting at 11:00 UTC 15 April 2019, from which we can tell the BL height started to rises at around 15:00 UTC and cirrus clouds were continuously present. Figure 5.10 shows the time-averaged optical profiles between 18:00 and 20:00 UTC 15 April. It can be seen that the optical properties are quite stable below 2.2 km the value of PLDR indicates the pure dust layer. We average the profiles between 2 and 2.2 km perform retrieval with BOREAL. Detailed information about Case 1 is summarized in Table 5.4 and can be further found in Hu et al. (2020).

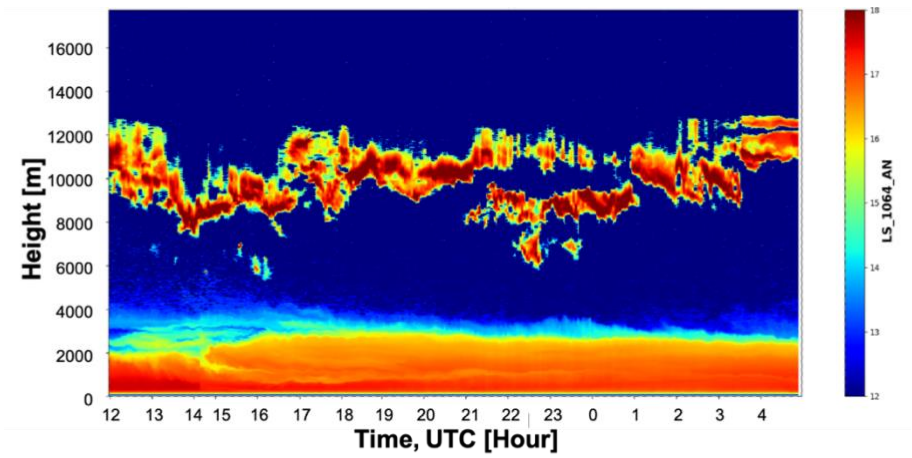


Figure 5.9. LILAS range-corrected backscattered signals at 1064 nm since 11:00 UTC 15 April 2019, at Kashi, China (cited from Hu et al. (2020), AUSTRAL processing).

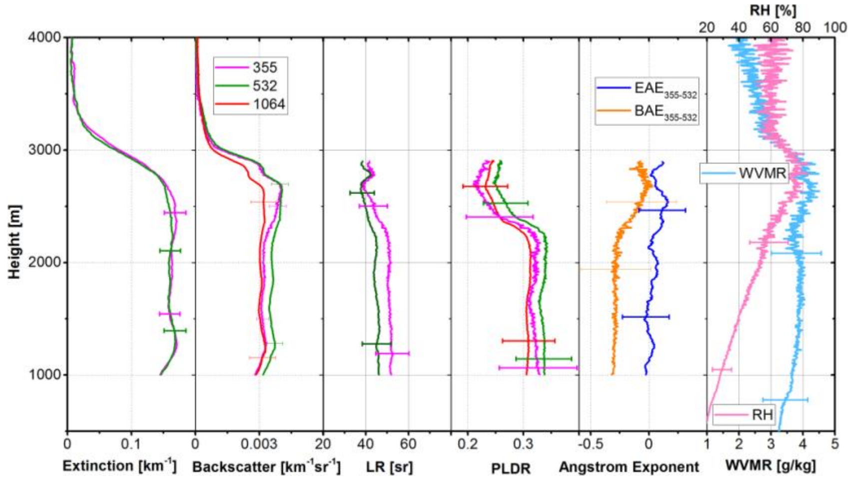


Figure 5.10. Optical profiles averaged for the period 18:00-20:00 UTC, 15 April 2019, at Kashi site: (from left to right) extinction coefficient (α), backscattering coefficient (β), lidar ratio (LR), PLDR, extinction Angstrom exponent (EAE) and backscattering Angstrom exponent (BAE), as well as water vapor mixing ratio and relative humidity (cited from Hu et al. (2020), AUSTRAL processing).

Case 2 records an aerosol event on 21 March 2022 observed at ATOLL/Lille. Figure 5.11 presents the range-corrected signals at 1064 nm on that day. An elevated but not continuous layer above the BL with high VDR can be identified. Figure 5.12 shows the time-averaged optical profiles between 20:00 and 23:00 UTC 21 March. Compared with Case 1, the PLDR at 355 in Case 2 is noisier in the whole layer. We average the profiles between 5.4 and 5.6 km perform retrieval with BOREAL. Back trajectory of this layer and the UVAI map from OMPS (Figure 5.13) indicate it could be related to a dust event in the Saharan region 7 days ago: the aerosol mass crossed over the Mediterranean circled above middle France before finally appeared over Lille. Detailed information about Case 2 is summarized in Table 5.4.

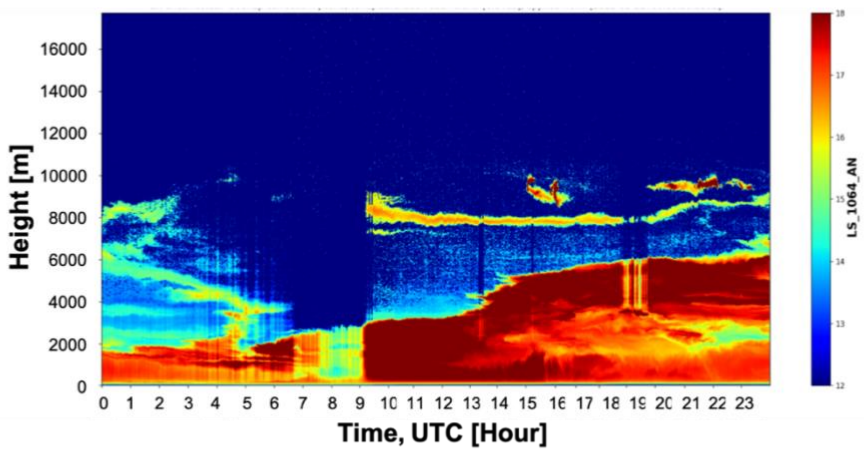


Figure 5.11. LILAS range-corrected backscattered signals (top) at 1064 nm on 21 March 2022, at ATOLL/Lille (AUSTRAL processing).

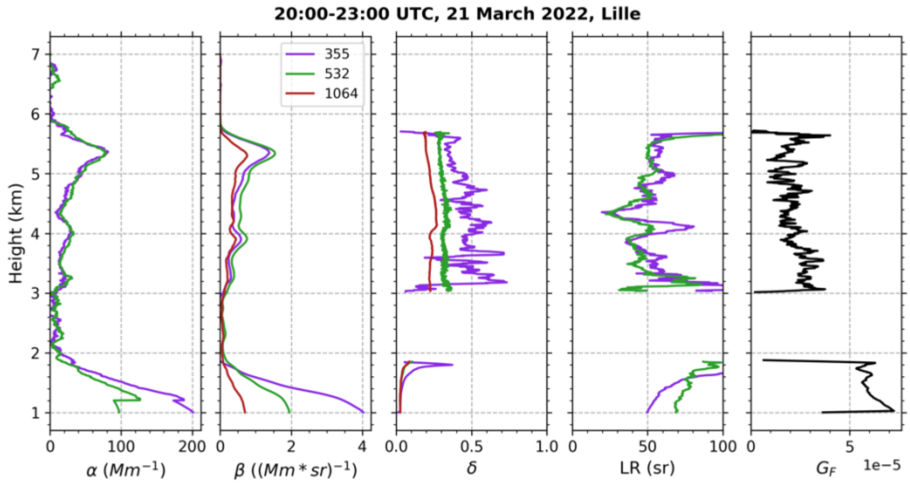


Figure 5.12. Optical profiles averaged for the period 20:00-23:00 UTC, 21 March 2022, at ATOLL/Lille: (from left to right) α, β, δ, LR and G_F (AUSTRAL processing).

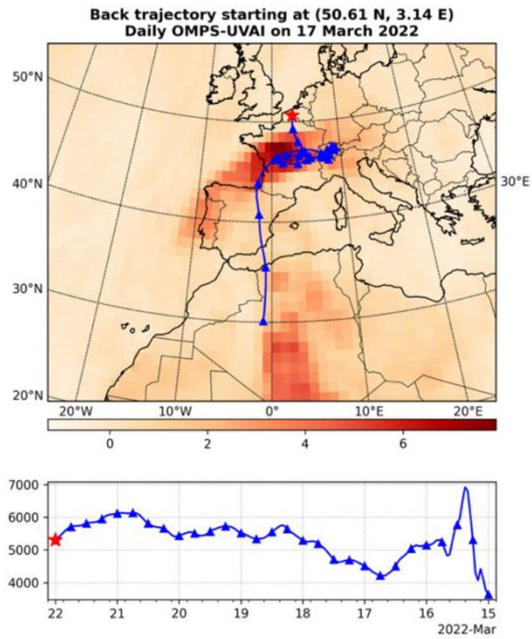


Figure 5.13. 7-day backward trajectory starting on 22 March 2022 at 5.3 km represented by the blue curve, over ATOLL/Lille (start marker), together with the UVAI measured by OMPS on 15 May 2023.

Table 5.4. Summary of the selected dust cases.

	Case 1	**Case 2**
Date	15 April 2019	21 March 2022
Time period (UTC)	18:00-20:00	20:00-23:00
Site	Kashi (39.50N, 75.93E)	Lille (50.61N, 3.14E)
Dust source	Taklamakan	Sahara
Ageing	Fresh	~ 7 days
Layer (km)	2-2.2	5.4-5.6
Layer AOD_{532}	0.032	0.014
AERONET retrievals	03:50, 15 April 2019	15:58, 21 March 2022
Columnar AOD_{440}	0.646	0.284

Like the BBA cases, we also present the AERONET retrievals passing the checking criteria in Sect. 5.1.1. The selected AERONET measurement times, together with the columnar AODs at 440 nm, are specified in Table 5.4.

5.2.2 Results and discussion

Figure 5.14 and Figure 5.15 show the results of retrieval (*a-d*) and measurement fitting (*e-f*) for the selected cases, together with the AERONET retrievals. The retrieved r_{eff}, V_{t} and $\varepsilon_{\mathrm{fit}}$ (RMS of the fitting error) are summarized in Table 5.5. Spherical and non-spherical models are used and measurements with and without 3δ are inverted. Among the selected cases, Case 2 has larger fitting error compared to Case 1, indicating a possible degradation of retrieval quality. In both cases, particle size is retrieved as a monomodal distribution no matter which retrieval configuration is employed. Comparing Case 1 and Case 2, the latter has a smaller r_{eff}, in accordance with previous studies on dust transport processes showing a decrease of particle size compared to freshly emitted particles due to deposition (Maring et al., 2003; Reid et al., 2008). When only (3β + 2α) are inverted, different scattering models result in limited changes in VSD, V_{t} and r_{eff}. After including 3δ, mode radius is retrieved larger in the transported dust case while keeps unchanged in the fresh dust case, whereas V_{t} and r_{eff} are apparently higher for both cases. With regard to the CRI retrieval, there are some features consistent with the simulation results. Firstly, using the Sphere model always leads to the lowest m_{R} and the highest m_{I}; secondly, considering additional 3δ measurements reduce m_{I}, except for the IH model in Case 1. The reduced m_{I}, in turns, results in an increase of SSA. The values derived from the (3β + 2α + 3δ, non-spherical model) and from the (3β + 2α, Sphere) configurations make the upper and lower bounds of retrieved SSA.

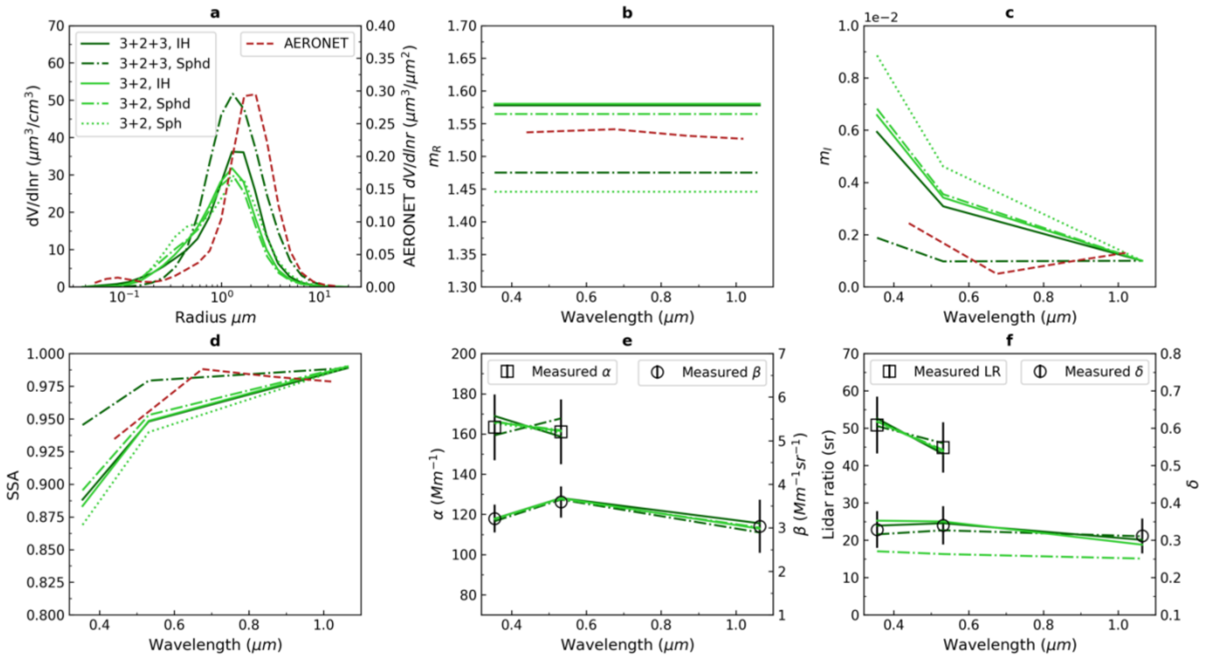


Figure 5.14. For Case 1 (15 April 2019, Kashi), retrieval of (a) VSD, (b) m_R, (c) m_I, (d) SSA and fitting of (e) α and β, (f) lidar ratio and δ. The results retrieved from (3β + 2α + 3δ) and (3β + 2α) measurements are shown in dark- and light- green lines, respectively. The results derived with the IH, Spheroid and Sphere models are shown in solid, dash-dot and dot lines, respectively. Refer to Table 4.10 and Table 4.11 for retrieval accuracy. AERONET retrieval indicated in Table 5.4 is also shown (red dashed lines) for comparison.

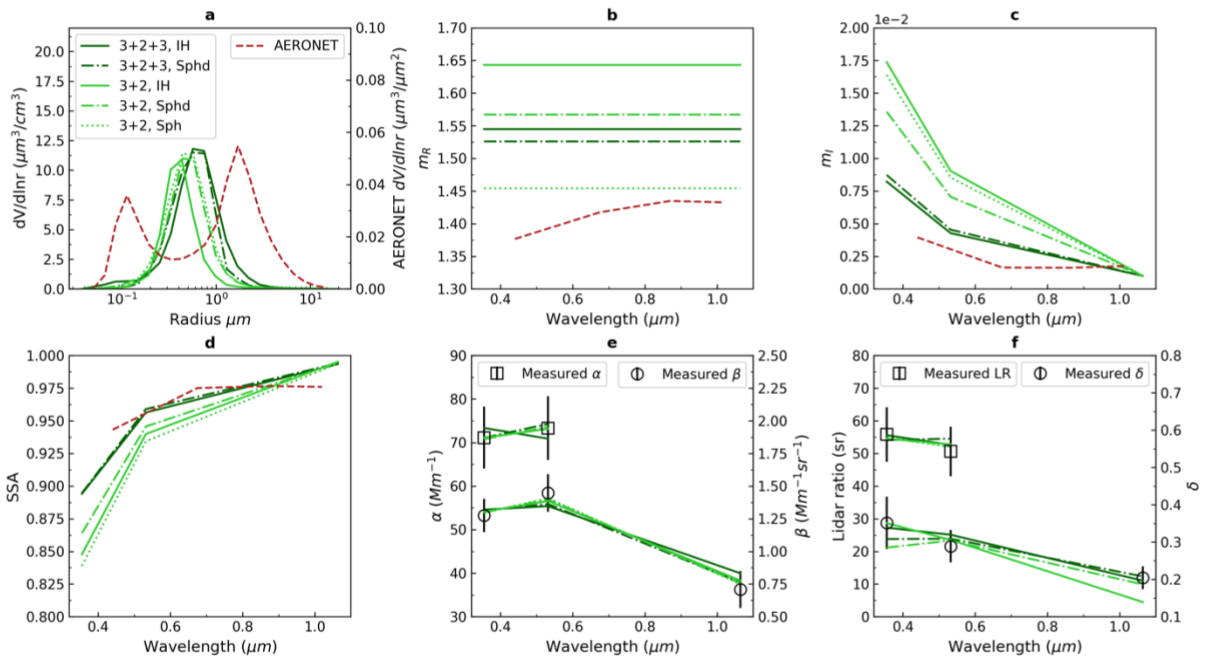


Figure 5.15. Same as Figure 5.14, but for Case 2 (21 March 2020, Lille).

Table 5.5. For each dust case, r_{eff}, V_{t} and $\varepsilon_{\text{fit}}(\%)$ retrieved with Sphere, Spheroid and IH models from (3β + 2α) and (3β + 2α + 3δ) measurements, together with the AERONET-retrieved r_{eff} for comparison.

Retrieval configuration		3β + 2α			3β + 2α + 3δ		**AERONET**
Retrieval results		**Sphere**	**Spheroid**	**IH**	**Spheroid**	**IH**	
$r_{\text{eff}}(\mu\text{m})$	**Case 1**	0.72	0.71	0.78	1.12	0.8	1.68
	Case 2	0.46	0.48	0.47	0.54	0.55	1.45
$V_{\text{t}}(\mu\text{m}^3/\text{cm}^3)$	**Case 1**	57	51	53	83	58	
	Case 2	13	12	11	14	14	
$\varepsilon_{\text{fit}}(\%)$	**Case 1**	1.2	1.4	1.6	3.2	2.6	
	Case 2	3.1	3.4	5.6	6.5	7.9	

In spite of the fact that fitting errors for the Spheroid and IH models are not significant and similar with each other, one should be aware of the differences in retrieval results due to the use of different scattering models. Discrepancy between the non-spherical models in the results retrieved from (3β + 2α + 3δ) should be larger than the results from (3β + 2α) measurements because the main contrast between the IH and Spheroid models shows in the reproduction of δ (see Figure 4.19 and Figure 4.20). This has been proved in Case 1, whereas in Case 2 the situation is reversed: the inversion of (3β + 2α) brings larger discrepancy between the two non-spherical models. The reason is not currently understood but might be partly related to the measurements with degraded quality – as shown in Figure 5.11, δ_{355} is quite noisy for Case 2. In this regard, more cases of transported dust with better measurement quality are needed. However, comparing the two cases shows that when inverting (3β + 2α + 3δ), difference between the IH and Spheroid models becomes larger for larger particle r_{eff} (supported by Figure 4.20). In addition, the positive correlation between m_{R} and m_{I}, and the negative correlation between m_{R} and V_{t} are also consistent with rules of cross-talk discussed in Chap. 4.

For Case 1, the difference between the two non-spherical models could be further examined via calculating intensive lidar quantities (i.e., lidar ratio and PLDR) from the AERONET-retrieved microphysical properties. Note that for AERONET retrieval, the results derived from using the Spheroid model and IH models as the forward model should be enough similar. It is because the measurements to invert – i.e., the sun irradiance and sky radiance are only related to $\langle C_{\text{ext}}\rangle$, $\langle C_{\text{sca}}\rangle$ and $\langle P_{11}\rangle$, which are calculated to be similar by the two non-spherical models, unless for the backward direction. The calculation results, together with the lidar measurements are shown in Figure 5.16. It can be seen that the lidar ratios reproduced by the two models are quite consistent with each other and with the measurements in the UV-VIS region. However, as the wavelength increases, the IH model generates higher lidar ratio but both models underestimate the lidar measurement. The PLDR reproduced by the Spheroid model is obviously smaller than that by the IH model, as well as smaller than the measurements, especially in the UV-VIS region. Accordingly, compared to the IH model, the Spheroid model retrieves lower m_R and m_I in order to produce higher values of lidar ratio and PLDR that fit the measurements.

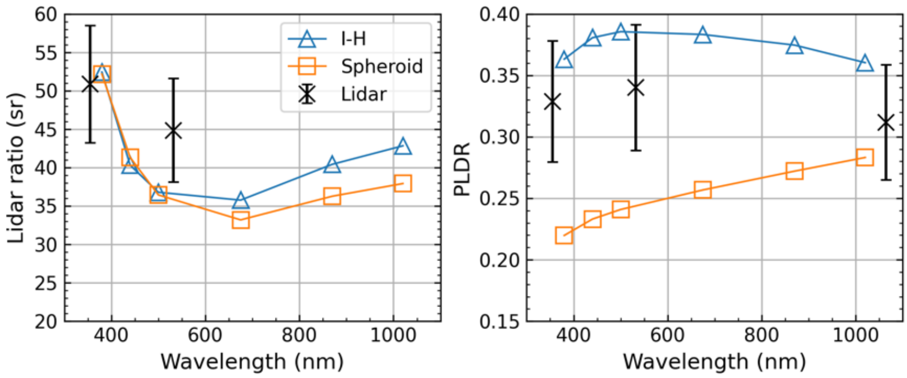


Figure 5.16. Reproduced lidar ratio and PLDR by the IH model (blue solid lines with the triangle marker) and Spheroid model (orange solid lines with the square marker) from the AERONET retrieval in Case 1, as well as the lidar measurements with measurement errors for the Case 1.

The AERONET-retrieved VSDs contain fine and coarse modes in both cases. Compared with the Case 1, a larger fraction of the fine-mode particles is found in Case 2, which could be from the BL. Both AERONET retrievals present coarse modes with larger r_V and r_{eff} compared to the BOREAL retrievals. This could be partly explained by the lack of sensitivity of the lidar-related optical Kernels to large particles, as have been demonstrated in Chap. 3. The angular scattering measurements of the sun-sky photometer is crucial for retrieving the size distribution of large particles (Eck et al., 2008). Müller et al. (2013) also reported smaller dust r_{eff} retrieved from ($3\beta + 2\alpha + 1\delta$) lidar measurements, compared with the aircraft in-situ measurements (Weinzierl et al., 2009). With regard to CRI and SSA, for Case 1, the AERONET-retrieved values are between those derived from the configurations of ($3\beta + 2\alpha + 1\delta$, IH) and ($3\beta + 2\alpha + 1\delta$, Spheroid). Whereas in Case 2, the AERONET retrieval may be more impacted by the particles within the BL, showing a columnar m_R lower than typical pure dust aerosol.

5.3 Continental aerosol pollutants

The location of northern France makes it not only a crossroad of various transported aerosols but also a pool of multiple particulate pollutants. The Hauts-de-France region is the

2nd most densely populated area in France with 6 million inhabitants, affected by intense local traffic, residential, industrial and agricultural emissions as well as continental pollution from central Europe (Roig Rodelas et al., 2019). Long-term in-situ measurements of submicron (PM_1) particles at the ATOLL platform indicate a dominated contribution of organic aerosols (OA), followed by, from high to low, contributions of nitrate, ammonium, sulfate and BC (Velazquez-Garcia et al., 2023; Chebaicheb et al., 2023). Here, we present a case study of retrieving the microphysical properties of continental aerosol pollutants observed by LILAS at ATOLL. Detailed information is shown in Table 5.6. On 4 March 2022, dense aerosol loading in the BL was detected (Figure 5.17). Daily AOD_{440} measured by AERONET was up to 0.44 and extinction Angstrom exponent ($EAE_{440\text{-}870}$) was 1.6, suggesting the dominance of fine-mode particles. Figure 5.18 shows the measured optical profiles averaged for the period 20:00-22:00 UTC, 4 March. We average the layer between 1.5 and 1.7 km to perform the retrieval. The low PLDR and fluorescence capacity indicate the presence of urban-type aerosol (Veselovskii et al., 2022).

Table 5.6. Information about the continental aerosol case

Date	4 March 2022
Time period (UTC)	20:00-22:00
Site	Lille (50.61N, 3.14E)
Aerosol source	Northern German
Layer (km)	1.5-1.7
Layer AOD_{532}	0.044
AERONET retrievals	15:20 UTC, 4 March 2022
Columnar AOD_{440}	0.449

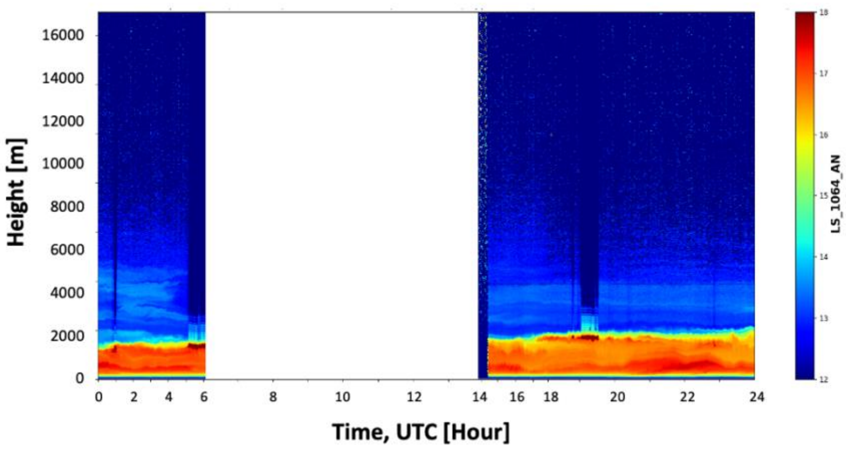


Figure 5.17. LILAS range-corrected signals at 1064 nm from 00:00 to 24:00 UTC 4 March 2022, at ATOLL/Lille (AUSTRAL processing).

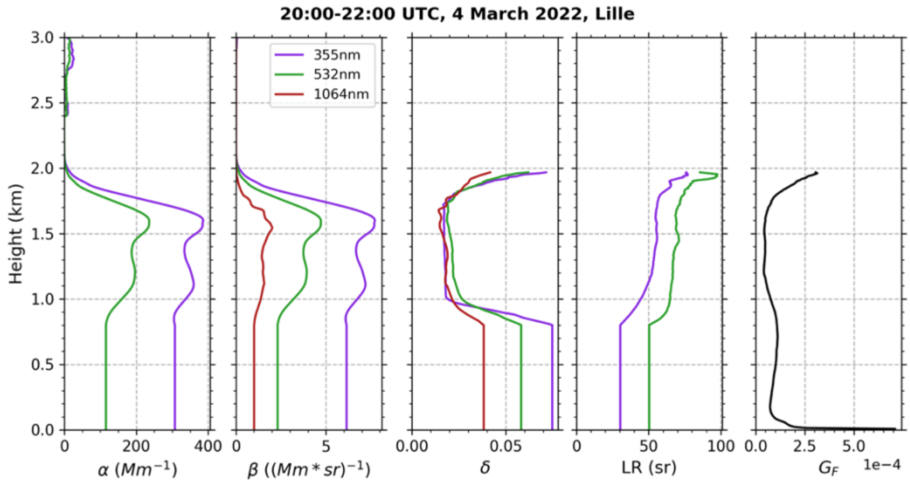


Figure 5.18. Same as Figure 5.12 but for the period 20:00-22:00 UTC, 4 March 2022, at ATOLL/Lille

Figure 5.19 shows the retrieved VSD, CRI and SSA of the selected layer, together with the AERONET level 2 retrieval product derived from the closest measurement (15:20 UTC, 4 March). Compared to the transported BBA and DA cases, here the results retrieved with BOREAL and AERONET should be more comparable because from Figure 5.17 one can see most of the aerosol particles were concentrated in the BL. It can be seen that the selected layer is dominated by fine particles with a r_{eff} (0.179 μm) consistent with AERONET-retrieved fine-mode r_{eff} (0.177 μm). On the other hand, compared with the lidar-retrieved values, the AERONET-retrieved m_{R} is lower, suggesting the existence of water-soluble particles, while the m_{I} is higher, especially towards the longer wavelengths. As a result, the AERONET-retrieved SSA is smaller than the lidar-retrieved value, especially towards the longer wavelengths. As discussed in Chap. 4, due to the influence of the *a priori* constraint on m_{R}, BOREAL tends to overestimate the real part when particles' real and imaginary parts are both low. However, the discrepancy in CRI between lidar and AERONET retrievals might also result from the coarse-mode particles which are retrieved by AERONET only.

The CRI and SSA retrieved with BOREAL from lidar measurements show the particles are not absorbing, which is in line with the results of in-situ measurements. Analyses of multi-year Aerosol Chemical Speciation Monitor (ACSM) measurements at ATOLL point out that in spring, the emitted carbonaceous components (OC and BC, which are main contributors to the absorption coefficient of the regional aerosol) usually decreases to a relatively low level due to the reduction of residential heating activities compared to winter; whereas the amount of nitrate and ammonium aerosols (less absorbing compared to OC and BC) significantly increases due to the use of fertilizers in agricultural activities and gas-to-particle conversion of NOx from traffic emission under favorable photochemical conditions (Velazquez-Garcia et al., 2023;

Chebaicheb et al., 2023). Furthermore, the 2-day back trajectory in Figure 5.20 shows the retrieved aerosol mass was transported from the northeastern side to the region, where significant contributions of ammonium nitrate have been identified (Roig Rodelas et al., 2019).

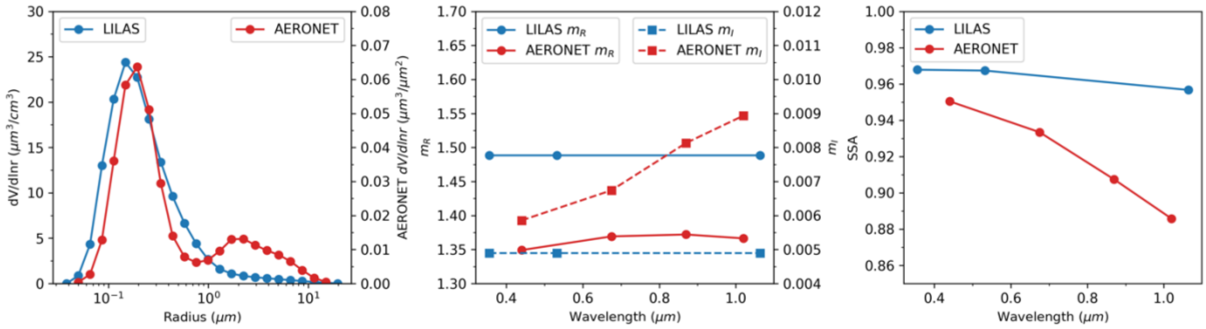


Figure 5.19. VSD, CRI and SSA of the selected layer retrieved with BOREAL from LILAS measurements, together with the AERONET level 2 retrieval product derived from the closest measurement (15:20 UTC, 4 March). Refer to Table 4.8 for BOREAL retrieval accuracy, and Table 5.6 for layer and columnar AODs.

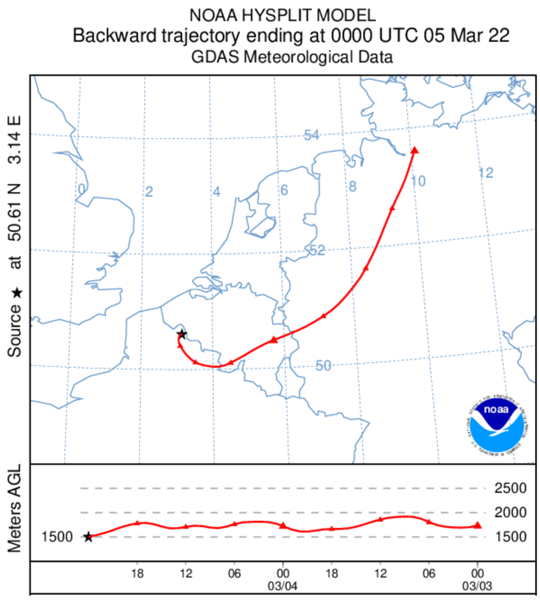


Figure 5.20. 2-day back trajectory starting on 00:00 UTC 5 March 2022 over ATOLL/Lille.

5.4 Chapter summary

In this chapter, we applied BOREAL to lidar measurements to retrieve and interpret typical aerosol types, including biomass burning aerosol (BBA), dust aerosol and continental polluted aerosol. Several cases were studied and compared with results from AERONET, previous studies and meteorology data to partly validate the reliability of BOREAL. In the cases of BBA, we retrieved microphysical properties of aged BBA particles consistent with previous studies: for instance, increasing particle size and SSA compared to freshly emitted BBA particles. We are also able to identify the influences of the burning source and ambient conditions on the microphysical properties of BBA. The retrievals of fresh and transported dust aerosols show a decrease of particle size for the latter, which is in accordance with previous studies showing it is caused by deposition of the coarse mode during the transport process. All three models are able to fit the measurement well. However, one should be aware of the differences in retrieval results caused by these models, which is another evidence of the underdetermination. The influence of assuming spherical particles in the retrieval is consistent with our simulation results in the last chapter: for instance, decrease of the real part and increase of the imaginary part. The difference between the Spheroid and IH models when inverting ($3\beta + 2\alpha + 3\delta$) measurements increases with the increase of particle size. More dust measurements, especially for transported dust, with high signal-to-noise ratio (SNR) are needed to conduct comprehensive comparison of the two non-spherical models. In the continental aerosol case, we successfully located the source of the aerosol mass and retrieved the microphysical properties with BOREAL, from which we infer it could be related to agricultural activities (fertilizing) or traffic emission (gas-to-particle conversion) in neighbor countries. In addition, comparison of lidar retrievals with AERONET retrievals is nontrivial because of differences in measurement time and vertical resolution. Comparisons with the results retrieved with other lidar-aerosol retrieval algorithms and/or with air-borne in-situ measurements are needed to further validate BOREAL retrievals.

6 Conclusions and perspectives

6.1 Conclusions

Height-resolved information on aerosol optical and microphysical properties provided by lidar is of significant importance to understand the vertical distribution, transport and evolution of aerosols, which in turns helps in data assimilation and better constraining atmospheric models. Compared with aerosol optical properties which are directly measured by lidar, acquisition of aerosol microphysical properties has to resort to "retrieval methods" and depends on the sensitivity of the measured optical properties, making it a challenging task. Regarding more optical quantities detectable by advanced lidar systems and increasing requirement of more detailed aerosol microphysical properties, this thesis aims at retrieving heigh-resolved aerosol microphysical properties from measurements of multi-wavelength Mie-Raman-depolarization lidars. As a result, the Basic algOrithm for REtrieval of Aerosol with Lidar (BOREAL) algorithm has been developed, tested and applied. Compared with traditional linear retrieval algorithms which prevailed in the past two decades, it is based on maximum likelihood estimation and highlighted by its flexibility, automaticity, as well as capabilities of inverting non-linear models and incorporating *a priori* constraints of different kinds.

Detailed description of algorithmic principles as well as implementation is given in Chap. 3. BOREAL inverts (3β + 2α) or (3β + 2α + 3δ) measurements into particle volume size distribution (VSD), real part (m_R) and imaginary part (m_I) of the complex refractive index (CRI), from which particle total volume concentration (V_t), effective radius (r_{eff}) and bulk single scattering albedo (SSA) are then calculated. The core of the inversion procedure is non-linear fitting to the measurement and constraint terms weighted by the errors. Three physical models, i.e., the Sphere, Spheroid and Irregular-Hexahedral (IH) models, are integrated into BOREAL for forward calculation (i.e., forward model) of the fitting process. These models describe scattering processes of spherical and non-spherical particles, respectively. To completely cover the size range of the aerosol to be retrieved, retrievals are performed for a series of pre-defined inversion windows to derive a set of individual solutions. The final solution is the mean of the qualified individual solutions (i.e., solution space) selected by criteria based on fitting error and VSD shape. Two kinds of constraints, namely the smoothing constraint on size distribution and *a priori* constraint on CRI, are adopted. Different modes of the *a priori* CRI constraints are included in BOREAL, allowing free switch according to the objective aerosol type (absorbing, non-absorbing and dust). Furthermore, it is worth pointing out that the non-linear strategy used in BOREAL allows it to flexibly incorporate measurements related to particle microphysical properties through non-linear physical models.

The results of sensitivity study and the assessment of BOREAL retrieval accuracy based on simulated data are presented in Chap. 4. When retrieving spherical particles from conventional ($3\beta + 2\alpha$) measurements, the best accuracy is achieved for particle ensembles with fine-mode dominated VSD and relatively medium or high m_R and m_I. One of the main factors that influence retrieval stability turns out to be the cross-talk between state variables, which stems from both lack of sensitivity and underdetermination of the retrieval system. It stresses the importance of the *a priori* constraint on CRI, particularly on m_I to improve the accuracy. On the other hand, the results of retrieving non-spherical particles demonstrate that: (1) incorporating 3δ measurements is of significant importance to improve the retrieval accuracy, particularly the accuracy of the CRI; (2) assuming non-spherical particles to be spherical undermines retrieval accuracy in two possible ways depending on the true CRI: underestimates the real part or overestimates the imaginary part and, at the same time, severely underestimate the volume concentration. Results of random noise tests show for both spherical and non-spherical particle cases, the reduction of retrieval accuracy due to measurement noise is less than or comparable with the measurement standard deviation, indicating its robustness on resisting measurement noise.

Application of BOREAL to real lidar measurements in Chap. 5 generally shows consistency between BOREAL-retrieved results and those from AERONET inversion and previous studies. The inversions of observations of long-range transported biomass burning aerosols (BBAs) derive typical microphysical properties of aged BBA. In the retrievals of dust aerosol cases, a shift of the distribution mode to smaller size for transported dust compared to fresh dust is observed, which is an evidence of particle deposition during the transport process. The retrieval of aerosol pollutants concentrated in the boundary layer (BL) shows they were composed of fine-mode particles with less absorption. Combining the back trajectory analysis and previous studies on seasonal variation of aerosol components in this region, we infer there is a large fraction of transported nitrate and ammonium which are mainly emitted from springtime agricultural activates and converted from NOx precursors of traffic emission.

This study proves the flexibility of BOREAL in incorporating multi-types of information apart from conventional extinction and backscattering measurements, which is important to better constrain and stabilize aerosol microphysical property retrieval using lidar measurements. On the other hand, the algorithm is highly unsupervised since, by default, the only argument to specify is the aerosol type and then the retrieval process is fully automated. Therefore, it is particularly useful for inverting massive measurements of a certain aerosol type and combines with lidar-aerosol typing algorithms like NATALI (Nicolae et al., 2018) or FLARE (Miri al., submitted). Last but not least, we emphasize that BOREAL is a generalized lidar-aerosol retrieval algorithm which can be applied to not only to LILAS data but also to ($3\beta + 2\alpha + 3\delta$) from any lidar system with measurement accuracy comparable to LILAS. Its flexibility also allows one to easily test reduced dataset (e.g., $3\beta + 2\alpha + 1\delta$ with 1δ at 532 nm) or augmented dataset (e.g., $3\beta + 2\alpha + 3\delta$ with an extra α at 1064 nm), as well as different forward models. Thus, it has potentiality to be a basic retrieval algorithm for lidar communities (e.g., EARLINET). In collaboration with the AUSTRAL development team in LOA, the first version of BOREAL has been integrated into AUSTRAL as retrieval module following its lidar data processing chain, which will make it an efficient tool to analyze massive lidar measurements in an automated manner.

6.2 Perspectives

There are several interesting perspectives following this work. Firstly, we are aware that there remains large potentiality of improvement for BOREAL. For example, one problem when retrieving aerosol with BOREAL is that the retrieved CRI is actually bounded by the ranges of m_{R} and m_{I} prescribed by the forward model which, however, cannot be taken into account in the optimization procedure. As a result, overflow occurs sometimes during the iteration if the m_{R}, m_{I} values are close to the edges of the domain. Thus, some constrained optimization approaches for nonlinear minimization subject to bounds might be considered in future. We saw a trend of underestimating particle size in the retrievals of large particles and we owed it to the lack of sensitivity of available lidar measurements. It might be partly solved by making use of the "nonlinear" characteristic of BOREAL and retrieving parameterized size distribution (e.g., retrieving r_{V}, σ_{g} of a lognormal distribution) which, on the one hand, is easier to be constrained by *a priori* knowledge on aerosol types and, on the other hand, reduces the total number of retrieved variables. In this study, aerosol mixture was not considered, which is, however, a quite common status of aerosols in the nature. To retrieve the mixing state of aerosol, future work should be planned for finding optimal variables which could represent mixing components and, at the same time, do not increase system underdetermination and make measurements have enough sensitivity. For this purpose, aerosol mixing rules (e.g., Maxwell Garnett effective medium approximation, Bruggeman approximation and so on) might be applied.

This study figured out the importance of reliable *a priori* information to enhance aerosol characterization from lidars. Accordingly, improvements in lidar-aerosol typing algorithms are necessary to establish a type-resolved *a priori* database of microphysical properties. For example, an aerosol typing algorithm (FLARE) based on machine learning, making use of the extinction, backscattering, depolarization, fluorescence and relative humidity from LILAS, is being developed in LOA (Miri et al., submitted). Meanwhile, we are developing a more powerful lidar system, LIFE (Laser-Induced Fluorescence Explorer), with an additional 1064 nm extinction channel and 4-5 fluorescence channels, whose measurements allow more possibility in aerosol characterization with BOREAL (Boissière et al., in preparation).

Furthermore, possible synergy with other instruments (e.g., sun-sky photometers) is also attractive given that information contents of stand-alone lidar measurements are still insufficient to accurately retrieve aerosol microphysical properties in some cases (e.g., size distribution of large particles).

Both validation and application purposes necessitate more case studies. In particular, more dust cases with high measurement quality are needed for a more comprehensive assessment of non-spherical scattering models. Retrieval of UTLS BBA and comparison with tropospheric BBA are also of great interest (Hu et al., in preparation) in order to evaluate its impacts on radiative balance, as a contribution to the PyroSrat project (Khaykin et al., 2020). In addition, we are planning to invert lidar measurements taken in periods when simultaneous airborne in-situ measurements are available so as to compare the inversion results with the in-situ measured results as a way of validation. For example, we could benefit from the AERO-HDF campaign conducted by LOA in July 2023, over the Hauts-de-France region.

Bibliography

Abel, S. J., Haywood, J. M., Highwood, E. J., Li, J., and Buseck, P. R.: Evolution of biomass burning aerosol properties from an agricultural fire in southern Africa: SMOKE PROPERTIES FROM BIOMASS BURNING, Geophys. Res. Lett., 30, https://doi.org/10.1029/2003GL017342, 2003.

Alados-Arboledas, L., Müller, D., Guerrero-Rascado, J. L., Navas-Guzmán, F., Pérez-Ramírez, D., and Olmo, F. J.: Optical and microphysical properties of fresh biomass burning aerosol retrieved by Raman lidar, and star-and sun-photometry, Geophysical Research Letters, 38, https://doi.org/10.1029/2010GL045999, 2011.

Albrecht, B. A.: Aerosols, Cloud Microphysics, and Fractional Cloudiness, Science, 245, 1227–1230, https://doi.org/10.1126/science.245.4923.1227, 1989.

Alfaro, S. C., Gaudichet, A., Gomes, L., and Maillé, M.: Modeling the size distribution of a soil aerosol produced by sandblasting, Journal of Geophysical Research: Atmospheres, 102, 11239–11249, https://doi.org/10.1029/97JD00403, 1997.

Ali, A., Amin, S. E., Ramadan, H. H., and Tolba, M. F.: Enhancement of OMI aerosol optical depth data assimilation using artificial neural network, Neural Comput & Applic, 23, 2267–2279, https://doi.org/10.1007/s00521-012-1178-9, 2013.

Andreae, M. O.: Emission of trace gases and aerosols from biomass burning – an updated assessment, Atmospheric Chemistry and Physics, 19, 8523–8546, https://doi.org/10.5194/acp-19-8523-2019, 2019.

Ansmann, A. and Müller, D.: Lidar and Atmospheric Aerosol Particles, in: Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, Springer Science and Business Media, Inc., Singapore, 2005.

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, https://doi.org/10.1007/BF00348608, 1992.

Bender, F. A.-M., Frey, L., McCoy, D. T., Grosvenor, D. P., and Mohrmann, J. K.: Assessment of aerosol–cloud–radiation correlations in satellite observations, climate models and reanalysis, Clim Dyn, 52, 4371–4392, https://doi.org/10.1007/s00382-018-4384-z, 2019.

Berdnik, V. V. and Loiko, V. A.: Neural networks for aerosol particles characterization, Journal of Quantitative Spectroscopy and Radiative Transfer, 184, 135–145, https://doi.org/10.1016/j.jqsrt.2016.06.034, 2016.

Bi, L., Yang, P., Kattawar, G. W., and Kahn, R.: Modeling optical properties of mineral aerosol particles by using nonsymmetric hexahedra, Appl. Opt., AO, 49, 334–342, https://doi.org/10.1364/AO.49.000334, 2010.

Böckmann, C.: Hybrid regularization method for the ill-posed inversion of multiwavelength lidar data in the retrieval of aerosol size distributions, Appl. Opt., 40, 1329–1342, https://doi.org/10.1364/AO.40.001329, 2001.

Böckmann, C. and Kirsche, A.: Iterative regularization method for lidar remote sensing, Computer Physics Communications, 174, 607–615, https://doi.org/10.1016/j.cpc.2005.12.019, 2006.

Böckmann, C. and Osterloh, L.: Runge-Kutta type regularization method for inversion of spheroidal particle distribution from limited optical data, Inverse Problems in Science and Engineering, 22, 150–165, https://doi.org/10.1080/17415977.2013.830615, 2014.

Böckmann, C., Mironova, I., Müller, D., Schneidenbach, L., and Nessler, R.: Microphysical aerosol parameters from multiwavelength lidar, J. Opt. Soc. Am. A, 22, 518–528, https://doi.org/10.1364/JOSAA.22.000518, 2005.

Bohren, C. F. and Huffman, D. R.: Absorption and scattering of light by small particles, Wiley-VCH, Weinheim, 530 pp., 2004.

Boucher, O., Randall, D., Artaxo, P., and Bretherton, C.: Clouds and Aerosols, in: Climate Change 2013 – The Physical Science Basis, Cambridge University Press, 571–658, https://doi.org/10.1017/CBO9781107415324.016, 2014.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, Atmospheric Measurement Techniques, 5, 73–98, https://doi.org/10.5194/amt-5-73-2012, 2012.

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$\beta +$\,2α lidar measurement system for aerosol microphysical retrievals, Atmospheric Measurement Techniques, 9, 5555–5574, https://doi.org/10.5194/amt-9-5555-2016, 2016.

Buseck, P. R. and Adachi, K.: Nanoparticles in the Atmosphere, Elements, 4, 389–394, https://doi.org/10.2113/gselements.4.6.389, 2008.

Chaikovsky, A., Dubovik, O., Holben, B., Bril, A., Goloub, P., Tanré, D., Pappalardo, G., Wandinger, U., Chaikovskaya, L., Denisov, S., Grudo, J., Lopatin, A., Karol, Y., Lapyonok, T., Amiridis, V., Ansmann, A., Apituley, A., Allados-Arboledas, L., Binietoglou, I., Boselli, A., D'Amico, G., Freudenthaler, V., Giles, D., Granados-Muñoz, M. J., Kokkalis, P., Nicolae, D., Oshchepkov, S., Papayannis, A., Perrone, M. R., Pietruczuk, A., Rocadenbosch, F., Sicard, M., Slutsker, I., Talianu, C., De Tomasi, F., Tsekeri, A., Wagner, J., and Wang, X.: Lidar-Radiometer Inversion Code (LIRIC) for the retrieval of vertical aerosol properties from combined lidar/radiometer data: development and distribution in EARLINET, Atmos. Meas. Tech., 9, 1181–1205, https://doi.org/10.5194/amt-9-1181-2016, 2016.

Chang, Y., Hu, Q., Goloub, P., Veselovskii, I., and Podvin, T.: Retrieval of Aerosol Microphysical Properties from Multi-Wavelength Mie–Raman Lidar Using Maximum

Likelihood Estimation: Algorithm, Performance, and Application, Remote Sensing, 14, 6208, https://doi.org/10.3390/rs14246208, 2022.

Chebaicheb, H., F. de Brito, J., Chen, G., Tison, E., Marchand, C., Prévôt, A. S. H., Favez, O., and Riffault, V.: Investigation of four-year chemical composition and organic aerosol sources of submicron particles at the ATOLL site in northern France, Environmental Pollution, 330, 121805, https://doi.org/10.1016/j.envpol.2023.121805, 2023.

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, Appl. Opt., 53, 7252–7266, https://doi.org/10.1364/AO.53.007252, 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, https://doi.org/10.1364/AO.55.002188, 2016.

Chen, C., Dubovik, O., Schuster, G. L., Chin, M., Henze, D. K., Lapyonok, T., Li, Z., Derimian, Y., and Zhang, Y.: Multi-angular polarimetric remote sensing to pinpoint global aerosol absorption and direct radiative forcing, Nature Communications, 13, 7459, https://doi.org/10.1038/s41467-022-35147-y, 2022.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., Martin, R. V., Logan, J. A., Higurashi, A., and Nakajima, T.: Tropospheric Aerosol Optical Thickness from the GOCART Model and Comparisons with Satellite and Sun Photometer Measurements, Journal of the Atmospheric Sciences, 59, 461–483, https://doi.org/10.1175/1520-0469(2002)059<0461:TAOTFT>2.0.CO;2, 2002.

China, S., Mazzoleni, C., Gorkowski, K., Aiken, A. C., and Dubey, M. K.: Morphology and mixing state of individual freshly emitted wildfire carbonaceous particles, Nat Commun, 4, 2122, https://doi.org/10.1038/ncomms3122, 2013.

Chong, E. K. P. and Żak, S. H.: An introduction to optimization, Fourth edition., Wiley, Hoboken, New Jersey, 622 pp., 2013.

Cox, C. and Munk, W.: Measurement of the Roughness of the Sea Surface from Photographs of the Sun's Glitter, J. Opt. Soc. Am., JOSA, 44, 838–850, https://doi.org/10.1364/JOSA.44.000838, 1954.

D'Almeida, G. A., Koepke, P., and Shettle, E. P.: Atmospheric aerosols : global climatology and radiative characteristics, A. Deepak Pub. Hampton, Va., USA, Hampton, Va., USA, 1991.

Davies, C. N.: Size distribution of atmospheric particles, Journal of Aerosol Science, 5, 293–300, https://doi.org/10.1016/0021-8502(74)90063-9, 1974.

Deirmendjian, D.: Electromagnetic Scattering on Spherical Polydispersions, American Elsevier Publishing Company, Inc., New York, 1969.

Di Biagio, C., Formenti, P., Balkanski, Y., Caponi, L., Cazaunau, M., Pangui, E., Journet, E., Nowak, S., Andreae, M. O., Kandler, K., Saeed, T., Piketh, S., Seibert, D., Williams, E., and Doussin, J.-F.: Complex refractive indices and single-scattering albedo of global dust aerosols in the shortwave spectrum and relationship to size and iron content, Atmospheric Chemistry and Physics, 19, 15503–15531, https://doi.org/10.5194/acp-19-15503-2019, 2019.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmospheric Measurement Techniques, 8, 281–299, https://doi.org/10.5194/amt-8-281-2015, 2015.

Diner, D. J., Abdou, W. A., Ackerman, T. P., Crean, K., Gordon, H. R., Kahn, R. A., Martonchik, J. V., Paradise, S. R., Pinty, B., Verstraete, M. M., Wang, M., and West, R. A.: Multi- angle Imaging SpectroRadiometer Level 2 Aerosol Retrieval Algoritlun Theoretical Basis, 2008.

Donovan, D. P. and Carswell, A. I.: Principal component analysis applied to multiwavelength lidar aerosol backscatter and extinction measurements, Appl. Opt., 36, 9406–9424, https://doi.org/10.1364/AO.36.009406, 1997.

Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, Journal of Geophysical Research: Atmospheres, 105, 20673–20696, https://doi.org/10.1029/2000JD900282, 2000.

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker, I.: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations, Journal of the Atmospheric Sciences, 59, 590–608, https://doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2, 2002.

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Muñoz, O., Veihelmann, B., van der Zande, W. J., Leon, J.-F., Sorokin, M., and Slutsker, I.: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2005JD006619, 2006.

Dubovik, O., Fuertes, D., Litvinov, P., Lopatin, A., Lapyonok, T., Doubovik, I., Xu, F., Ducos, F., Chen, C., Torres, B., Derimian, Y., Li, L., Herreras-Giralda, M., Herrera, M., Karol, Y., Matar, C., Schuster, G. L., Espinosa, R., Puthukkudy, A., Li, Z., Fischer, J., Preusker, R., Cuesta, J., Kreuter, A., Cede, A., Aspetsberger, M., Marth, D., Bindreiter, L., Hangler, A., Lanzinger, V., Holter, C., and Federspiel, C.: A Comprehensive Description of Multi-Term LSM for Applying Multiple a Priori Constraints in Problems of Atmospheric Remote Sensing: GRASP Algorithm, Concept, and Applications, Frontiers in Remote Sensing, 2, https://doi.org/10.3389/frsen.2021.706851, 2021.

Ducos, F., Hu, Q., and Popovici, I.: AUSTRAL User's Guide -- AUtomated Server for the TReatment of Atmospheric Lidars, 2022.

Eadie, W. T., Drijard, D., and James, F. E.: Statistical methods in experimental physics, North-Holland Publishing Company, New York, 155 pp., 1971.

Eck, T. F., Holben, B. N., Reid, J. S., Sinyuk, A., Dubovik, O., Smirnov, A., Giles, D., O'Neill, N. T., Tsay, S.-C., Ji, Q., Al Mandoos, A., Ramzan Khan, M., Reid, E. A., Schafer, J. S., Sorokine, M., Newcomb, W., and Slutsker, I.: Spatial and temporal variability of column-integrated aerosol optical properties in the southern Arabian Gulf and United Arab Emirates in summer, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD008944, 2008.

Elder, A., Vidyasagar, S., and DeLouise, L.: Physicochemical factors that affect metal and metal oxide nanoparticle passage across epithelial barriers, WIREs Nanomedicine and Nanobiotechnology, 1, 434–450, https://doi.org/10.1002/wnan.44, 2009.

Fedarenka, A., Dubovik, O., Goloub, P., Li, Z., Lapyonok, T., Litvinov, P., Barel, L., Gonzalez, L., Podvin, T., and Crozel, D.: Utilization of AERONET polarimetric measurements for improving retrieval of aerosol microphysics: GSFC, Beijing and Dakar data analysis, Journal of Quantitative Spectroscopy and Radiative Transfer, 179, 72–97, https://doi.org/10.1016/j.jqsrt.2016.03.021, 2016.

Fiocco, G. and Smullin, L. D.: Detection of Scattering Layers in the Upper Atmosphere (60–140 km) by Optical Radar, Nature, 199, 1275–1276, https://doi.org/10.1038/1991275a0, 1963.

Fisher, R. A. and Russell, E. J.: On the mathematical foundations of theoretical statistics, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222, 309–368, https://doi.org/10.1098/rsta.1922.0009, 1922.

Fletcher, R.: MODIFIED MARQUARDT SUBROUTINE FOR NON-LINEAR LEAST SQUARES., Atomic Energy Research Establishment, Harwell (England), 1971.

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., Lunt, D. J., Mauritsen, T., Palmer, M. D., Watanabe, M., Wild, M., and Zhang, H.: The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity, edited by: Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., Caud, N., Chen, Y., Goldfarb, L., Gomis, M. I., Huang, M., Leitzell, K., Lonnoy, E., Matthews, J. B. R., Maycock, T. K., Waterfield, T., Yelekçi, O., Yu, R., and Zhou, B., Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, 923–1054, https://doi.org/10.1017/9781009157896.009, 2021.

Freudenthaler, V., Esselborn, M., Wiegner, M., Heese, B., Tesche, M., Ansmann, A., MüLLER, D., Althausen, D., Wirth, M., Fix, A., Ehret, G., Knippertz, P., Toledano, C., Gasteiger, J., Garhammer, M., and Seefeldner, M.: Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006, Tellus B: Chemical and Physical Meteorology, 61, 165–179, https://doi.org/10.1111/j.1600-0889.2008.00396.x, 2009.

Freudenthaler, V., Linné, H., Chaikovski, A., Rabus, D., and Groß, S.: EARLINET lidar quality assurance tools, Aerosols/Remote Sensing/Instruments and Platforms, https://doi.org/10.5194/amt-2017-395, 2018.

Gasteiger, J. and Freudenthaler, V.: Benefit of depolarization ratio at λ = 1064 nm for the retrieval of the aerosol microphysics from lidar measurements, Atmospheric Measurement Techniques, 7, 3773–3781, https://doi.org/10.5194/amt-7-3773-2014, 2014.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., Wargan, K., Coy, L., Cullather, R., Draper, C., Akella, S., Buchard, V., Conaty, A., Silva, A. M. da, Gu, W., Kim, G.-K., Koster, R., Lucchesi, R., Merkova, D., Nielsen, J. E., Partyka, G., Pawson, S., Putman, W., Rienecker, M., Schubert,

S. D., Sienkiewicz, M., and Zhao, B.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2), Journal of Climate, 30, 5419–5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Ginoux, P., Prospero, J. M., Gill, T. E., Hsu, N. C., and Zhao, M.: Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products: ANTHROPOGENIC AND NATURAL DUST SOURCES, Rev. Geophys., 50, https://doi.org/10.1029/2012RG000388, 2012.

Gomes, L., Bergametti, G., Coudé-Gaussen, G., and Rognon, P.: Submicron desert dusts: A sandblasting process, Journal of Geophysical Research: Atmospheres, 95, 13927–13935, https://doi.org/10.1029/JD095iD09p13927, 1990.

Graaf, M., Donovan, D., and Apituley, A.: Aerosol microphysical properties from inversion of tropospheric optical Raman lidar data, 2009.

Granados-Muñoz, M. J., Bravo-Aranda, J. A., Baumgardner, D., Guerrero-Rascado, J. L., Pérez-Ramírez, D., Navas-Guzmán, F., Veselovskii, I., Lyamani, H., Valenzuela, A., Olmo, F. J., Titos, G., Andrey, J., Chaikovsky, A., Dubovik, O., Gil-Ojeda, M., and Alados-Arboledas, L.: A comparative study of aerosol microphysical properties retrieved from ground-based remote sensing and aircraft in situ measurements during a Saharan dust event, Atmospheric Measurement Techniques, 9, 1113–1133, https://doi.org/10.5194/amt-9-1113-2016, 2016.

Gryspeerdt, E., Quaas, J., Ferrachat, S., Gettelman, A., Ghan, S., Lohmann, U., Morrison, H., Neubauer, D., Partridge, D. G., Stier, P., Takemura, T., Wang, H., Wang, M., and Zhang, K.: Constraining the instantaneous aerosol influence on cloud albedo, Proc. Natl. Acad. Sci. U.S.A., 114, 4899–4904, https://doi.org/10.1073/pnas.1617765114, 2017.

Gupta, P. and Christopher, S. A.: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach, J. Geophys. Res., 114, D20205, https://doi.org/10.1029/2008JD011497, 2009.

Haarig, M., Ansmann, A., Engelmann, R., Baars, H., Toledano, C., Torres, B., Althausen, D., Radenz, M., and Wandinger, U.: First triple-wavelength lidar observations of depolarization and extinction-to-backscatter ratios of Saharan dust, Atmospheric Chemistry and Physics, 22, 355–369, https://doi.org/10.5194/acp-22-355-2022, 2022.

Hansen, J., Nazarenko, L., Ruedy, R., Sato, M., Willis, J., Del Genio, A., Koch, D., Lacis, A., Lo, K., Menon, S., Novakov, T., Perlwitz, J., Russell, G., Schmidt, G. A., and Tausnev, N.: Earth’s Energy Imbalance: Confirmation and Implications, Science, 308, 1431–1435, https://doi.org/10.1126/science.1110252, 2005.

Hansen, J. E. and Travis, L. D.: Light scattering in planetary atmospheres, Space Sci Rev, 16, 527–610, https://doi.org/10.1007/BF00168069, 1974.

Harris, D. C. and Bertolucci, M. D.: Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy, Dover Publications, New York, 550 pp., 1989.

Hess, M., Koepke, P., and Schult, I.: Optical Properties of Aerosols and Clouds: The Software Package OPAC, Bulletin of the American Meteorological Society, 79, 831–844, https://doi.org/10.1175/1520-0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

Hinds, W. C. and Zhu, Y.: Aerosol Technology - Properties, Behavior and Measurements of Airnorne Particles., Third., John Wiley & Sons, Inc., 448 pp., 2022.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization, Remote Sensing of Environment, 66, 1–16, https://doi.org/10.1016/S0034-4257(98)00031-5, 1998.

Hu, Q.: Advanced aerosol characterization using sun/sky photometer and multi-wavelength Mie-Raman lidar measurements, These de doctorat, Université de Lille (2018-2021), 2018.

Hu, Q., Goloub, P., Veselovskii, I., Bravo-Aranda, J.-A., Popovici, I. E., Podvin, T., Haeffelin, M., Lopatin, A., Dubovik, O., Pietras, C., Huang, X., Torres, B., and Chen, C.: Long-range-transported Canadian smoke plumes in the lower stratosphere over northern France, Atmospheric Chemistry and Physics, 19, 1173–1193, https://doi.org/10.5194/acp-19-1173-2019, 2019.

Hu, Q., Wang, H., Goloub, P., Li, Z., Veselovskii, I., Podvin, T., Li, K., and Korenskiy, M.: The characterization of Taklamakan dust properties using a multiwavelength Raman polarization lidar in Kashi, China, Atmospheric Chemistry and Physics, 20, 13817–13834, https://doi.org/10.5194/acp-20-13817-2020, 2020.

Hu, Q., Goloub, P., Veselovskii, I., and Podvin, T.: The characterization of long-range transported North American biomass burning plumes: what can a multi-wavelength Mie–Raman-polarization-fluorescence lidar provide?, Atmospheric Chemistry and Physics, 22, 5399–5414, https://doi.org/10.5194/acp-22-5399-2022, 2022.

van de Hulst, H. C.: Light Scattering by Small Particles., John Wiley and Sons, New York, 1957.

Johnson, B. R.: Invariant imbedding T matrix approach to electromagnetic scattering, Appl. Opt., 27, 4861, https://doi.org/10.1364/AO.27.004861, 1988.

Junge, C.: THE SIZE DISTRIBUTION AND AGING OF NATURAL AEROSOLS AS DETERMINED FROM ELECTRICAL AND OPTICAL DATA ON THE ATMOSPHERE, J. Meteor., 12, 13–25, https://doi.org/10.1175/1520-0469(1955)012<0013:TSDAAO>2.0.CO;2, 1955.

Kaufman, Y. J., Tanre, D., Leon, J.-F., and Pelon, J.: Retrievals of profiles of fine and coarse aerosols using lidar and radiometric space measurements, IEEE Transactions on Geoscience and Remote Sensing, 41, 1743–1754, https://doi.org/10.1109/TGRS.2003.814138, 2003.

Khaykin, S., Legras, B., Bucci, S., Sellitto, P., Isaksen, L., Tencé, F., Bekki, S., Bourassa, A., Rieger, L., Zawada, D., Jumelet, J., and Godin-Beekmann, S.: The 2019/20 Australian wildfires generated a persistent smoke-charged vortex rising up to 35 km altitude, Commun Earth Environ, 1, 1–12, https://doi.org/10.1038/s43247-020-00022-5, 2020.

Klett, J. D.: Lidar inversion with variable backscatter/extinction ratios, Appl. Opt., AO, 24, 1638–1643, https://doi.org/10.1364/AO.24.001638, 1985.

Kok, J. F.: A scaling theory for the size distribution of emitted dust aerosols suggests climate models underestimate the size of the global dust cycle, Proceedings of the National Academy of Sciences, 108, 1016–1021, https://doi.org/10.1073/pnas.1014798108, 2011.

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 2: simulations with synthetic optical data, Appl. Opt., 55, 9850–9865, https://doi.org/10.1364/AO.55.009850, 2016.

Kolgotin, A., Müller, D., Chemyakin, E., Romanov, A., and Alehnovich, V.: Improved identification of the solution space of aerosol microphysical properties derived from the inversion of profiles of lidar optical data, part 3: case studies, Appl. Opt., 57, 2499–2513, https://doi.org/10.1364/AO.57.002499, 2018.

Kreidenweis, S. M., Remer, L. A., Bruintjes, R., and Dubovik, O.: Smoke aerosol from biomass burning in Mexico: Hygroscopic smoke optical model, Journal of Geophysical Research: Atmospheres, 106, 4831–4844, https://doi.org/10.1029/2000JD900488, 2001.

Lenoble, J., Remer, L. A., and Tanré, D.: Introduction, in: Aerosol Remote Sensing, edited by: Lenoble, J., Remer, L., and Tanre, D., Springer Berlin Heidelberg, Berlin, Heidelberg, 1–11, https://doi.org/10.1007/978-3-642-17725-5_1, 2013.

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: Second-generation operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance, Journal of Geophysical Research: Atmospheres, 112, https://doi.org/10.1029/2006JD007811, 2007.

Li, Z., Goloub, P., Dubovik, O., Blarel, L., Zhang, W., Podvin, T., Sinyuk, A., Sorokin, M., Chen, H., Holben, B., Tanré, D., Canini, M., and Buis, J.-P.: Improvements for ground-based remote sensing of atmospheric aerosol properties by additional polarimetric measurements, Journal of Quantitative Spectroscopy and Radiative Transfer, 110, 1954–1961, https://doi.org/10.1016/j.jqsrt.2009.04.009, 2009.

Li, Z., Hou, W., Hong, J., Zheng, F., Luo, D., Wang, J., Gu, X., and Qiao, Y.: Directional Polarimetric Camera (DPC): Monitoring aerosol spectral optical properties over land from satellite observation, Journal of Quantitative Spectroscopy and Radiative Transfer, 218, 21–37, https://doi.org/10.1016/j.jqsrt.2018.07.003, 2018.

Likens, G. E., Driscoll, C. T., and Buso, D. C.: Long-Term Effects of Acid Rain: Response and Recovery of a Forest Ecosystem, Science, 272, 244–246, https://doi.org/10.1126/science.272.5259.244, 1996.

Liou, K. and Lahore, H.: Laser Sensing of Cloud Composition: A Backscattered Depolarization Technique, Journal of Applied Meteorology and Climatology, 13, 257–263, https://doi.org/10.1175/1520-0450(1974)013<0257:LSOCCA>2.0.CO;2, 1974.

Liou, K.-N.: An introduction to atmospheric radiation, 2nd ed., Academic Press, Amsterdam ; Boston, 583 pp., 2002.

Lopatin, A., Dubovik, O., Chaikovsky, A., Goloub, P., Lapyonok, T., Tanré, D., and Litvinov, P.: Enhancement of aerosol characterization using synergy of lidar and sun-photometer coincident observations: the GARRLiC algorithm, Atmospheric Measurement Techniques, 6, 2065–2088, https://doi.org/10.5194/amt-6-2065-2013, 2013.

Lucht, W., Schaaf, C. B., and Strahler, A. H.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Trans. Geosci. Remote Sensing, 38, 977–998, https://doi.org/10.1109/36.841980, 2000.

Lund, M. T., Myhre, G., Haslerud, A. S., Skeie, R. B., Griesfeller, J., Platt, S. M., Kumar, R., Myhre, C. L., and Schulz, M.: Concentrations and radiative forcing of anthropogenic aerosols from 1750 to 2014 simulated with the Oslo CTM3 and CEDS emission inventory, Geosci. Model Dev., 11, 4909–4931, https://doi.org/10.5194/gmd-11-4909-2018, 2018.

Ma, X., Yu, F., and Quaas, J.: Reassessment of satellite-based estimate of aerosol climate forcing: SATELLITE-BASED AEROSOL FORCING, J. Geophys. Res. Atmos., 119, 10,394-10,409, https://doi.org/10.1002/2014JD021670, 2014.

Mahrt, F., Marcolli, C., David, R. O., Grönquist, P., Barthazy Meier, E. J., Lohmann, U., and Kanji, Z. A.: Ice nucleation abilities of soot particles determined with the Horizontal Ice Nucleation Chamber, Atmos. Chem. Phys., 18, 13363–13392, https://doi.org/10.5194/acp-18-13363-2018, 2018.

Mamouri, R. E. and Ansmann, A.: Fine and coarse dust separation with polarization lidar, Atmospheric Measurement Techniques, 7, 3717–3735, https://doi.org/10.5194/amt-7-3717-2014, 2014.

Marbach, T., Riedi, J., Lacan, A., and Schlüssel, P.: The 3MI mission: multi-viewing-channel-polarisation imager of the EUMETSAT polar system: second generation (EPS-SG) dedicated to aerosol and cloud monitoring, SPIE Optical Engineering + Applications, San Diego, California, United States, 961310, https://doi.org/10.1117/12.2186978, 2015.

Maring, H., Savoie, D. L., Izaguirre, M. A., Custals, L., and Reid, J. S.: Mineral dust aerosol size distribution change during atmospheric transport, Journal of Geophysical Research: Atmospheres, 108, https://doi.org/10.1029/2002JD002536, 2003.

Marquardt, D. W.: An Algorithm for Least-Squares Estimation of Nonlinear Parameters, Journal of the Society for Industrial and Applied Mathematics, 11, 431–441, 1963.

McCormick, M. P., Hamill, P., Pepin, T. J., Chu, W. P., Swissler, T. J., and McMaster, L. R.: Satellite Studies of the Stratospheric Aerosol, Bulletin of the American Meteorological Society, 60, 1038–1046, 1979.

Mishchenko, M. I., Hovenier, J. W., and Travis, L. D.: Concepts, Terms, Notation, in: Light scattering by nonspherical particles: theory, measurements, and applications, Academic Press, San Diego, 690, 2000.

Mishchenko, M. I., Travis, L. D., and Lacis, A. A.: Scattering, Absorption, and Emission of Light by Small Particles, Cambridge University Press, Cambridge, 2002.

Moré, J. J.: The Levenberg-Marquardt algorithm: Implementation and theory, in: Numerical Analysis, Berlin, Heidelberg, 105–116, https://doi.org/10.1007/BFb0067700, 1978.

Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory, Appl. Opt., 38, 2346–2357, https://doi.org/10.1364/AO.38.002346, 1999.

Müller, D., Mattis, I., Wandinger, U., Ansmann, A., Althausen, D., and Stohl, A.: Raman lidar observations of aged Siberian and Canadian forest fire smoke in the free troposphere over Germany in 2003: Microphysical particle characterization, Journal of Geophysical Research: Atmospheres, 110, https://doi.org/10.1029/2004JD005756, 2005.

Müller, D., Veselovskii, I., Kolgotin, A., Tesche, M., Ansmann, A., and Dubovik, O.: Vertical profiles of pure dust and mixed smoke–dust plumes inferred from inversion of multiwavelength Raman/polarization lidar data and comparison to AERONET retrievals and in situ observations, Appl. Opt., 52, 3178–3202, https://doi.org/10.1364/AO.52.003178, 2013.

Müller, D., Hostetler, C. A., Ferrare, R. A., Burton, S. P., Chemyakin, E., Kolgotin, A., Hair, J. W., Cook, A. L., Harper, D. B., Rogers, R. R., Hare, R. W., Cleckner, C. S., Obland, M. D., Tomlinson, J., Berg, L. K., and Schmid, B.: Airborne Multiwavelength High Spectral Resolution Lidar (HSRL-2) observations during TCAP 2012: vertical profiles of optical and microphysical properties of a smoke/urban haze plume over the northeastern coast of the US, Atmospheric Measurement Techniques, 7, 3487–3496, https://doi.org/10.5194/amt-7-3487-2014, 2014.

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, https://doi.org/10.5194/amt-9-5007-2016, 2016.

Müller, D., Chemyakin, E., Kolgotin, A., Ferrare, R. A., Hostetler, C. A., and Romanov, A.: Automated, unsupervised inversion of multiwavelength lidar data with TiARA: assessment of retrieval performance of microphysical parameters using simulated data, Appl. Opt., 58, 4981–5008, https://doi.org/10.1364/AO.58.004981, 2019.

Nepomuceno Pereira, S., Preißler, J., Guerrero-Rascado, J. L., Silva, A. M., and Wagner, F.: Forest Fire Smoke Layers Observed in the Free Troposphere over Portugal with a Multiwavelength Raman Lidar: Optical and Microphysical Properties, The Scientific World Journal, 2014, 1–11, https://doi.org/10.1155/2014/421838, 2014.

Nicolae, D., Vasilescu, J., Talianu, C., Binietoglou, I., Nicolae, V., Andrei, S., and Antonescu, B.: A neural network aerosol-typing algorithm based on lidar data, Atmospheric Chemistry and Physics, 18, 14511–14537, https://doi.org/10.5194/acp-18-14511-2018, 2018.

Nielsen, M. A.: Neural Networks and Deep Lerning, Determination Press, 2015.

Orlandini, S. and Morlini, I.: Artificial neural network estimation of rainfall intensity from radar observations, Journal of Geophysical Research: Atmospheres, 105, 24849–24861, https://doi.org/10.1029/2000JD900408, 2000.

Osborne, S. R., Johnson, B. T., Haywood, J. M., Baran, A. J., Harrison, M. a. J., and McConnell, C. L.: Physical and optical properties of mineral dust aerosol during the Dust and Biomass-burning Experiment, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009551, 2008.

Ott, W. R.: A Physical Explanation of the Lognormality of Pollutant Concentrations, Journal of the Air & Waste Management Association, 40, 1378–1383, https://doi.org/10.1080/10473289.1990.10466789, 1990.

Pan, Y.-L.: Detection and characterization of biological and other organic-carbon aerosol particles in atmosphere using fluorescence, Journal of Quantitative Spectroscopy and Radiative Transfer, 150, 12–35, https://doi.org/10.1016/j.jqsrt.2014.06.007, 2015.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmospheric Measurement Techniques, 7, 2389–2409, https://doi.org/10.5194/amt-7-2389-2014, 2014.

Piironen, P. and Eloranta, E. W.: Demonstration of a high-spectral-resolution lidar based on an iodine absorption filter, Opt. Lett., 19, 234–236, https://doi.org/10.1364/OL.19.000234, 1994.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical recipes in C. Book: William H. Press, 2. ed., repr., Cambridge Univ. Pr, Cambridge, 994 pp., 2002.

Rault, D. F. and Loughman, R. P.: The OMPS Limb Profiler Environmental Data Record Algorithm Theoretical Basis Document and Expected Performance, IEEE Transactions on Geoscience and Remote Sensing, 51, 2505–2527, https://doi.org/10.1109/TGRS.2012.2213093, 2013.

Reid, J. S., Koppmann, R., Eck, T. F., and Eleuterio, D. P.: A review of biomass burning emissions part II: intensive physical properties of biomass burning particles, Atmospheric Chemistry and Physics, 5, 799–825, https://doi.org/10.5194/acp-5-799-2005, 2005.

Reid, J. S., Reid, E. A., Walker, A., Piketh, S., Cliff, S., Al Mandoos, A., Tsay, S.-C., and Eck, T. F.: Dynamics of southwest Asian dust particle size characteristics with implications for global dust research, Journal of Geophysical Research: Atmospheres, 113, https://doi.org/10.1029/2007JD009752, 2008.

Remer, L. A., Kaufman, Y. J., Tanré, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R.-R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS Aerosol Algorithm, Products, and Validation, Journal of the Atmospheric Sciences, 62, 947–973, https://doi.org/10.1175/JAS3385.1, 2005.

Roberts, J. B., Clayson, C. A., Robertson, F. R., and Jackson, D. L.: Predicting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach, J. Geophys. Res., 115, 2009JD013099, https://doi.org/10.1029/2009JD013099, 2010.

Rodgers, C. D.: Inverse methods for atmospheric sounding: theory and practice, World Scientific, Singapore ; [River Edge, N.J.], 238 pp., 2000.

Roig Rodelas, R., Perdrix, E., Herbin, B., and Riffault, V.: Characterization and variability of inorganic aerosols and their gaseous precursors at a suburban site in northern France over one year (2015–2016), Atmospheric Environment, 200, 142–157, https://doi.org/10.1016/j.atmosenv.2018.11.041, 2019.

Saito, M., Yang, P., Ding, J., and Liu, X.: A Comprehensive Database of the Optical Properties of Irregular Aerosol Particles for Radiative Transfer Simulations, Journal of the Atmospheric Sciences, 78, 2089–2111, https://doi.org/10.1175/JAS-D-20-0338.1, 2021.

Sassen, K.: Lidar Backscatter Depolarization Technique for Cloud and Aerosol Research, in: Light scattering by nonspherical particles: theory, measurements, and applications, Academic Press, San Diego, 690, 2000.

Schoeberl, M. R.: The afternoon constellation: a formation of Earth observing systems for the atmosphere and hydrosphere, in: IEEE International Geoscience and Remote Sensing Symposium, IEEE International Geoscience and Remote Sensing Symposium. IGARSS 2002, Toronto, Ont., Canada, 354–356, https://doi.org/10.1109/IGARSS.2002.1025038, 2002.

Schoennagel, T., Balch, J. K., Brenkert-Smith, H., Dennison, P. E., Harvey, B. J., Krawchuk, M. A., Mietkiewicz, N., Morgan, P., Moritz, M. A., Rasker, R., Turner, M. G., and Whitlock, C.: Adapt to more wildfire in western North American forests as climate changes, Proc. Natl. Acad. Sci. U.S.A., 114, 4582–4590, https://doi.org/10.1073/pnas.1617464114, 2017.

Seifert, A., Heus, T., Pincus, R., and Stevens, B.: Large-eddy simulation of the transient and near-equilibrium behavior of precipitating shallow convection, J Adv Model Earth Syst, 7, 1918–1937, https://doi.org/10.1002/2015MS000489, 2015.

Shao, Y.: Physics and modelling of wind erosion, 2. rev. & exp. ed., Springer, S.l., 452 pp., 2008.

Sherwood, S. C., Bony, S., Boucher, O., Bretherton, C., Forster, P. M., Gregory, J. M., and Stevens, B.: Adjustments in the Forcing-Feedback Framework for Understanding Climate Change, Bulletin of the American Meteorological Society, 96, 217–228, https://doi.org/10.1175/BAMS-D-13-00167.1, 2015.

Smirnov, A., Holben, B. N., Kaufman, Y. J., Dubovik, O., Eck, T. F., Slutsker, I., Pietras, C., and Halthore, R. N.: Optical Properties of Atmospheric Aerosol in Maritime Environments, Journal of the Atmospheric Sciences, 59, 501–523, https://doi.org/10.1175/1520-0469(2002)059<0501:OPOAAI>2.0.CO;2, 2002.

Smith, V. H. and Schindler, D. W.: Eutrophication science: where do we go from here?, Trends in Ecology & Evolution, 24, 201–207, https://doi.org/10.1016/j.tree.2008.11.009, 2009.

Spanu, A., Dollner, M., Gasteiger, J., Bui, T. P., and Weinzierl, B.: Flow-induced errors in airborne in situ measurements of aerosols and clouds, Atmos. Meas. Tech., 13, 1963–1987, https://doi.org/10.5194/amt-13-1963-2020, 2020.

Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA’s HYSPLIT Atmospheric Transport and Dispersion Modeling System, Bulletin of the American Meteorological Society, 96, 2059–2077, https://doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Tanré, D., Bréon, F. M., Deuzé, J. L., Dubovik, O., Ducos, F., François, P., Goloub, P., Herman, M., Lifermann, A., and Waquet, F.: Remote sensing of aerosols by using polarized, directional and spectral measurements within the A-Train: the PARASOL mission, Atmospheric Measurement Techniques, 4, 1383–1395, https://doi.org/10.5194/amt-4-1383-2011, 2011.

Tarantola, A.: Inverse problem theory and methods for model parameter estimation, SIAM, Society for Industrial and Applied Mathematics, Philadelphia, 342 pp., 2005.

Tesche, M., Ansmann, A., Müller, D., Althausen, D., Engelmann, R., Freudenthaler, V., and Groß, S.: Vertically resolved separation of dust and smoke over Cape Verde using multiwavelength Raman and polarization lidars during Saharan Mineral Dust Experiment 2008, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2009JD011862, 2009.

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, https://doi.org/10.5194/amt-12-4421-2019, 2019.

Twomey, S.: The nuclei of natural cloud formation part II: The supersaturation in natural clouds and the variation of cloud droplet concentration, Geofisica Pura e Applicata, 43, 243–249, https://doi.org/10.1007/BF01993560, 1959.

Twomey, S.: Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurements, Elsevier Sci, New York, 243 pp., 1977.

Ullrich, R., Hoose, C., Möhler, O., Niemand, M., Wagner, R., Höhler, K., Hiranuma, N., Saathoff, H., and Leisner, T.: A New Ice Nucleation Active Site Parameterization for Desert Dust and Soot, Journal of the Atmospheric Sciences, 74, 699–717, https://doi.org/10.1175/JAS-D-16-0074.1, 2017.

Velazquez-Garcia, A., Crumeyrolle, S., de Brito, J. F., Tison, E., Bourrianne, E., Chiapello, I., and Riffault, V.: Deriving composition-dependent aerosol absorption, scattering and extinction mass efficiencies from multi-annual high time resolution observations in Northern France, Atmospheric Environment, 298, 119613, https://doi.org/10.1016/j.atmosenv.2023.119613, 2023.

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, Appl. Opt., 41, 3685–3699, https://doi.org/10.1364/AO.41.003685, 2002.

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, Appl. Opt., 43, 1180–1195, https://doi.org/10.1364/AO.43.001180, 2004.

Veselovskii, I., Dubovik, O., Kolgotin, A., Lapyonok, T., Di Girolamo, P., Summa, D., Whiteman, D. N., Mishchenko, M., and Tanré, D.: Application of randomly oriented spheroids for retrieval of dust particle parameters from multiwavelength lidar measurements, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2010JD014139, 2010.

Veselovskii, I., Dubovik, O., Kolgotin, A., Korenskiy, M., Whiteman, D. N., Allakhverdiev, K., and Huseyinoglu, F.: Linear estimation of particle bulk parameters from multi-wavelength lidar measurements, Atmospheric Measurement Techniques, 5, 1135–1145, https://doi.org/10.5194/amt-5-1135-2012, 2012.

Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Korenskiy, M., Pujol, O., Dubovik, O., and Lopatin, A.: Combined use of Mie–Raman and fluorescence lidar observations for improving aerosol characterization: feasibility experiment, Atmospheric Measurement Techniques, 13, 6691–6701, https://doi.org/10.5194/amt-13-6691-2020, 2020.

Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Choël, M., Visez, N., and Korenskiy, M.: Mie–Raman–fluorescence lidar observations of aerosols during pollen season in the north of France, Atmospheric Measurement Techniques, 14, 4773–4786, https://doi.org/10.5194/amt-14-4773-2021, 2021.

Veselovskii, I., Hu, Q., Goloub, P., Podvin, T., Barchunov, B., and Korenskii, M.: Combining Mie–Raman and fluorescence observations: a step forward in aerosol classification with lidar technology, Atmospheric Measurement Techniques, 15, 4881–4900, https://doi.org/10.5194/amt-15-4881-2022, 2022.

Volten, H., Muñoz, O., Rol, E., de Haan, J. F., Vassen, W., Hovenier, J. W., Muinonen, K., and Nousiainen, T.: Scattering matrices of mineral aerosol particles at 441.6 nm and 632.8 nm, Journal of Geophysical Research: Atmospheres, 106, 17375–17401, https://doi.org/10.1029/2001JD900068, 2001.

Wandinger, U., Müller, D., Böckmann, C., Althausen, D., Matthias, V., Bösenberg, J., Weiß, V., Fiebig, M., Wendisch, M., Stohl, A., and Ansmann, A.: Optical and microphysical characterization of biomass- burning and industrial-pollution aerosols from- multiwavelength lidar and aircraft measurements, Journal of Geophysical Research: Atmospheres, 107, LAC 7-1-LAC 7-20, https://doi.org/10.1029/2000JD000202, 2002.

Wandinger, U., Freudenthaler, V., Baars, H., Amodeo, A., Engelmann, R., Mattis, I., Groß, S., Pappalardo, G., Giunta, A., D'Amico, G., Chaikovsky, A., Osipenko, F., Slesar, A., Nicolae, D., Belegante, L., Talianu, C., Serikov, I., Linné, H., Jansen, F., Apituley, A., Wilson, K. M., De Graaf, M., Trickl, T., Giehl, H., Adam, M., Comerón, A., Muñoz-Porcar, C., Rocadenbosch, F., Sicard, M., Tomás, S., Lange, D., Kumar, D., Pujadas, M., Molero, F., Fernández, A. J., Alados-Arboledas, L., Bravo-Aranda, J. A., Navas-Guzmán, F., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Preißler, J., Wagner, F., Gausa, M., Grigorov, I., Stoyanov, D., Iarlori, M., Rizi, V., Spinelli, N., Boselli, A., Wang, X., Lo Feudo, T., Perrone, M. R., De Tomasi, F., and Burlizzi, P.: EARLINET instrument intercomparison campaigns: overview on strategy and results, Atmos. Meas. Tech., 9, 1001–1023, https://doi.org/10.5194/amt-9-1001-2016, 2016.

Wandinger, U., De Mazière, M., Goloub, P., Mihalopoulos, N., Mona, L., O'Connor, E., Reimann, S., and van Pinxteren, D.: Deliverable D5.1: Documentation on technical and requirements for ACTRIS Observations Platforms, 2018.

Weinzierl, B., Petzold, A., Esselborn, M., Wirth, M., Rasp, K., Kandler, K., Schütz, L., Koepke, P., and Fiebig, M.: Airborne measurements of dust layer properties, particle size distribution and mixing state of Saharan dust during SAMUM 2006, Tellus B, 61, 96–117, https://doi.org/10.1111/j.1600-0889.2008.00392.x, 2009.

Whitby, K. T.: THE PHYSICAL CHARACTERISTICS OF SULFUR AEROSOLS, in: Sulfur in the Atmosphere, edited by: Husar, R. B., Lodge, J. P., and Moore, D. J., Pergamon, 135–159, https://doi.org/10.1016/B978-0-08-022932-4.50018-5, 1978.

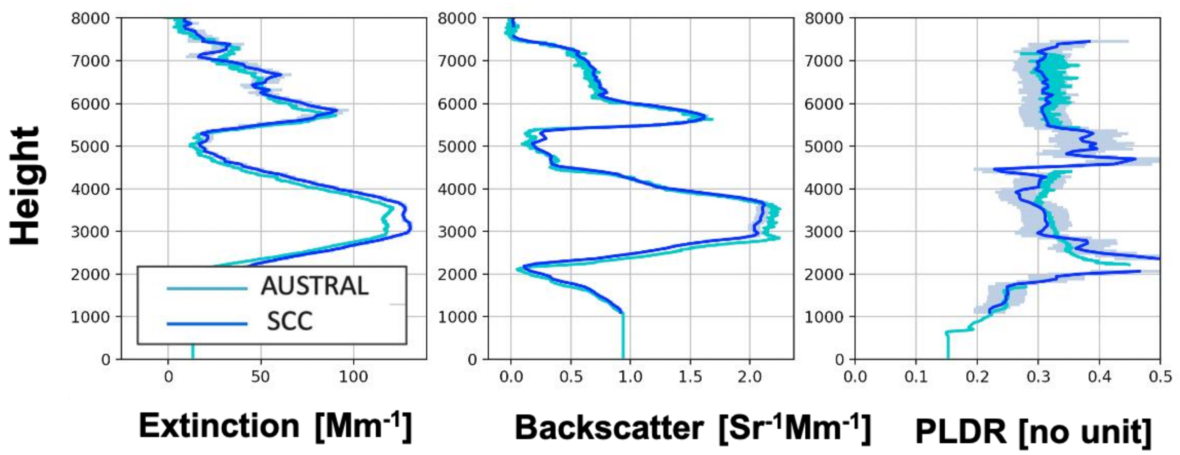
Yagola, A. G.: Ill-Posed Problems and Methods for Their Numerical Solution, in: Optimization and regularization for computational inverse problems and applications, Springer Verlag, Heidelberg, 350, 2010.

Yamasoe, M. A., Kaufman, Y. J., Dubovik, O., Remer, L. A., Holben, B. N., and Artaxo, P.: Retrieval of the real part of the refractive index of smoke particles from Sun/sky measurements during SCAR-B, Journal of Geophysical Research: Atmospheres, 103, 31893–31902, https://doi.org/10.1029/98JD01211, 1998.

Yang, P. and Liou, K. N.: Geometric-optics–integral-equation method for light scattering by nonspherical ice crystals, Appl. Opt., 35, 6568–6584, https://doi.org/10.1364/AO.35.006568, 1996.

Appendix A

Comparison between AUSTRAL and SCC. Extinction, backscattering coefficients and particle linear depolarization ratio (PLDR) at 532 nm measured by LILAS and processed by AUSTRAL and SCC, for time period 21:00-22:00 UTC, March 2, 2021, at ATOLL/Lille (Hu et al., report in ACTRIS-France, 2022).



Appendix B

Article published on *Remote Sensing*

remote sensing

MDPI

Article

Retrieval of Aerosol Microphysical Properties from Multi-Wavelength Mie–Raman Lidar Using Maximum Likelihood Estimation: Algorithm, Performance, and Application

Yuyang Chang [1,*], Qiaoyun Hu [1], Philippe Goloub [1], Igor Veselovskii [2] and Thierry Podvin [1]

1 UMR 8518–LOA–Laboratoire d'Optique Atmosphérique, University of Lille, CNRS, 59650 Villeneuve d'Ascq, France
2 Prokhorov General Physics Institute, Russian Academy of Sciences, Moscow 119991, Russia
* Correspondence: yuyang.chang@univ-lille.fr

Abstract: Lidar plays an essential role in monitoring the vertical variation of atmospheric aerosols. However, due to the limited information that lidar measurements provide, ill-posedness still remains a big challenge in quantitative lidar remote sensing. In this study, we describe the Basic algOrithm for REtrieval of Aerosol with Lidar (BOREAL), which is based on maximum likelihood estimation (MLE), and retrieve aerosol microphysical properties from extinction and backscattering measurements of multi-wavelength Mie–Raman lidar systems. The algorithm utilizes different types of a priori constraints to better constrain the solution space and suppress the influence of the ill-posedness. Sensitivity test demonstrates that BOREAL could retrieve particle volume size distribution (VSD), total volume concentration (V_t), effective radius (R_{eff}), and complex refractive index (CRI = $n - ik$) of simulated aerosol models with satisfying accuracy. The application of the algorithm to real aerosol events measured by LIlle Lidar AtmosphereS (LILAS) shows it is able to realize fast and reliable retrievals of different aerosol scenarios (dust, aged-transported smoke, and urban aerosols) with almost uniform and simple pre-settings. Furthermore, the algorithmic principle allows BOREAL to incorporate measurements with different and non-linearly related errors to the retrieved parameters, which makes it a flexible and generalized algorithm for lidar retrieval.

Keywords: maximum likelihood estimation; retrieval of height-resolved aerosol microphysical properties; analysis of lidar measurements

check for updates

Citation: Chang, Y.; Hu, Q.; Goloub, P.; Veselovskii, I.; Podvin, T. Retrieval of Aerosol Microphysical Properties from Multi-Wavelength Mie–Raman Lidar Using Maximum Likelihood Estimation: Algorithm, Performance, and Application. *Remote Sens.* **2022**, *14*, 6208. https://doi.org/10.3390/rs14246208

Academic Editor: Alain Miffre

Received: 31 August 2022
Accepted: 5 December 2022
Published: 7 December 2022

1. Introduction

Atmospheric aerosols play a significant role in the Earth's climate change and radiative forcing. They can not only change the scattering and absorption of incident solar irradiance, but also affect the formation and optical properties of clouds through aerosol-cloud interactions [1]. The tempo–spatial variation of aerosols properties and sources makes up a dominant source of uncertainty for the assessment of the Earth's radiative forcing and the global temperature projection, although important progress has been made since the last decade thanks to the progress in both atmospheric modeling and observation systems [2].

Remote sensing is an effective way to continuously monitor the temporal and spatial distributions of aerosols and access their microphysical properties, such as particle volume size distribution (VSD), complex refractive index (CRI = $n - ik$), morphologic parameters, and so on. Passive remote sensing, like regional global surface networks [3] or space-borne instruments [4,5], is capable of providing long-term aerosol monitoring with global coverage, whereas it cannot derive height-resolved aerosol properties, which are important for accurately assessing aerosol radiative forcing [6]. In this context, light detection and ranging (lidar) technology has been widely used for atmospheric remote sensing [7].

However, at the early stage of lidar remote sensing, due to relatively high measurement uncertainty [8], the retrieval of aerosol microphysical properties was usually conducted under the assumption of the Junge VSD or known CRI, which limits the application to real measurements [9–11].

More general retrieval methods were proposed since the 1990s when the technique of Raman lidar [12,13] or High Spectral Resolution Lidar (HSRL) [14], which is capable of simultaneously measuring the extinction coefficient (α) and backscattering coefficient (β) with enough accuracy, was developed. Based on such an instrumental leap, on the one hand, lidar networks on a continental scale—such as the European Aerosol Research Lidar Network under the framework of the Aerosol, Clouds, and Trace gases Research InfraStructure (EARLINET/ACTRIS), Micro Pulse Lidar Network (MPLNET), and Asian Dust Lidar Network (ADNET)—have been established since 2000 to extend the spatial coverage of ground-based lidar observations [15–17]. On the other hand, a number of algorithms aimed at retrieving tropospheric aerosols from lidar measurements have been proposed. For example, a well-known method is to linearly inverse the Fredholm integral composed of VSD and optical kernels for a series of CRIs and certain radius ranges using regularization or principal component analysis. Then, a family of solutions which minimize the so-called *discrepancy* will be selected and averaged [18–23]. Another method is based on Look up Tables (LUTs), such as the arrange and average method [24], which utilizes combined measurements of α, β, and lidar ratio (LR) at several wavelengths.

Previous studies demonstrated that "$3\beta + 2\alpha$", i.e., backscattering coefficients at 355 nm, 532 nm, and 1064 nm and extinction coefficients at 355 nm and 532 nm, is the least lidar measurements to retrieve aerosol microphysical properties [20,25–27]. However, this inversion system is ill-posed because, on the one hand, the measurements are highly interdependent on each other and, on the other hand, the number of retrieval parameters in which we are interested is usually more than the number of the measurements. Chemyakin et al. [28] pointed out that the main difficulty in lidar inversion is the non-uniqueness of the solution. Indeed, such difficulty is faced by some retrieval algorithms, such as the linear regularization method and principal component analysis method, which have to identify the proper solution space from all the "solutions" derived by performing linear inversion at every point in the searching domain composed of all non-linear parameters (e.g., CRI) [18–23]. For example, the minimum discrepancy method [20] could find two "qualified" solutions corresponding to different local minima far from each other. In this circumstance, additional constraints on the searching domain must be applied [29]. However, with the development of more advanced lidar systems, as well as the increasing need of synergy with other types of instruments, more aerosol microphysical properties non-linearly coupled with each other are expected to be retrieved quantitively. As a result, traditional linear retrieval algorithms will suffer from both increase of computational burden and algorithmic complexity. In this regard, we propose a non-linear retrieval algorithm, BOREAL (Basic algOrithm for REtrieval of Aerosol with Lidar), based on maximum likelihood estimation (MLE) to reduce the ill-posedness of $3\beta + 2\alpha$ and improve the identification of solution space by incorporating a priori constraints from multi sources. Although the statistical optimization strategy used in BOREAL allows flexibility to incorporate different types of measurements, for example, profiles of depolarization ratio (δ) and fluorescence, only $3\beta + 2\alpha$ data are inverted at this preliminary stage. This study will contribute to the development of an automated aerosol retrieval of LIlle Lidar AtmosphereS (LILAS), operated under the frame of ACTRIS/EARLINET [30,31], and other LILAS-like lidar systems.

The following sections of this paper are organized as follows: in Section 2, we demonstrate the principle and implementation of the BOREAL algorithm; in Section 3, we test the algorithm through sensitivity study using synthetic data; in Section 4, to further evaluate BOREAL's performance, it is applied to a set of real aerosol events (dust, aged smoke, and urban aerosols) detected by LILAS during the SHADOW-2 campaign and in operation at the Atmospheric Observatory of LilLe (ATOLL); and Section 5 concludes this paper.

2. BOREAL Algorithm

2.1. Modeling the Problem

The optical data consisting of extinction and backscattering coefficients measured by lidars can be modeled through particle bulk single-scattering properties:

$$\alpha_\lambda = \int_{\ln r_{\min}}^{\ln r_{\max}} \frac{3\sigma_{\text{ext}}(\lambda, n, k, \ln r)}{4\pi r^3} v(\ln r) d\ln r + \varepsilon_{\alpha_\lambda} \tag{1}$$

$$\beta_\lambda = \int_{\ln r_{\min}}^{\ln r_{\max}} \frac{3\sigma_{\text{bac}}(\lambda, n, k, \ln r)}{4\pi r^3} v(\ln r) d\ln r + \varepsilon_{\beta_\lambda} \tag{2}$$

where σ_{ext} and σ_{bac} are extinction and backscattering cross sections of a single spherical particle, respectively, functions of wavelength λ, real part of the CRI (n), imaginary part of the CRI (k), and particle radius (r). The particle VSD, $v(\ln r)$, is a continuous function of $\ln r$, and $\varepsilon_{\ldots}$ stands for the error in extinction or backscattering measurements. σ_{ext} and σ_{bac} can be calculated from various scattering models, such as Mishchenko et al. [32] and Yang and Liou [33].

Because of the finite number of measurements, $v(\ln r)$ is approximated by the linear combination of a set of base functions $\{\phi_j(\ln r)\}$:

$$v(\ln r) \approx \sum_{j=1}^{N} v_j \phi_j(\ln r) \tag{3}$$

where v_j is the weight factor of ϕ_j. A smooth function with continuous second derivative can be approximated by a piecewise cubic polynomial, which can be expressed as a linear combination of a B-spline basis [34,35]. On the basis of previous studies and for the sake of simplifying the computation [18,20,36,37], we utilize the B-splines of the first degree as the base functions which have the following definition:

$$\phi_j(\ln r) = \begin{cases} 0, & \ln r \le \ln r_{j-1} \\ \frac{\ln r - \ln r_{j-1}}{\ln r_j - \ln r_{j-1}}, & \ln r_{j-1} < \ln r \le \ln r_j \\ \frac{\ln r_{j+1} - \ln r}{\ln r_{j+1} - \ln r_j}, & \ln r_j < \ln r \le \ln r_{j+1} \\ 0, & \ln r > \ln r_{j+1} \end{cases} \tag{4}$$

where the piecewise nodal grids are logarithmic equidistant and $r_1 = r_{\min}, r_N = r_{\max}$. Correspondingly, v_j is equal to $v(\ln r_j)$. With the increase of the number of B-spline functions, i.e., the increase of N in Equation (3), both approximation accuracy and ill-posedness will increase. To balance the two competing factors, N is set to 8 in this study. We found this to be the smallest value to represent aerosol bimodal size distributions with acceptable accuracy and, at the same time, not to cause too large ill-posedness in the inversion procedure. N = 8 was also adopted by other studies where linear inversion methods were used for the 3β + 2α data [18,21,25,38]. With Equations (3) and (4), Equations (1) and (2) can be written in the vector–matrix form:

$$\mathbf{y}_1 = \mathbf{f}_1(\mathbf{x}) + \boldsymbol{\varepsilon}_1 = \mathbf{K}(n,k)\mathbf{v} + \boldsymbol{\varepsilon}_1 \tag{5}$$

where $\mathbf{y}_1$ is the vector of lidar measurements. For a typical aerosol lidar with a Nd: YAG laser, $\mathbf{y}_1 = [\alpha_{355}, \alpha_{532}, \beta_{355}, \beta_{532}, \beta_{1064}]^T$. $\boldsymbol{\varepsilon}_1$ represents the vector of measurement errors, $\mathbf{v} = [v_1, v_2, \cdots, v_N]^T$ collects the weight factors, and $\mathbf{x} = [\mathbf{v}^T, n, k]^T$. $\mathbf{K}$ is the kernel matrix with the elements

$$\{\mathbf{K}(n,k)\}_{ij} = \int_{\ln r_{j-1}}^{\ln r_{j+1}} \frac{3\sigma_i(n, k, \ln r)}{4\pi r^3} \phi_j(\ln r) d\ln r \tag{6}$$

where i corresponds to the element of $\mathbf{y}$. At current stage, we use the database of Dubovik et al. [37], where the kernel matrices of spherical particles and spheroidal particles with a fixed axis ratio distribution were precalculated, for fast calculation of α and β. Other scattering models for specific non-spherical particles, such as the super-spheroid model and the advanced bulk optical model [39,40], will be implemented in the next step.

According to the definition by Hadamard [41], Equation (5) is ill-posed, as there are typically 5 lidar measurements, but 10 parameters to be retrieved. Since most of realistic size distributions of aerosol particles are smooth functions (with continuous second derivatives), we introduce the following smoothing constraint on VSD:

$$\mathbf{y}_2 = 0 = \mathbf{f}_2(\mathbf{x}) + \varepsilon_2 = \mathbf{H}\mathbf{v} + \varepsilon_2 \tag{7}$$

where $\mathbf{H}$ is the operator to calculate the second-order difference of $\mathbf{v}$. ε_2 acts as a weight factor of the constraint.

To further reduce the ill-posedness, a priori constraints are also applied to the real and imaginary parts of the CRI [42]:

$$\mathbf{y}_3 = n_\mathrm{a} = \mathbf{f}_3(\mathbf{x}) + \varepsilon_3 = n + \varepsilon_{n_\mathrm{a}} \tag{8}$$

$$\mathbf{y}_4 = k_\mathrm{a} = \mathbf{f}_4(\mathbf{x}) + \varepsilon_4 = k + \varepsilon_{k_\mathrm{a}} \tag{9}$$

where the subscript 'a' means the a priori value and $\varepsilon_{\cdots_\mathrm{a}}$ the a priori standard deviation, also acting as weight factors of the corresponding constraints. It has been proved in many studies that the $3\beta + 2\alpha$ measurements do not have enough sensitivity to accurately retrieve the CRI, especially to the imaginary part [20,21,23,43]. The introduction of the a priori constraints on CRI is in fact equivalent to prescribing a reasonable range for the retrieved CRI (centered at n_a, k_a with spread of $\varepsilon_{n_\mathrm{a}}$ and $\varepsilon_{k_\mathrm{a}}$, respectively). This strategy is feasible in most cases because the aerosol type can be determined before the retrieval from lidar observations [44–47] and supplementary information (satellite data, atmospheric transport model, etc.) and type-resolved aerosol CRIs from in situ measurements or multi-angle passive remote sensing [48,49] are available.

For clarity, we rewrite Equations (5) and (7)–(9) into a uniform form:

$$\mathbf{y}_l = \mathbf{f}_l(\mathbf{x}) + \varepsilon_l, (l = 1, 2, 3, 4) \tag{10}$$

where l = 1 represents the equations describing the lidar measurements and l = 2, 3, 4 represent the equations about a priori constraints. If we assume ε_l values are independent of each other and follow the Gaussian probability density function, the likelihood function [50] of the retrieval parameter vector $\mathbf{x}$ can be expressed as

$$L(\mathbf{x}) = \prod_l P(\mathbf{y}_l|\mathbf{x}) = \prod_l \frac{1}{(2\pi)^{n/2}|\mathbf{C}_l|^{1/2}} \exp\left\{-\frac{1}{2}[\mathbf{y}_l - \mathbf{f}_l(\mathbf{x})]^T \mathbf{C}_l^{-1} [\mathbf{y}_l - \mathbf{f}_l(\mathbf{x})]\right\} \tag{11}$$

where $P(\mathbf{y}_l|\mathbf{x})$ represents the conditional probability of $\mathbf{y}_l$, and $\mathbf{C}_l$ is the covariance matrix of ε_l. $|\cdot|$ represents the determinant operator. According to the MLE, the value of $\mathbf{x}$ maximizing the likelihood function is the maximum likelihood estimate of $\mathbf{x}$, which is equivalent to minimizing the following cost function:

$$\chi^2(\mathbf{x}) = \sum_{l=1}^{4} [\mathbf{y}_l - \mathbf{f}_l(\mathbf{x})]^T \mathbf{C}_l^{-1} [\mathbf{y}_l - \mathbf{f}_l(\mathbf{x})] \tag{12}$$

In this way, the search of the retrieval parameter vector $\mathbf{x}$ is converted to an optimal problem. Since negative values of the retrieval parameters do not carry any physical meaning, we implement logarithmic transformation to avoid negative values in the retrieval parameters [36] and rewrite Equation (12) as below:

$$\chi^2(\mathbf{X}) = \sum_{l=1}^{4} [\mathbf{Y}_l - \mathbf{F}_l(\mathbf{X})]^T \mathbf{S}_l^{-1} [\mathbf{Y}_l - \mathbf{F}_l(\mathbf{X})] \tag{13}$$

where $\mathbf{X} = \ln \mathbf{x}, \mathbf{Y}_l = \ln \mathbf{y}_l$, and $\mathbf{F}(\mathbf{X}) = \ln\left[\mathbf{f}_l\left(\mathrm{e}^{\mathbf{X}}\right)\right]$. The measurement variances after the transformation (i.e., the diagonal elements of $\mathbf{S}_l$) are related with their relative variances. For instance, in the term representing the lidar measurements (l = 1):

$$S_i = \ln\left[\frac{1}{2}\left(1 + \sqrt{1 + \frac{4C_i}{y_i^2}}\right)\right] \approx \frac{C_i}{y_i^2}, \quad (C_i \ll 1) \tag{14}$$

Note that by converting Equation (12) to Equation (13), we assume the measurements conform to the multivariate lognormal probability density function. For measurement noise of positively defined characteristics, this assumption is supported by both theoretical analysis and experimental results [51], and for a very small variance, lognormal distribution is almost equivalent to normal distribution.

2.2. Optimization Procedure

The minimization of Equation (13) is in fact a multi-term nonlinear least-square fitting weighted by the corresponding covariance matrices. It is implemented by the Levenberg–Marquardt iteration [52] as below:

$$\begin{aligned} \mathbf{X}^{(u+1)} &= \mathbf{X}^{(u)} + \Delta\mathbf{X}^{(u)}, \\ \Delta\mathbf{X}^{(u)} &= \mathbf{G}_u^{-1}\mathbf{b}_u \end{aligned} \tag{15}$$

where

$$\begin{aligned} \mathbf{G}_u &= \sum_{l=1}^{4} \mathbf{J}_{l,\mathbf{X}^{(u)}}^T \mathbf{S}_l^{-1} \mathbf{J}_{l,\mathbf{X}^{(u)}} + \gamma^{(u)}\mathbf{D}, \\ \mathbf{b}_u &= \sum_{l=1}^{4} \mathbf{J}_{l,\mathbf{X}^{(u)}}^T \mathbf{S}_l^{-1}\left[\mathbf{Y}_l - \mathbf{F}_l\left(\mathbf{X}^{(u)}\right)\right] \end{aligned} \tag{16}$$

and the superscript (u) represents the uth iteration. $\mathbf{J}_{l,\mathbf{X}^{(u)}}$ is the Jacobian matrix of $\mathbf{F}_l$ at $\mathbf{X}^{(u)}$, $\mathbf{D}$ is a scaling matrix controlling the relative iteration steps, and $\gamma^{(u)}$ is the overall scalar factor controlling the speed of the convergence of the iteration process. The value of $\gamma^{(u)}$ should be adjusted in each iteration to ensure the reduction of the cost function and the non-singularity of $\mathbf{G}$. We adopt the following strategy to update $\gamma^{(u)}$ [36]:

$$\gamma^{(u)} = \frac{2\chi^2\left(\mathbf{X}^{(u)}\right)}{p - q} \tag{17}$$

where p and q are the number of total general measurements and the number of retrieval parameters and, in our case, p = 13 and q = 10.

A study of Veselovskii et al. [25] shows that both extinction kernels and backscattering kernels become highly interdependent and asymptotic to zero at large radii, which means the large particles of the size distribution contribute less to the total optical data. Thus, it is quite possible that the iteration can converge to an unrealistic but smooth monotonic VSD curve with large values at large radii, which simultaneously fits all the terms in Equation (13) quite well. We call such a VSD curve 'oversmoothed' and the cost function 'overfitted'. To avoid this issue, we set the stopping conditions to

1. $\chi^2\left(\mathbf{X}^{(u)}\right) < p - q$,
2. the number of iteration u reaches the prescribed maximum value, and the iteration will stop if either of the above conditions is met. Condition 1 is based on the statistical principle. Since we have assumed each $\mathbf{Y}_l$ conforms to a Gaussian distribution, χ^2 conforms to a chi-square distribution with a degree of freedom (DOF) of p–q. A 'good' fit is derived if the ratio of χ^2 and DOF is just not greater than 1 [53].

Setting an initial guess near the solution could accelerate the speed of convergence compared with setting it arbitrarily. The type-resolved a priori value on CRI should not be far from the actual value if the type of the aerosol could be correctly identified before retrieval. Thus, we set the initial guesses of n, k to n_a, k_a, respectively. Correspondingly, the initial guess of VSD is set to a constant function derived by fitting α_{532}.

2.3. The Selection of Individual Solutions

The optimization procedure gives a solution for a specific aerosol size range, i.e., $[r_{\min}, r_{\max}]$, which is called an inversion window hereafter. A solution corresponding to a specific inversion window is referred to as an individual solution. A proper inversion window covering the real aerosol size range is important for deriving a realistic solution. However, such a priori information is hardly available. Therefore, we decide to perform the inversion for a set of pre-defined inversion windows covering the radius range of 0.05–15 μm and then select the qualified individual solutions by some extra constraints. Due to the simultaneous retrieval of the VSD and CRI, there is only one individual solution for an inversion window rather than several hundred derived by linear methods [20,21,23], which simplifies the selection procedure.

Plenty of previous research reveals that, in most cases, the VSD of atmospheric aerosol ensembles conforms to multi-mode lognormal distributions [49,54–57]. Thus, we take this conclusion as an extra a priori constraint on VSD to select proper inversion windows (i.e., qualified individual solutions) because unproper inversion windows (either too wide or too narrow) can cause significant oscillations of the retrieved VSD curve in the wing zones, making it deviate from a 'lognormal' shape. However, such "deformed" curves can have very low fitting error due to the ill-posedness of the system. Thus, in addition to selecting qualified individual solutions by judging their fitting errors, we also select by judging whether the retrieved VSD has a lognormal-like shape. Specifically speaking, the selection procedure consists of the following steps:

1. Select the individual solutions with fitting errors less than the prescribed measurement error (10% for all the measurement channels in this study);
2. Among the selected individual solutions, select those whose elements of **v** meet either of the following inequalities:

$$\begin{cases} v_1 < v_2 \\ v_1 < 0.7 v_{\max} \\ v_8 < v_7 \\ v_8 < 0.7 v_{\max} \end{cases} \text{or} \begin{cases} v_1 > v_2 \\ v_1 < 0.05 v_{\max} \\ v_8 > v_7 \\ v_8 < 0.05 v_{\max} \end{cases} \tag{18}$$

   where $v_{\max}$ means the maximum retrieved element in **v,** and the multiple factors are empirically chosen.
3. Among the selected individual solutions, select those whose standard deviations of the VSD are greater than 0.35. This criterion is based on the study of Tanré et al. [58]. The standard deviation of a distribution v (ln$r$) is calculated by:

$$\sigma_{\rm v} = \sqrt{\frac{\int_{r_{\min}}^{r_{\max}} (\ln r - \ln \mu)^2 v(\ln r) d\ln r}{\int_{r_{\min}}^{r_{\max}} v(\ln r) d\ln r}} \tag{19}$$

   where

$$\mu = \exp\left[\frac{\int_{r_{\min}}^{r_{\max}} \ln r \cdot v(\ln r) d\ln r}{\int_{r_{\min}}^{r_{\max}} v(\ln r) d\ln r}\right] \tag{20}$$

After determining the "qualified" individual solutions, we average them (the average of both retrieved VSDs and retrieved CRIs) to build the final averaged solution, which is regarded as the retrieval result of the case.

In addition, to describe the bulk properties of a particle ensemble, total volume concentration (V_t) and effective radius ($R_{\rm eff}$) can be calculated from the retrieved VSD:

$$V_{\rm t} = \int_{r_{\min}}^{r_{\max}} v(\ln r) d\ln r \tag{21}$$

$$R_{\text{eff}} = \frac{\int_{r_{\min}}^{r_{\max}} v(\ln r) d \ln r}{\int_{r_{\min}}^{r_{\max}} \frac{1}{r} v(\ln r) d \ln r} \tag{22}$$

2.4. Propagation of Measurement Error

In this part, we evaluate the influence of lidar measurement error on individual solutions. According to Equation (15), if the iteration stops at u, we have

$$\hat{\mathbf{X}} = \mathbf{X}^{(u)} = \mathbf{X}^{(u-1)} + \Delta\mathbf{X}^{(u-1)} \tag{23}$$

where $\hat{\mathbf{X}}$ means the retrieved value of $\mathbf{X}$. If the variation of $\hat{\mathbf{X}}$ due to a lidar measurement error $d\mathbf{Y}_1$ can be approximated to be linear, we derive

$$\frac{d\hat{\mathbf{X}}}{d\mathbf{Y}_1} = \frac{\partial \mathbf{X}^{(u)}}{\partial \mathbf{Y}_1} = \frac{\partial \mathbf{X}^{(u-1)}}{\partial \mathbf{Y}_1} + \frac{\partial \Delta\mathbf{X}^{(u-1)}}{\partial \mathbf{Y}_1} \tag{24}$$

According to the rules of nested matrix calculus, we have

$$\begin{aligned} \frac{\partial \Delta\mathbf{X}^{(u-1)}}{\partial \mathbf{Y}_1} &= \left(\mathbf{G}_{u-1}^{-1}\mathbf{b}_{u-1}\right)^T \otimes \left(-\mathbf{G}_{u-1}^{-1}\right) vec(\mathbf{D}) \left(\frac{\partial \gamma^{(u-1)}}{\partial \mathbf{Y}_1}\right)^T \\ &+ \mathbf{G}_{u-1}^{-1}\left[\mathbf{J}_{1,\mathbf{X}^{(u-1)}}^T \mathbf{S}_1^{-1} - \left(\sum_{l=1}^{4} \mathbf{J}_{l,\mathbf{X}^{(u-1)}}^T \mathbf{S}_l^{-1} \mathbf{J}_{l,\mathbf{X}^{(u-1)}}\right) \frac{\partial \mathbf{X}^{(u-1)}}{\partial \mathbf{Y}_1}\right] \end{aligned} \tag{25}$$

where

$$\frac{\partial \gamma^{(u-1)}}{\partial \mathbf{Y}_1} = \frac{4}{3}\left\{\mathbf{S}_1^{-1}\left[\mathbf{Y}_1 - \mathbf{F}_1\left(\mathbf{X}^{(u-1)}\right)\right] - \sum_{l=1}^{4}\left(\mathbf{J}_{l,\mathbf{X}^{(u-1)}} \frac{\partial \mathbf{X}^{(u-1)}}{\partial \mathbf{Y}_1}\right)^T \mathbf{S}_l^{-1}\left[\mathbf{Y}_l - \mathbf{F}_l\left(\mathbf{X}^{(u-1)}\right)\right]\right\} \tag{26}$$

The operator $\otimes$ represents the Kronecker product of two matrices and $vec(\cdot)$ means the vectorization of a matrix [59]. With Equations (25) and (26), we can calculate $d\hat{\mathbf{X}}/d\mathbf{Y}_1$ iteratively and note that $d\mathbf{X}^{(0)}/d\mathbf{Y}_1 = 0$. Correspondingly, the covariance matrix of $\hat{\mathbf{X}}$, denoted as $\hat{\mathbf{S}}$, can be calculated from

$$\hat{\mathbf{S}} = \left(\frac{d\hat{\mathbf{X}}}{d\mathbf{Y}_1}\right) \mathbf{S}_1 \left(\frac{d\hat{\mathbf{X}}}{d\mathbf{Y}_1}\right)^T \tag{27}$$

and since $\hat{\mathbf{x}} = \exp \hat{\mathbf{X}}$, the variation and covariance matrix of $\hat{\mathbf{x}}$, denoted as $d\hat{\mathbf{x}}$ and $\hat{\mathbf{C}}$, respectively, are

$$d\hat{\mathbf{x}} = \exp \hat{\mathbf{X}} d\hat{\mathbf{X}} \tag{28}$$

$$\{\hat{\mathbf{C}}\}_{ij} = E(\hat{x}_i) E(\hat{x}_j) \left[\exp\{\hat{\mathbf{S}}\}_{ij} - 1\right] \tag{29}$$

where $E(\hat{x}_i) = \exp\left(\hat{X}_i + \{\hat{\mathbf{S}}\}_{ii}/2\right)$ is the expectation of the ith element of $\hat{\mathbf{x}}$ and $E(\hat{x}_j)$ the expectation of the jth element. Likewise, the variety and covariance matrices of V_{t} and R_{eff} can be calculated from

$$dI = \frac{dI}{d\hat{\mathbf{x}}} d\hat{\mathbf{x}}, \quad (I = V_{\text{t}}, R_{\text{eff}}) \tag{30}$$

$$\mathbf{C}_I = \left(\frac{dI}{d\hat{\mathbf{x}}}\right) \mathbf{C}_{\hat{\mathbf{x}}} \left(\frac{dI}{d\hat{\mathbf{x}}}\right)^T, \quad (I = V_{\text{t}}, R_{\text{eff}}) \tag{31}$$

We are interested in deriving these above relations because they facilitate both the calculation of retrieval sensitivity in sensitivity study and the calculation of retrieval uncertainty in real application. However, their accuracies depend on the linearity of the system when lidar measurements vary in a range of measurement errors. In the next section, we will examine the feasibility of these relations by numerical simulations.

3. Sensitivity Study

The first part of this section is focused on assessing the performance of the BOREAL algorithm by inverting synthetic optical data generated by different aerosol models. We derive retrieval results for these aerosol models with and without considering measurement error, respectively, and compare them with their original values. In the second part of this section, we evaluate the feasibility of the error propagation model proposed in Section 2.4.

3.1. Data Preparation and Initialization

As indicated in Section 2.3, we use the lognormal distribution to model the VSD of aerosols composed of spherical particles:

$$v(\ln r) = \frac{dV_{\rm t}}{d\ln r} = \sum_{i={\rm f,c}} \frac{dV_i}{\sqrt{2\pi}\sigma_{{\rm v},i}} \exp\left[-\frac{(\ln r - \ln r_{{\rm v},i})^2}{2\sigma_{{\rm v},i}^2}\right] \tag{32}$$

where the subscript i indicates the fine mode (i = f) or coarse mode (i = c). In each mode, V_i represents the volume concentration, $\sigma_{{\rm v},i}$ the geometric standard deviation, and $r_{{\rm v},i}$ the mode radius. $V_{\rm t}$ is the total volume concentration, the same parameter defined by Equation (21).

According to previous characterization of aerosol types [30,49,60–63], we assumed 4 types of VSDs and 25 spectrally independent CRIs, as shown in Table 1. Synthetic optical data (3β + 2α), which are to be inverted with BOREAL, were calculated from these aerosol models with the Mie theory using the databank of Dubovik et al. [37].

Table 1. Aerosol models used for generating synthetic (3β + 2α) data. The definitions of the parameters describing the lognormal VSD can be found in Equation (32). Four VSD types (MF for mono-fine mode, MC for mono-coarse mode, BF for bimodal with fine-mode-dominant, and BC for bimodal with coarse-mode-dominant) and 25 combinations of complex refractive index (CRI = $n - ik$) are prescribed.

SD Type	$V_{\rm f}$	$r_{\rm v,f}$	$\sigma_{\rm v,f}$	$V_{\rm c}$	$r_{\rm v,c}$	$\sigma_{\rm v,c}$	$V_{\rm t}$	$R_{\rm eff}$
MF	1	0.2	0.4	0	0	0	1	0.18
MC	0	0	0	1	1.2	0.6	1	0.99
BF	2/3	0.2	0.4	1/3	2	0.6	1	0.26
BC	1/6	0.2	0.4	5/6	2	0.6	1	0.70
$n_{\rm ture}$	1.4, 1.45, 1.5, 1.55, 1.6							
$k_{\rm true}$	0.001, 0.005, 0.01, 0.015, 0.02							

We use $(n_{\rm a}, \varepsilon_{n_{\rm a}}) = (1.5, 0.1)$ as the a priori constraint on the real part of the CRI for all the cases; $(k_{\rm a}, \varepsilon_{k_{\rm a}}) = (0.005, 0.005)$ for non-absorbing cases, where $k_{\rm true} \leq 0.01$; and $(k_{\rm a}, \varepsilon_{k_{\rm a}}) = (0.015, 0.01)$ for absorbing cases, where $k_{\rm true} > 0.01$. We will also use this configuration for inverting real lidar measurements before an applicable aerosol typing method is developed and a correlated type-resolved database of the a priori constraints is established.

3.2. Evaluation of Retrieval Accuracy

Figure 1 shows the comparisons between the retrieved and true VSDs when $n_{\rm ture} = 1.6$ and $k_{\rm ture} = 0.01$. The left column (Figure 1a1–d1) represents the results when the synthetic optical data were free of error (error-free), while the right column (Figure 1a2–d2) shows the statistics of the results when measurement error is considered (error-contaminated), which is accomplished by adding the error vector to the optical data and inverting the error-contaminated optical data 100 times. The elements of the error vector are independent of each other and conform to the Gaussian distribution: $\sim N(0, 0.1)$. From Figure 1, it is seen that there are larger dispersions in the coarse mode than in the fine mode for both error-free and error-contaminated optical data. This can be explained by the fact that the backscattering kernels decrease rapidly if the particle radius exceeds 2–3 µm [25], which undermines the contribution of the coarse mode to total backscattering when both modes exist.

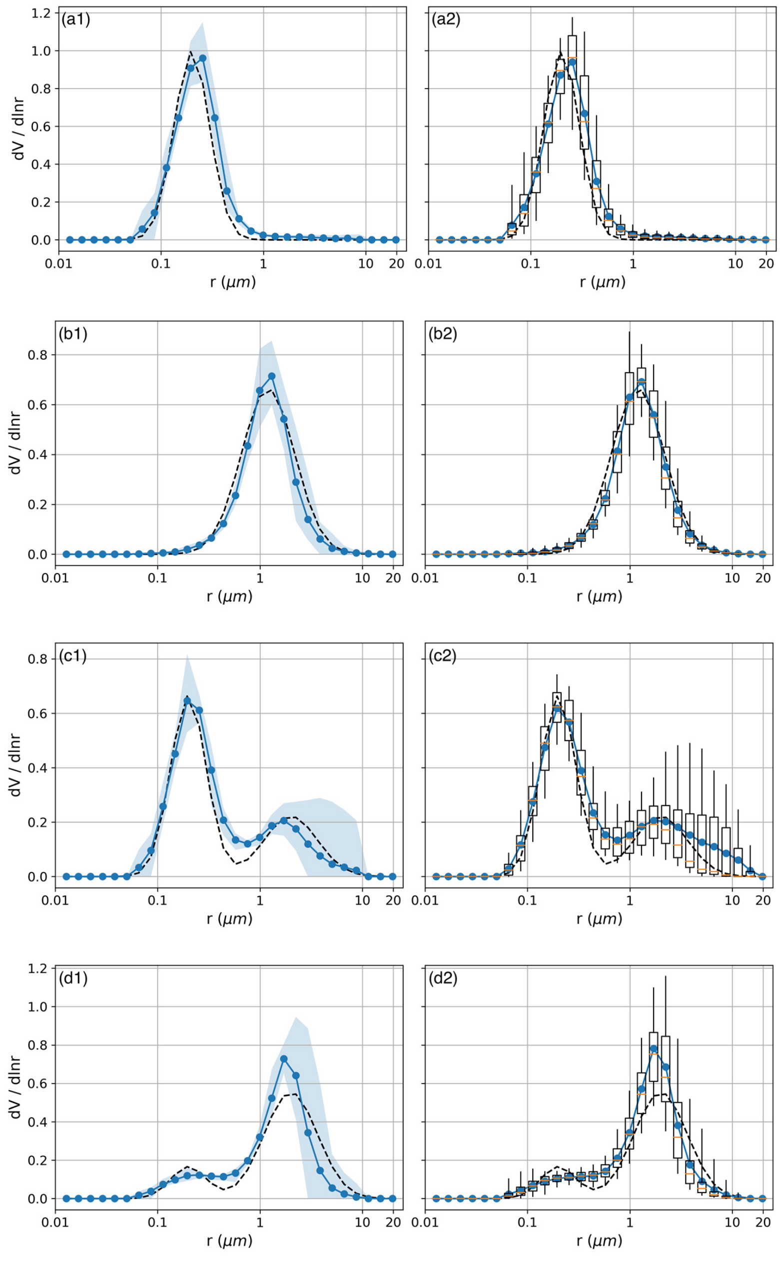


Figure 1. Comparisons of the original volume size distributions (VSDs) of the aerosol models and the retrieved VSDs. Four different VSDs in Table 1. (a. MF, b. MC, c. BF, d. BC) with complex refractive index (CRI) equal to 1.6 -i0.01 were considered. The left column (**a1**–**d1**) corresponds to the error-free optical data, where the true VSDs (dashed lines), upper and lower limits of the selected individual solutions (shaded areas), and the averaged solutions (circle solid lines) are shown. The right column (**a2**–**d2**) represents the statistics of the results when measurement error is considered, which is accomplished by adding the Gaussian error to the optical data and inverting the error-contaminated optical data 100 times. The box-and-whiskers plots show the distribution of the retrieval results, where the endpoints and horizontal lines from bottom to top correspond to the values below which 5%, 25%, 50%, 75%, and 95% of the results lie (namely, the percentile of the statistics). The blue solid lines connect the mean values of each bin.

Table 2 shows the retrieval differences, defined as the difference between the retrieved value and true value, in CRIs, V_t, and R_{eff} corresponding to the scenarios presented in Figure 1. For all these scenarios, both the real part and imaginary part are underestimated by approximately 0.05 and 50%, respectively. The retrieved imaginary parts are quite close to the a priori value ($k_a = 0.005$ for $k_{true} \leq 0.01$), and the retrieved real parts lie in the range $[n_a, n_{true}]$. If the a priori constraint $(k_a, \varepsilon_{k_a}) = (0.015, 0.01)$ is used, the imaginary parts for all the cases will be overestimated, with retrieved values also near k_a (not shown here). These facts indicate the influence of a priori constraints on the retrieval of CRI, especially on the imaginary part. To the contrary, the change of a priori constraints hardly affects the retrieval of VSDs for these scenarios. From Table 2, the retrievals of the VSDs for these scenarios have similar accuracies with δV_t ranging in [−8%, −13%] and δR_{eff} ranging in [−4%, −11%] if the optical data are error-free. On the other hand, measurement error affects these retrievals in two aspects. Firstly, it causes bias in some parameters, for example, the V_t, R_{eff} of Type BF, and Type BC. Such bias results from the overestimate of the VSD of the coarse mode, which can be inferred from Figure 1. Secondly, it disperses the retrieved parameters to a different extent, acting as statistical standard deviations shown in Table 2: the magnitudes of dispersions in k, V_t, R_{eff} are comparable with the measurement error, while those in n are much less than the measurement error.

Table 2. Retrieval differences [1], defined as the difference between the retrieved value and true value, in n, k, V_t, and R_{eff}, for the scenarios presented in Figure 1. For the error-contaminated optical data, mean differences and standard deviations (in parentheses) of the statistics are shown.

	Error-Free Optical Data				Error-Contaminated Optical Data			
	δn	δk	δV_t	δR_{eff}	δn	δk	δV_t	δR_{eff}
MF	−0.05	−53%	13%	11%	−0.05 (2%)	−52% (10%)	16 (11%)	11% (15%)
MC	−0.03	−49%	−8%	−4%	−0.03 (1%)	−51% (8%)	−9% (12%)	−6% (12%)
BF	−0.05	−49%	6%	4%	−0.05 (2%)	−47 (9%)	24% (19%)	15% (23%)
BC	−0.06	−44%	4%	−4%	−0.06 (1%)	−46% (9%)	10% (22%)	0% (26%)

[1] The retrieval differences in n are in absolute values, while those in k, V_t, and R_{eff} are in relative values.

Figure 2 shows the statistics of the absolute retrieval differences (the absolute value of retrieval difference, which is always positive) of CRIs, V_t, and R_{eff} for all the scenarios in Table 1. In general, compared with other aerosol types, retrieval differences for MF aerosols have the lowest medium values and smallest dispersions, representing the best retrieval accuracies among the four VSD types. On the other hand, BC aerosols have the largest dispersions in δV_t and δR_{eff}, which likely results from the errors in coarse-mode retrieval. For different retrieval parameters, measurement error enlarges the retrieval dispersion to various extents, influencing V_t and R_{eff} more than n and k. In particular, the dispersions of δk are nearly the same with and without measurement error. Figure 3 shows in detail the distribution of δk, from which we see $\delta k > 300\%$ when $k_{true} = 0.001$, regardless of the VSD type and n_{true}. This is because in these scenarios, the retrieved values of k are all close to k_a. Such retrieval difficulty is also faced in linear regularization methods [62]. Retrieval accuracy of k improves with the increase of k_{true} and with n_{true} getting close to n_a. For example, δk smaller than 10% can be derived when $k_{true} = 0.02$ and $n_{true} = n_a$ for all types of size distributions.

Table 3 summarizes the third quartiles of the retrieval differences corresponding to Figure 2, which we adopt as an overall estimate of retrieval accuracy with respect to the VSD type. The sensitivity study shows that using the configuration in Section 3.1, the values of VSD, V_t, R_{eff}, and CRI for typical aerosols could be retrieved with acceptable accuracies by BOREAL in the case of relative measurement uncertainty in each channel less than 10%. Note that the last quartile of δk corresponds to the scenarios where $k_{true} = 0.001$, which are all above 300%, according to Figure 3. Accordingly, once again, we emphasize the importance of the a priori information on CRI, especially on the imaginary part, to effectively constrain the final solution. Retrieval accuracy for monomodal aerosols is comparable to the result of Müller et al. [21], where a linear inversion algorithm was used to retrieve V_t, R_{eff}, and CRI.

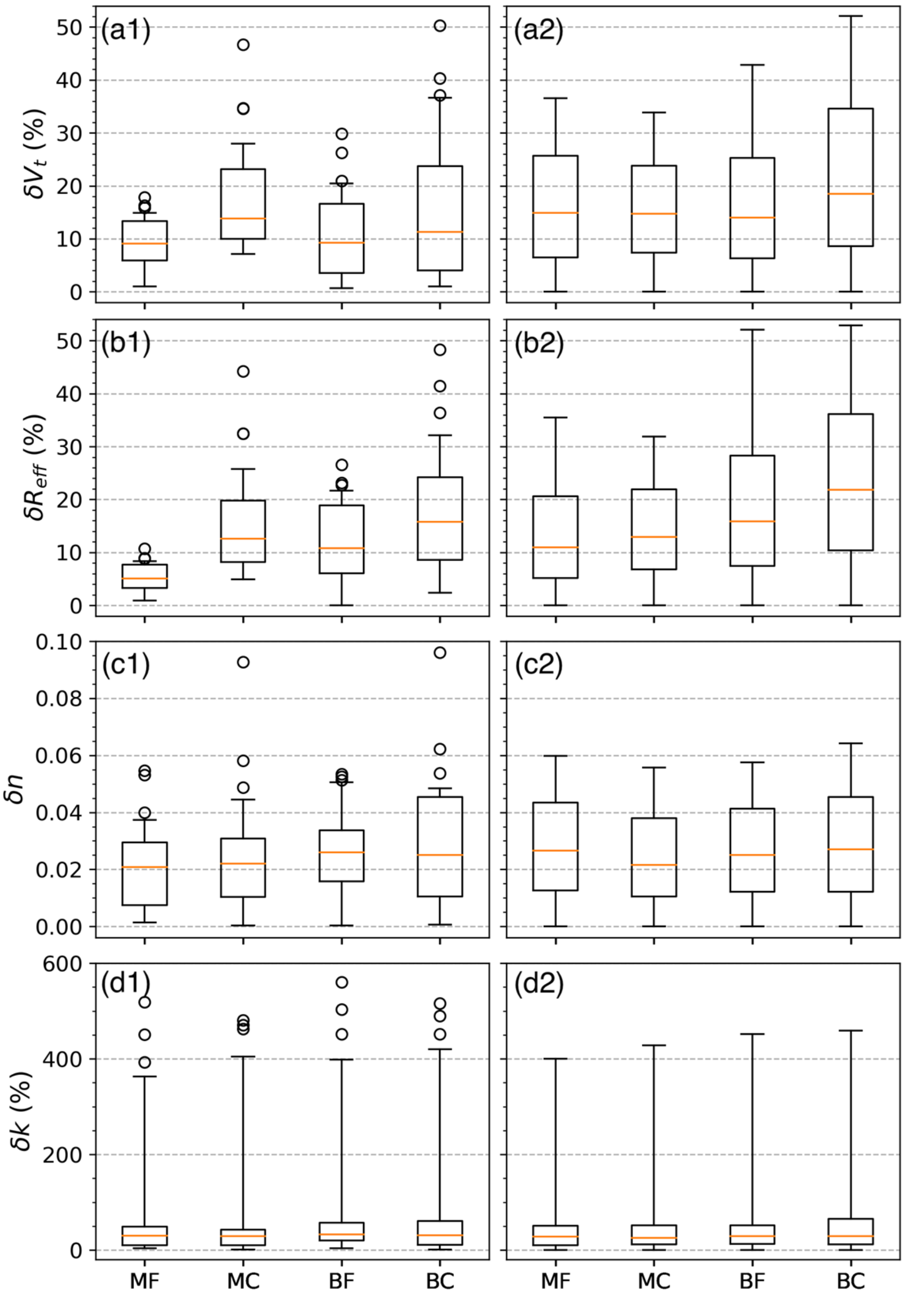


Figure 2. Box-and-whisker plots of retrieval differences, defined as the difference between the retrieved value and true value, in V_t (%), R_{eff} (%), n, and k (%) with respect to the VSD types for all the scenarios in Table 1. The left column (**a1**–**d1**) corresponds to the error-free optical data and the right column (**a2**–**d2**) to the error-contaminated optical data (i.e., each error-free scenario is perturbed by Gaussian error 100 times, thus, 10,000 scenarios in total). The hinges and horizontal lines from the bottom to top of the box-and-whiskers plots successively represent the 0, 25, 50, 75, and 90 percentiles of the dataset. Data beyond the top hinge are designated outliers and shown as hollow circles. Considering the size of the dataset, the outliers corresponding to the error-contaminated optical data are not shown.

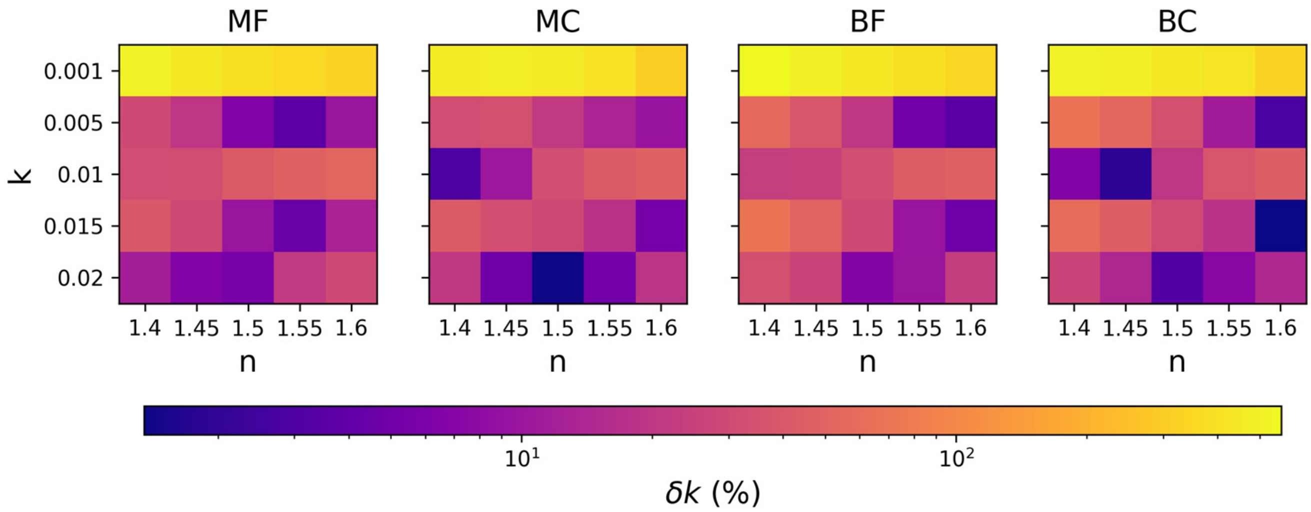


Figure 3. Distribution of δk for the retrieval scenarios in Table 1.

Table 3. Third quartiles of δV_t, δR_{eff}, δn, and δk corresponding to Figure 2.

	Error-Free Optical Data				**Error-Contaminated Optical Data**			
	δV_t	δR_{eff}	δn	δk	δV_t	δR_{eff}	δn	δk
MF	13%	8%	0.030	49%	26%	21%	0.045	51%
MC	24%	19%	0.031	43%	24%	22%	0.038	52%
BF	18%	16%	0.034	55%	25%	28%	0.040	52%
BC	23%	19%	0.042	55%	35%	36%	0.045	65%

3.3. Evaluation of the Error Propagation Model

In the second part of this section, we evaluate the feasibility of the error propagation model proposed in Section 2.4. Note that in this subsection, all the retrieval parameters are with respect to the individual solution.

Firstly, we evaluate when $\hat{\mathbf{x}}$, the function of lidar measurement $\mathbf{y}_1$, could be approximated to be linear as $\mathbf{y}_1$ varying in $\mathbf{y}_1 \pm \boldsymbol{\varepsilon}_1$. To this end, we define the relative approximation error (RAE) of a single retrieval parameter as

$$\rho = \left| \frac{\hat{x}_{pa} - \hat{x}_p}{x - \hat{x}_p} \right| \tag{33}$$

where x is the true value of the retrieval parameter, $\hat{x}_p$ is the retrieved value when a known perturbation is added to $\mathbf{y}_1$, and

$$\hat{x}_{pa} = \hat{x} + d\hat{x} \tag{34}$$

where $\hat{x}$ is the retrieved value when no perturbation is added to $\mathbf{y}_1$, and $d\hat{x}$ is calculated through the equations in Section 2.4. A low ρ indicates the linearization error is minor compared with the retrieval error caused by lidar measurement error and algorithmic error. In general, RAE should increase with the increase of measurement error because it could substantially change the path of the minimization procedure, for example, changing the iteration number from u to u', which enlarges the difference between $\hat{x}_p$ and $\hat{x}_{pa}$ since $d\hat{x}$ is evaluated for the iteration number u rather than u'.

For the scenarios in Table 1, we assigned suitable inversion windows corresponding to their VSDs. Then, we perform retrieval and calculate the RAEs of V_t, R_{eff}, n, and k when the error-free optical data are perturbed by 1%, 5%, and 10%, respectively. Optical data at different wavelengths are perturbed by the same magnitude but with different signs to imitate random effects, as shown in Table 1 of [21]. Figure 4 shows the statistical results for the MF aerosol, which are classified by whether the iteration number changes. As discussed above, RAEs for the scenarios where the iteration number changes are 3–5 times higher than those where the iteration number does not change. At the same time, the number of scenarios where the iteration number changes increases with the increase of the magnitude of perturbation. For a measurement uncertainty of 10% in each channel, (1) more than 80% of the scenarios have their iteration numbers changed with quite large RAEs and (2) among the scenarios with unchanged iteration numbers, more than 50% have the RAEs greater than 0.3, 0.4, 0.1, and 0.1 in V_t, R_{eff}, n, and k, respectively.

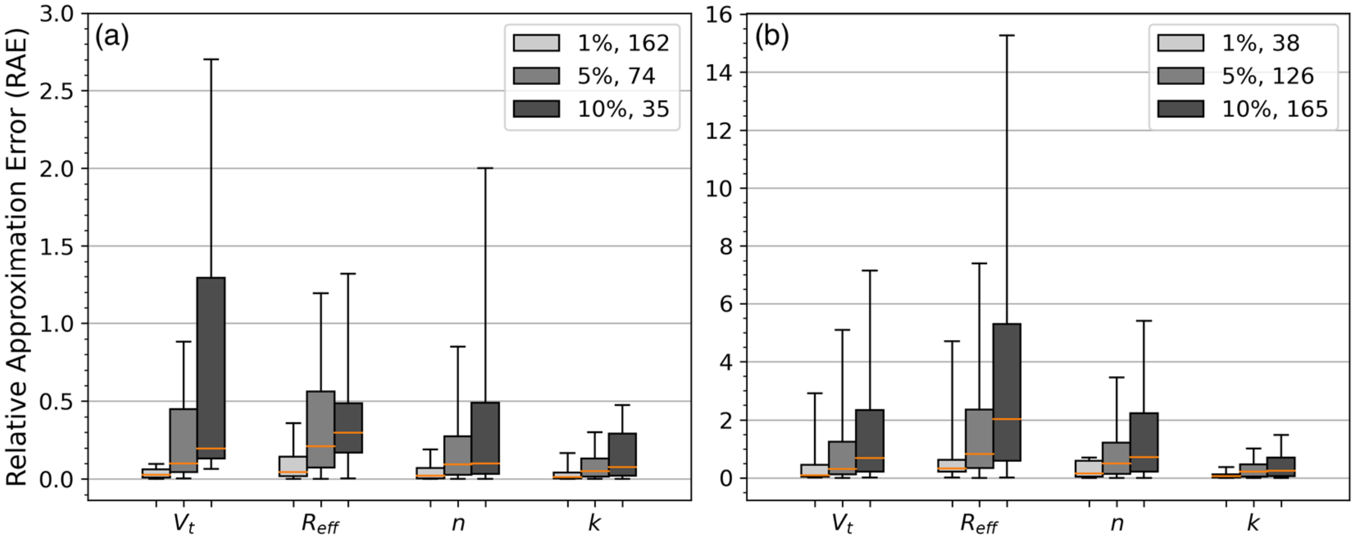


Figure 4. Relative approximation error (RAE) of V_t, R_{eff}, n, and k for MF aerosols in Table 1. (**a**) The results of which the iteration number does not change after the introduction of perturbation; (**b**) the results of which the iteration number changes after the introduction of perturbation. The magnitudes of perturbations (1%, 5%, and 10%) are labeled in the legend, followed by the counts of the cases. The hinges and horizontal lines from the bottom to top of the box-and-whiskers plots represent 0, 25, 50, 75, and 90 percentiles of the dataset.

Then, under inversion windows the same as those mentioned above, we evaluate the retrieval standard deviation (RStd) calculated with the error propagation model for a measurement uncertainty of 10%. Figure 5 shows a case-by-case comparison of the RStds of V_t, R_{eff}, n, and k calculated with the error propagation model (y-axes) and derived from the statistics of the 100 inversions of error-contaminated optical data (same as the method described in Section 3.2) (x-axes). From Figure 5, it is seen that the correlation of the RStd depends on the retrieval parameter and VSD type and, generally speaking, the difference between the calculation and experimental result is too large to allow the error propagation model to be applicable for estimating the retrieval uncertainty of the individual solution under 10% measurement uncertainty.

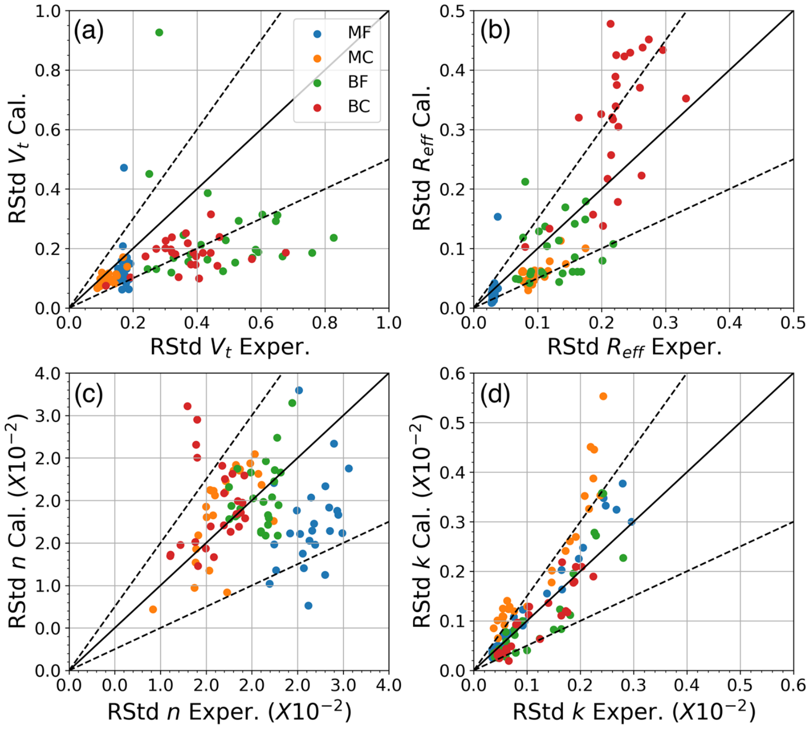


Figure 5. Case-by-case comparison of the retrieval standard deviation (RStd) of (**a**) V_t, (**b**) R_{eff}, (**c**) n, and (**d**) k calculated with the propagation model for a measurement uncertainty of 10% (y-axes) and derived from the statistics of the 100 inversions of error-contaminated optical data (same as the method described in Section 3.2) (x-axes). For each VSD type, individual solutions are derived for suitable inversion windows. In each panel, the black solid line represents the 1–1 line, and between the two dashed lines is the area where relative error is less than 50%.

4. Application to Real Lidar Measurements

To test the algorithmic performance on real aerosol events, we applied BOREAL algorithm to three representative aerosol events detected by LILAS (LIlle Lidar AtmosphereS). LILAS is a high-performance Mie–Raman–Fluorescent lidar system developed at Laboratoire d'Optique Atmosphérique as of 2013. It is capable of measuring $3\beta + 2\alpha + 3\delta + 1\beta_F$ simultaneously, where "3δ " is referred to as the particle depolarization ratio at 355 nm, 532 nm, and 1064 nm, while "$1\beta_F$" means the fluorescent backscattering coefficient centered at 466 nm. Detailed descriptions regarding the instrument and measurement uncertainties can be found in Hu et al. [30] and Veselovskii et al. [31]. The computer used for the retrievals is equipped with a 2.3 GHz Intel 8-Core i9 processor. Processing time of the CPU in each case was counted as an indicator of the algorithmic efficiency.

4.1. Case 1: 10 April 2015, Dakar

This observation was recorded in Dakar during SHADOW-2 (study of Saharan Dust Over West Africa) campaign in 2015. According to the analysis of Veselovskii et al. [64], on 10 April, dry dust transported from the Sahara Desert was dominant in the atmosphere. Here, we retrieved the aerosol properties in the period of 00:00–02:00 UTC using BOREAL and compared them with the results presented in Veselovskii et al. [64], where the regularization method [38] was used to retrieve the aerosol microphysical properties. Since the spheroids' volume fraction (SVF) on that day was higher than 98%, according to AERONET retrieval, we assumed the particles were totally spheroidal, which was also adopted in Veselovskii et al. [64].

Figure 6 shows the comparison of aerosol optical parameters from lidar measurements in the period of 00:00–02:00 UTC, 10 April 2015, and recalculated from the retrieval of BOREAL. The layer 1500–4400 m, where mineral dust was mainly concentrated, was selected and resampled for the retrieval. The total processing time was ~1 min. The overall difference between the lidar measurements and recalculated measurements was less than 10% for α and 5% for β. Figure 7 shows the comparison of the profiles of V_t, R_{eff}, and CRI retrieved by BOREAL and presented in Veselovskii et al. [64]. The V_t and R_{eff} derived from BOREAL were generally smaller but within the ranges of retrieval uncertainty provided by Veselovskii et al. [64]. The profiles of the real parts of the CRI, in Figure 6b are in good agreement. The increase of the extinction Angstrom exponent (EAE) and decrease of α indicate that particles became smaller and less concentrated upon 3300 m, which is reflected in V_t and R_{eff} in Figure 7.

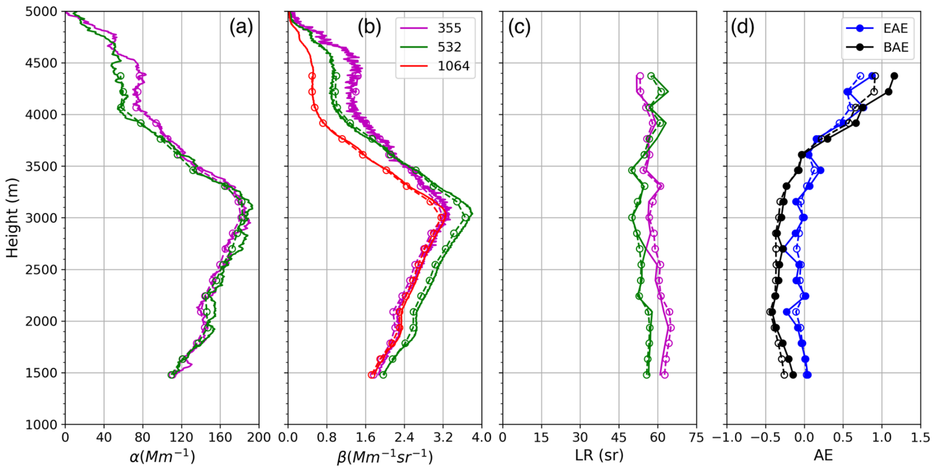


Figure 6. LILAS measurements (solid lines) and the measurements recalculated from the retrievals (dashed lines) on 10 April 2015, in the period of 00:00–02:00 UTC, at Dakar. (**a**) Extinction coefficients (α); (**b**) backscattering coefficients (β); (**c**) Lidar ratios (LRs), and (**d**) Angstrom exponents of 355 nm over 532 nm ($AE_{355-532}$), including extinction Angstrom exponent ($EAE_{355-532}$) and backscattering Angstrom exponent ($BAE_{355-532}$). The layer 1500–4400 m was selected and resampled for the retrieval. Measurements at different wavelengths are represented by the corresponding colors.

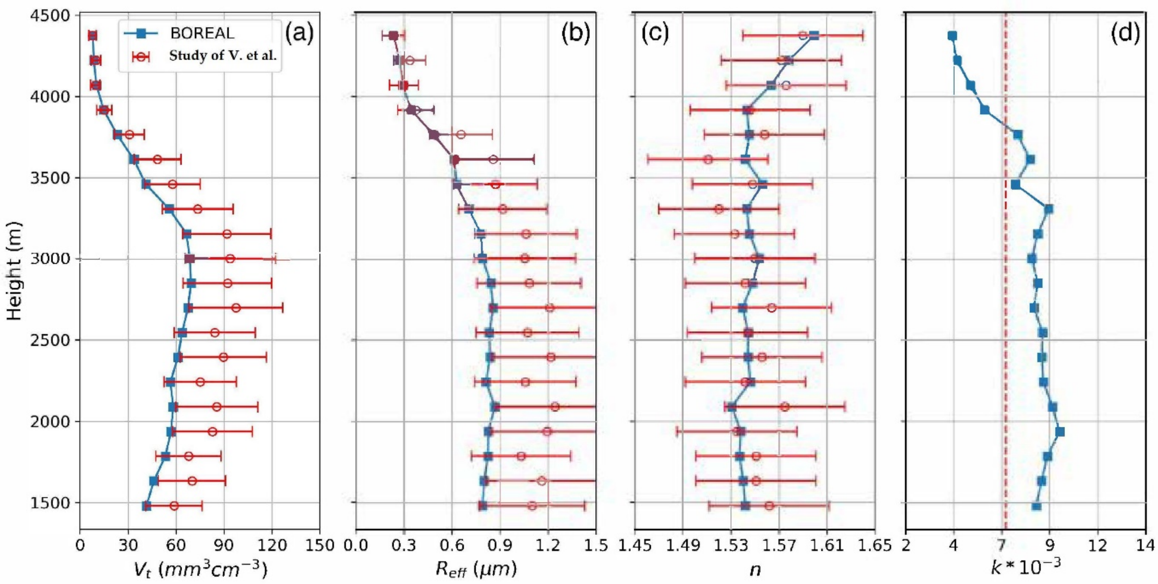


Figure 7. Comparison of retrieval results derived by BOREAL from Figure 6 (blue solid lines) and presented in Veselovskii et al. [64] (red hollow circles). (**a**) V_t; (**b**) R_{eff}; (**c**) n, and (**d**) k. The study in Veselovskii et al. [64] did not provide the profile of k but an approximated value of 0.007 for the whole dust layer (red dashed line). Because the particles are all assumed to be spheroids, results in Table 3 cannot be used here as estimates of retrieval accuracies.

To further investigate possible reasons for the underestimation of V_t and R_{eff} compared with the results in Veselovskii et al. [64], Figure 8a shows the retrieved VSDs at two heights where aerosols were concentrated. For each level, our result shows a strict mono-coarse mode with $r_v \approx 1\,\mu m$, while an extra mode with $r_v \approx 3\,\mu m$ is shown in Veselovskii et al. [64]. We attribute such differences to algorithmic principles. Due to the optimal searching strategy, BOREAL derives only one individual solution for a specific inversion window. Nevertheless, the linear regularization method [64] retrieves VSD for every combination of CRI and inversion window pre-defined in the searching domain. In addition, differences between inversion windows and selection criteria between the two algorithms could also explain the different final averaged solutions. Due to the underdetermination in lidar inverse problems, it is hard to judge which retrieval is closer to the true state without the comparison with appropriate in situ measurements, which is indeed needed for further validation. However, by checking the fitting errors shown in Figure 6, we argue that the BOREAL-derived retrieval is reasonable enough for reproducing $3\beta + 2\alpha$ lidar measurements. The right panel of Figure 8 shows a comparison of the VSDs retrieved from the vertical-integrated LILAS measurements and from AERONET. The two retrievals both have a single coarse mode with quite similar V_t. However, LILAS/BOREAL retrieval has smaller r_v and R_{eff} possibly due to: (1) the influence of retrieved CRI: the LILAS/BOREAL retrieval gives a spectral independent CRI of $n = 1.55$ and $k = 0.009$, while the AERONET retrieval gives a spectral independent n of 1.6 and a spectral dependent k (decreasing from slightly above 0.004 to below 0.001 with the increase of wavelength); (2) contributions of aerosols in the boundary layer are not taken into account in LILAS retrieval; and (3) temporal difference of 7 h between the two retrievals.

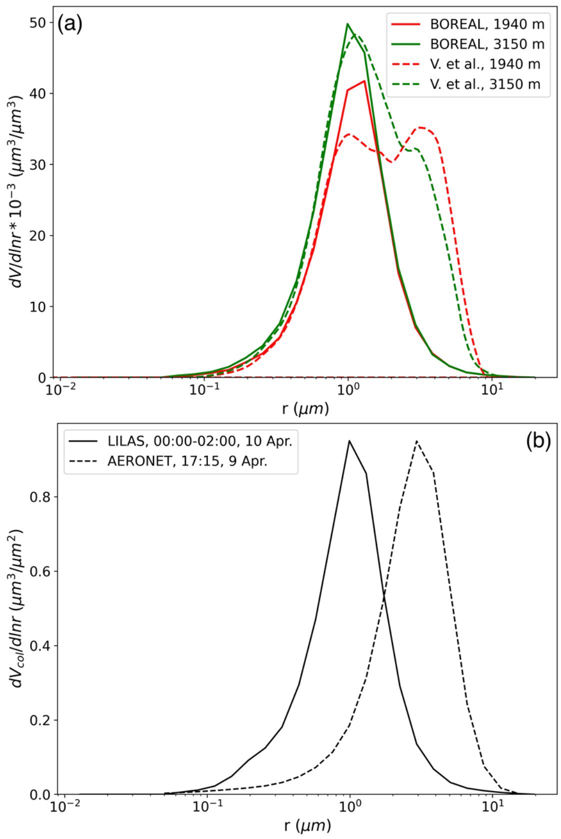
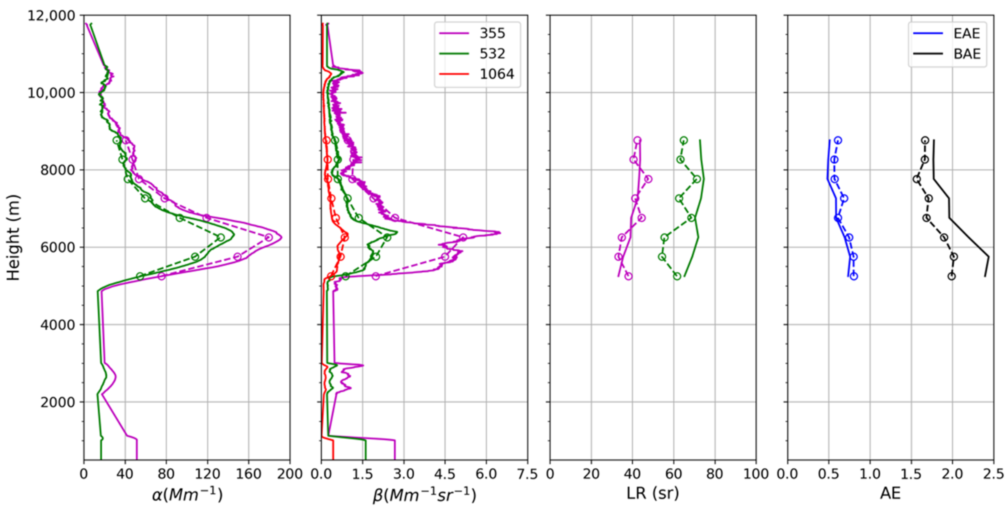


Figure 8. Comparison of VSD retrieval. (**a**) Comparison between the VSDs retrieved by BOREAL (solid lines) and presented in Veselovskii et al. [64] (dashed lines) at 2 concentrated levels, the "*" in the label of the ordinate means the multiplication symbol; (**b**) VSDs retrieved from the vertical-integrated LILAS measurements (1500–4500 m, solid line) and from AERONET measurement at 17:15 UTC, 9 April (dashed line).

4.2. Case 2: 11–12 September 2020, Lille

During this period, aged biomass burning aerosols (BBA) originating from California wildfires were observed by LILAS in operation at ATOLL [65]. Here, we averaged the measurements in the period of between 22:30–03:00 UTC, 11–12 September 2020, and retrieved the layer 5000–9000 m. Note that the vertical resolution in this case was reduced to 500 m due to the low SNR in the upper troposphere. Spherical and absorbing particle assumptions were used in the retrieval.

Figure 9 shows the LILAS measurements and the recalculated measurements in that period. The total processing time was ~1 min. The overall fitting error was less than 10% for α and 5% for β. Figure 10a,b show the retrieved profiles of, V_t, R_{eff}, and CRI. The range of EAE suggests that the aerosol layer contained mainly fine mode particles, which is reflected in R_{eff} in Figure 10a,b. The profile of V_t reveals the particles were concentrated mainly below 6500 m. The real part of CRI, n, varied between 1.51 and 1.60 while, the imaginary part, k, between 0.012 and 0.015, which are in accordance with previous remote or in situ measurements of transported biomass burning aerosols [49,66–68].



(**a**) (**b**) (**c**) (**d**)

Figure 9. Same as Figure 6 but for Case 2: 22:30–03:00 UTC, 11–12 September 2020, Lille. (**a**) α; (**b**) β; (**c**) LR, and (**d**) $AE_{355–532}$. The layer 5000–9000 m was selected and resampled for the retrieval.

Figure 10c shows the retrieved VSDs at 5250 m, 6255 m, and 8265 m, together with the AERONET level 2.0 retrieval on 11 September at 13:55 UTC. It can be seen that the selected layer contains mostly fine mode particles which are well consistent with the fine mode retrieved by AERONET. However, AERONET shows an extra coarse mode accounting for approximately 30% of the total volume concentration. To determine possible reasons for such a difference, note that the columnar $EAE_{340–500}$ measured by AERONET was 0.8 [66], while the $EAE_{355–532}$ of the selected layer measured by LILAS was 0.6. The decrease of EAE could be due to an increase of particle size (i.e., there should be a coarse mode in that layer) or an increase of the imaginary part (k) of the CRI when the fine-mode fraction predominates [69]. The later could be reasonable in this case because the AERONET retrieval returned a value of ~0.002, much lower than that retrieved by BOREAL. As mentioned in Section 3.1, here we use $(k_a, \varepsilon_{k_a}) = (0.015, 0.01)$ as the a priori constraint on k because we inferred, with the help of fluorescent measurements of LILAS, that absorbing BBA is concentrated in this layer. However, we also found during the sensitivity study that backscattering kernels corresponding to the coarse-mode region decrease with the increase of k, which means β_λ is less sensitive to the coarse-mode particles, resulting in suppression of the coarse mode under measurement noise. Another possibility results from potential uncertainty of the AERONET retrieval since it is the level 1.5 product (level 2.0 retrieval is unavailable). Therefore, the comparison here is only qualitative.

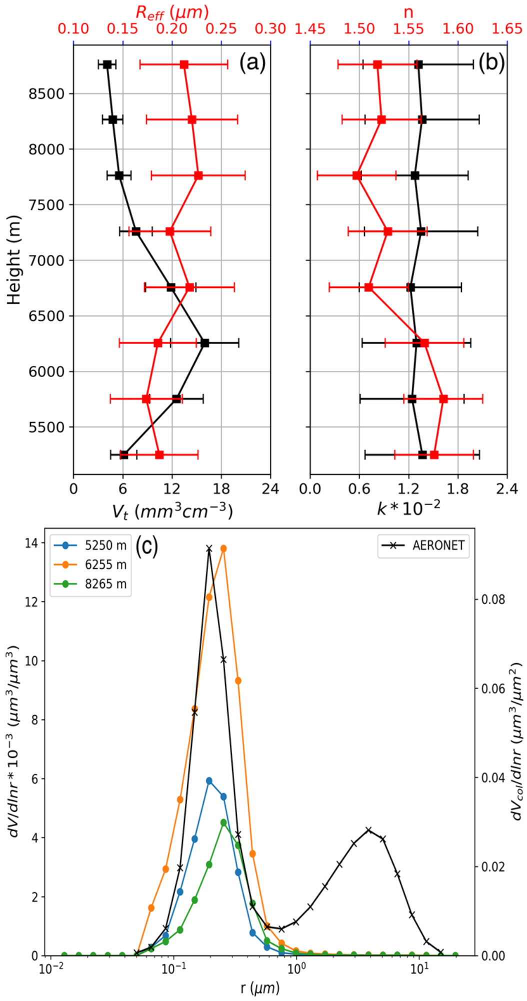


Figure 10. Retrievals for Case 2. (**a**) Profiles of V_t and R_{eff}; (**b**) profiles of n and k; and (**c**) comparison of layer-resolved VSDs from the LILAS/BOREAL retrieval and column-integral VSD from the AERONET retrieval at 13:55 UTC, 11 September 2020. The error bars in (**a**,**b**) are extracted from Table 3.

4.3. Case 3: 30–31 May 2020, Lille

On the night of 30–31 May in the period of 21:00–02:00 UTC, a mixture of pollen grains and urban aerosols from 500 m to 2500 m was observed by LILAS [70]. Considering the wavelength limit of the $3\beta + 2\alpha$ measurements, we retrieved the layer between 1300 m and 2200 m where background urban aerosols mainly concentrated, according to the aerosol classification based on depolarization and fluorescence observations [47]. Spherical and non-absorbing particle assumptions are used in the retrieval.

Figure 11 shows the LILAS measurements and the recalculated measurements in that period. The total processing time was 24 s. Compared with the measurements in previous two cases, the stable and low signals in this case suggest background aerosols, which could consist of fine-mode particles according to the $EAE_{355–532}$. Figure 12a,b show the retrieved profiles of V_t, R_{eff}, and CRI. The R_{eff} varied between 0.12 µm and 0.15 µm, which explains the range of EAE shown in Figure 11. The real part of CRI decreased from 1.57 to 1.50 with the increase of altitude, while the imaginary part of CRI varied between 0.0042 and 0.0049, slightly lower than the a priori value 0.005. Figure 12c shows the retrieved VSDs at different heights, together with the AERONET level 2.0 retrieval on 30 May at 16:28 UTC. The VSDs from the LILAS/BOREAL retrieval are predominated by fine particles with 0.1 µm $< r_v <$ 0.2 µm, which are well consistent with the fine mode from the AERONET retrieval. The predominated coarse-mode retrieved by AERONET was very likely to be pollen grains because: (1) the daily cycle of pollen grains, where maximum emission occurs near noon and less emission happens during the night, was validated by in situ measurements [70]; (2) the selected layer excluded the influence of pollen grains according to the aerosol classification result [47]; and (3) the $EAE_{355–532}$ measured by LILAS (~2) was larger than the $EAE_{340–500}$ measured by AERONET (~1.5), and the low value of the imaginary part was retrieved, which indicate the lack of coarse-mode particles in the selected layer.

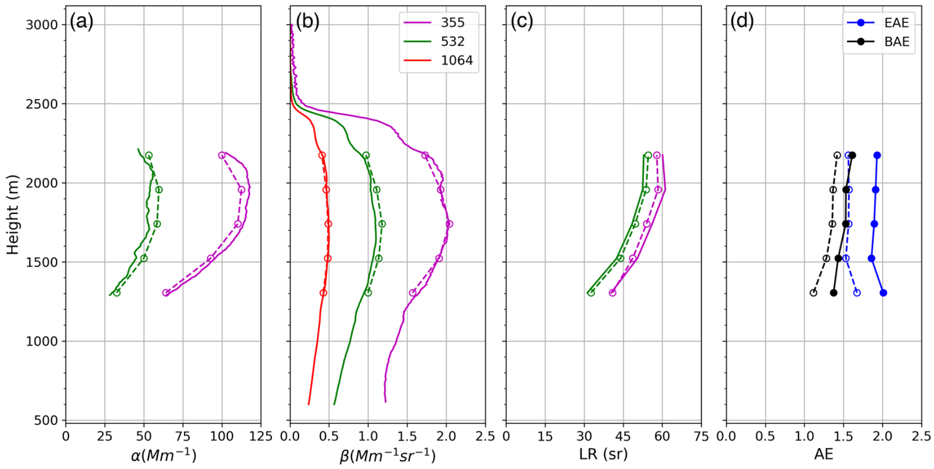


Figure 11. Same as Figure 6 but for Case 3: 21:00–03:00 UTC, 11–12 September 2020, Lille. (**a**) α; (**b**) β; (**c**) LR, and (**d**) $AE_{355–532}$. The layer 1300–2200 m was selected and resampled for the retrieval.

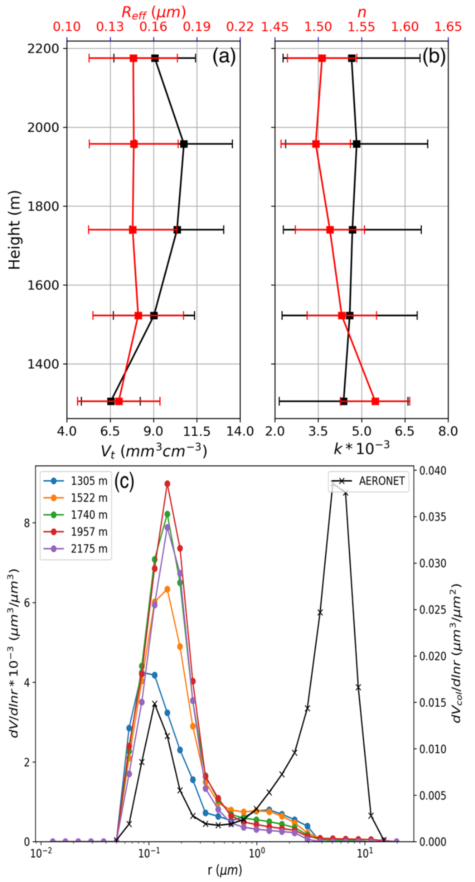


Figure 12. Retrievals for Case 3. (**a**) Profiles of V_t and R_{eff}; (**b**) profiles of n and k; and (**c**) comparison of layer-resolved VSDs from the LILAS/BOREAL retrieval and column-integral VSD from the AERONET retrieval at 16:28 UTC, 30 May 2020. The error bars in (**a**,**b**) are extracted from Table 3.

5. Conclusions

The retrieval of height-resolved aerosol microphysical properties is of ever-increasing interest in the field of aerosol remote sensing with the development of lidar networks based on high-performance Mie–Raman lidar systems. In this study, we developed BOREAL, a non-linear inversion algorithm based on maximum likelihood estimation (MLE) to retrieve particle VSD, V_t, R_{eff}, and CRI ($n - ik$) from the 3β (backscattering coefficient at 355 nm, 532 nm, and 1064 nm) + 2α (extinction coefficient at 355 nm and 532 nm) measured by the Mie–Raman lidar. Compared with other linear retrieval algorithms such as the regularization method and principal component analysis method, BOREAL simultaneously retrieves VSD and CRI by performing optimal searching rather than walking through the whole searching domain, which is evidently more efficient. Based on statistical principles, it is seen that measurement errors are well considered, and their magnitudes serve as scaling factors for the corresponding measurements. At the same time, a priori constraints are treated as virtual measurements with straightforward statistical meaning. Furthermore, the general form of the algorithm will remain unchanged, and computational burden will not evidently increase if more measurements with non-linear forward models are incorporated. To realize stable and realistic retrieval from the ill-posed inversion system, BOREAL (1) utilizes the smoothing constraint on VSD and the a priori constraint on CRI, (2) sets up stopping conditions on the basis of statistical properties, and (3) selects qualified individual solutions derived from a series inversion windows.

We used synthetic optical data $(3\beta + 2\alpha)$ generated by different aerosol models to test the performance of BOREAL when one set of a priori constraint on the real part and two sets of a priori constraints on the imaginary part of the CRI were employed. Sensitivity tests show the robustness of the algorithm. For monomodal and bimodal aerosols with n varying in 1.4–1.6 and k varying in 0.005–0.02, VSD, V_t, R_{eff}, and CRI could be retrieved with acceptable accuracies when measurement uncertainty in each channel is up to 10%. We conclude that $3\beta + 2\alpha$ measurements have limit sensitivity to very low imaginary parts (k~0.001) and large particles, which, in turn, increases retrieval uncertainty for these parameters. At the same time, insufficient information content of $3\beta + 2\alpha$ measurements on the imaginary part increases the influence of the a priori constraint.

We proposed and evaluated an error propagation model, aiming to provide rigorous and real-time estimate of the retrieval covariance matrix, which is a function of the measurement covariance matrix. However, simulation results show that measurement errors in $3\beta + 2\alpha$ data are too large to obey a linear propagation rule, which makes the error propagation model not applicable enough for most cases.

We applied BOREAL to several representative aerosol events: Saharan dust, transported smoke, and background urban aerosols during a pollen season detected by LILAS. The retrieval of the dust case shows good consistency with the result presented by [64], except for overestimates in V_t and R_{eff}, which we attribute to the differences in algorithmic principles. The comparisons with AERONET illustrate the advantages and limits of lidar and sun photometer measurements and demonstrate that the aerosol events could be well interpreted by our retrievals.

The next step will also focus on improving the retrieval of CRI, especially the imaginary part. This might be accomplished by further constraining CRI with aerosol-typing results using lidar measurements (for example, see Veselovskii et al. [47]). Another perspective is to incorporate spectral depolarization measurements into the inversion scheme to realize accurate retrieval of non-spherical particles. For this purpose, application of scattering models accurately describing the backscattering of non-spherical particles and assessment of information content of depolarization measurements are needed. At the same time, the first version of BOREAL is being implemented into the AUSTRAL (Automated Server for the Treatment of Atmospheric Lidars) [71] processing and inversion framework to more efficiently evaluate the code with real lidar data and, finally, to implement automated aerosol retrieval and further services.

Author Contributions: Conceptualization, Y.C., Q.H. and P.G.; methodology, Y.C., Q.H. and P.G.; software, Y.C.; validation, Y.C.; formal analysis, Y.C., Q.H. and P.G.; investigation, Y.C.; resources, Q.H., P.G., I.V. and T.P.; data curation, Y.C.; writing—original draft preparation, Y.C.; writing—review and editing, Y.C., Q.H., P.G., I.V. and T.P.; visualization, Y.C.; supervision, P.G.; project administration, P.G.; funding acquisition, P.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by French National Research Agency (ANR) through the PIA (Programme d'Investissement d'Avenir) under contract "ANR-11-LABX-0005-01" and by the Regional Council "Hauts-de-France" and the "European Funds for Regional Economic Development" (FEDER).

Data Availability Statement: Database of the spherical–spheroidal scattering model can be found in the GRASP-OPEN server https://www.grasp-open.com/products/spheroid-package-release (accessed on 4 December 2022), AERONET data can be found in the AERONET NASA server https://aeronet.gsfc.nasa.gov (accessed on 4 December 2022), and the LILAS measurements and retrievals presented in this study are available on request from the corresponding author.

Acknowledgments: The authors acknowledge the support of ESA/QA4EO program, CNRS-INSU National Observation Service PHOTONS-AERONET/EARLINET teams, and Lille University for the observation activity at LOA. The authors thank the PIs of the relevant sites of AERONET for establishing and maintaining the sites used in this investigation. The authors also thank the GRASP team for providing the database of the particle scattering model.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Charlson, R.J.; Schwartz, S.E.; Hales, J.M.; Cess, R.D.; Coakley, J.A.; Hansen, J.E.; Hofmann, D.J. Climate Forcing by Anthropogenic Aerosols. *Science* **1992**, *255*, 423–430. [CrossRef] [PubMed]
2. Forster, P.; Storelvmo, T.; Armour, K.; Collins, W.; Dufresne, J.-L.; Frame, D.; Lunt, D.J.; Mauritsen, T.; Palmer, M.D.; Watanabe, M.; et al. The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*; Masson-Delmotte, V., Zhai, P., Eds.; Cambridge University Press: Cambridge, UK; New York, NY, USA, 2021; pp. 923–1054.
3. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanré, D.; Buis, J.P.; Setzer, A.; Vermote, E.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [CrossRef]
4. Deschamps, P.-Y.; Breon, F.-M.; Leroy, M.; Podaire, A.; Bricaud, A.; Buriez, J.-C.; Seze, G. The POLDER Mission: Instrument Characteristics and Scientific Objectives. *IEEE Trans. Geosci. Remote Sens.* **1994**, *32*, 598–615. [CrossRef]
5. Salomonson, V.V.; Barnes, W.L.; Maymon, P.W.; Montgomery, H.E.; Ostrow, H. MODIS: Advanced Facility Instrument for Studies of the Earth as a System. *IEEE Trans. Geosci. Remote Sens.* **1989**, *27*, 145–153. [CrossRef]
6. Haywood, J.M.; Roberts, D.L.; Slingo, A.; Edwards, J.M.; Shine, K.P. General Circulation Model Calculations of the Direct Radiative Forcing by Anthropogenic Sulfate and Fossil-Fuel Soot Aerosol. *J. Clim.* **1997**, *10*, 1562–1577. [CrossRef]
7. Fiocco, G.; Smullin, L.D. Detection of Scattering Layers in the Upper Atmosphere (60–140 Km) by Optical Radar. *Nature* **1963**, *199*, 1275–1276. [CrossRef]
8. Klett, J.D. Lidar Inversion with Variable Backscatter/Extinction Ratios. *Appl. Opt.* **1985**, *24*, 1638–1643. [CrossRef]
9. Zuev, V.E.; Naats, I.E. *Inverse Problems of Lidar Sensing of the Atmosphere*; Springer: Berlin, Germany, 1983.
10. Müller, H.; Quenzel, H. Information Content of Multispectral Lidar Measurements with Respect to the Aerosol Size Distribution. *Appl. Opt.* **1985**, *24*, 648–654. [CrossRef]
11. Qing, P.; Nakane, H.; Sasano, Y.; Kitamura, S. Numerical Simulation of the Retrieval of Aerosol Size Distribution from Multiwavelength Laser Radar Measurements. *Appl. Opt.* **1989**, *28*, 5259–5265. [CrossRef]
12. Ansmann, A.; Riebesell, M.; Wandinger, U.; Weitkamp, C.; Voss, E.; Lahmann, W.; Michaelis, W. Combined Raman Elastic-Backscatter LIDAR for Vertical Profiling of Moisture, Aerosol Extinction, Backscatter, and LIDAR Ratio. *Appl. Phys. B* **1992**, *55*, 18–28. [CrossRef]
13. Althausen, D.; Müller, D.; Ansmann, A.; Wandinger, U.; Hube, H.; Clauder, E.; Zörner, S. Scanning 6-Wavelength 11-Channel Aerosol Lidar. *J. Atmos. Ocean. Technol.* **2000**, *17*, 1469–1482. [CrossRef]
14. Piironen, P.; Eloranta, E.W. Demonstration of a High-Spectral-Resolution Lidar Based on an Iodine Absorption Filter. *Opt. Lett.* **1994**, *19*, 234–236. [CrossRef] [PubMed]
15. Rocadenbosch, F.; Mattis, I.; Ansmann, A.; Wandinger, U.; Bockmann, C.; Pappalardo, G.; Amodeo, A.; Bosenberg, J.; Alados-Arboledas, L.; Apituley, A.; et al. The European Aerosol Research Lidar Network (EARLINET): An Overview. In Proceedings of the IGARSS 2008—2008 IEEE International Geoscience and Remote Sensing Symposium, Boston, MA, USA, 8–11 July 2008; Volume 2, pp. II-410–II-413.

16. Welton, E.J.; Campbell, J.R.; Spinhirne, J.D.; Scott, V.S., III. Global Monitoring of Clouds and Aerosols Using a Network of Micropulse Lidar Systems. In *Lidar Remote Sensing for Industry and Environment Monitoring*; Singh, U.N., Asai, K., Ogawa, T., Singh, U.N., Itabe, T., Sugimoto, N., Eds.; SPIE: Bellingham, WA, USA, 2001; Volume 4153, pp. 151–158.
17. Nishizawa, T.; Sugimoto, N.; Matsui, I.; Shimizu, A.; Higurashi, A.; Jin, Y. The Asian Dust and Aerosol Lidar Observation Network (AD-NET): Strategy and Progress. In *EPJ Web of Conferences*; EDP Sciences: Les Ulis, France, 2016; Volume 119, p. 19001. [CrossRef]
18. Müller, D.; Wandinger, U.; Ansmann, A. Microphysical Particle Parameters from Extinction and Backscatter Lidar Data by Inversion with Regularization: Theory. *Appl. Opt.* **1999**, *38*, 2346–2357. [CrossRef] [PubMed]
19. Böckmann, C. Hybrid Regularization Method for the Ill-Posed Inversion of Multiwavelength Lidar Data in the Retrieval of Aerosol Size Distributions. *Appl. Opt.* **2001**, *40*, 1329–1342. [CrossRef]
20. Veselovskii, I.; Kolgotin, A.; Griaznov, V.; Müller, D.; Wandinger, U.; Whiteman, D.N. Inversion with Regularization for the Retrieval of Tropospheric Aerosol Parameters from Multiwavelength Lidar Sounding. *Appl. Opt.* **2002**, *41*, 3685–3699. [CrossRef] [PubMed]
21. Müller, D.; Chemyakin, E.; Kolgotin, A.; Ferrare, R.A.; Hostetler, C.A.; Romanov, A. Automated, Unsupervised Inversion of Multiwavelength Lidar Data with TiARA: Assessment of Retrieval Performance of Microphysical Parameters Using Simulated Data. *Appl. Opt.* **2019**, *58*, 4981–5008. [CrossRef] [PubMed]
22. Donovan, D.P.; Carswell, A.I. Principal Component Analysis Applied to Multiwavelength Lidar Aerosol Backscatter and Extinction Measurements. *Appl. Opt.* **1997**, *36*, 9406–9424. [CrossRef]
23. Veselovskii, I.; Dubovik, O.; Kolgotin, A.; Korenskiy, M.; Whiteman, D.N.; Allakhverdiev, K.; Huseyinoglu, F. Linear Estimation of Particle Bulk Parameters from Multi-Wavelength Lidar Measurements. *Atmos. Meas. Tech.* **2012**, *5*, 1135–1145. [CrossRef]
24. Chemyakin, E.; Müller, D.; Burton, S.; Kolgotin, A.; Hostetler, C.; Ferrare, R. Arrange and Average Algorithm for the Retrieval of Aerosol Parameters from Multiwavelength High-Spectral-Resolution Lidar/Raman Lidar Data. *Appl. Opt.* **2014**, *53*, 7252–7266. [CrossRef]
25. Veselovskii, I.; Kolgotin, A.; Griaznov, V.; Müller, D.; Franke, K.; Whiteman, D.N. Inversion of Multiwavelength Raman Lidar Data for Retrieval of Bimodal Aerosol Size Distribution. *Appl. Opt.* **2004**, *43*, 1180–1195. [CrossRef]
26. Veselovskii, I.; Kolgotin, A.; Müller, D.; Whiteman, D. Information Content of Multiwavelength Lidar Data with Respect to Microphysical Particle Properties Derived from Eigenvalue Analysis. *Appl. Opt.* **2005**, *44*, 5292–5303. [CrossRef] [PubMed]
27. Burton, S.P.; Chemyakin, E.; Liu, X.; Knobelspiesse, K.; Stamnes, S.; Sawamura, P.; Moore, R.H.; Hostetler, C.A.; Ferrare, R.A. Information Content and Sensitivity of the $3\beta + 2\alpha$ Lidar Measurement System for Aerosol Microphysical Retrievals. *Atmos. Meas. Tech.* **2016**, *9*, 5555–5574. [CrossRef]
28. Chemyakin, E.; Burton, S.; Kolgotin, A.; Müller, D.; Hostetler, C.; Ferrare, R. Retrieval of Aerosol Parameters from Multiwavelength Lidar: Investigation of the Underlying Inverse Mathematical Problem. *Appl. Opt.* **2016**, *55*, 2188. [CrossRef] [PubMed]
29. Kolgotin, A.; Müller, D.; Chemyakin, E.; 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. *Appl. Opt.* **2016**, *55*, 9839–9849. [CrossRef] [PubMed]
30. Hu, Q.; Goloub, P.; Veselovskii, I.; Bravo-Aranda, J.-A.; Popovici, I.E.; Podvin, T.; Haeffelin, M.; Lopatin, A.; Dubovik, O.; Pietras, C.; et al. Long-Range-Transported Canadian Smoke Plumes in the Lower Stratosphere over Northern France. *Atmos. Chem. Phys.* **2019**, *19*, 1173–1193. [CrossRef]
31. Veselovskii, I.; Hu, Q.; Goloub, P.; Podvin, T.; Korenskiy, M.; Pujol, O.; Dubovik, O.; Lopatin, A. Combined Use of Mie–Raman and Fluorescence Lidar Observations for Improving Aerosol Characterization: Feasibility Experiment. *Atmos. Meas. Tech.* **2020**, *13*, 6691–6701. [CrossRef]
32. Mishchenko, M.I.; Travis, L.D. T-Matrix Computations of Light Scattering by Large Spheroidal Particles. *Opt. Commun.* **1994**, *109*, 16–21. [CrossRef]
33. Yang, P.; Liou, K.N. Geometric-Optics–Integral-Equation Method for Light Scattering by Nonspherical Ice Crystals. *Appl. Opt.* **1996**, *35*, 6568–6584. [CrossRef]
34. DeBoor, C. *A Practical Guide to B-Splines*; Springer: Berlin/Heidelberg, Germany, 1978; Volume 27.
35. O'Sullivan, F. A Statistical Perspective on Ill-Posed Inverse Problems. *Stat. Sci.* **1986**, *1*, 502–518.
36. Dubovik, O.; King, M.D. A Flexible Inversion Algorithm for Retrieval of Aerosol Optical Properties from Sun and Sky Radiance Measurements. *J. Geophys. Res. Atmos.* **2000**, *105*, 20673–20696. [CrossRef]
37. Dubovik, O.; Sinyuk, A.; Lapyonok, T.; Holben, B.N.; Mishchenko, M.; Yang, P.; Eck, T.F.; Volten, H.; Muñoz, O.; Veihelmann, B.; et al. Application of Spheroid Models to Account for Aerosol Particle Nonsphericity in Remote Sensing of Desert Dust. *J. Geophys. Res. Atmos.* **2006**, *111*. [CrossRef]
38. Veselovskii, I.; Dubovik, O.; Kolgotin, A.; Lapyonok, T.; Di Girolamo, P.; Summa, D.; Whiteman, D.N.; Mishchenko, M.; Tanré, D. Application of Randomly Oriented Spheroids for Retrieval of Dust Particle Parameters from Multiwavelength Lidar Measurements. *J. Geophys. Res. Atmos.* **2010**, *115*. [CrossRef]
39. Bi, L.; Lin, W.; Wang, Z.; Tang, X.; Zhang, X.; Yi, B. Optical Modeling of Sea Salt Aerosols: The Effects of Nonsphericity and Inhomogeneity. *J. Geophys. Res. Atmos.* **2018**, *123*, 543–558. [CrossRef]
40. Saito, M.; Yang, P.; Ding, J.; Liu, X. A Comprehensive Database of the Optical Properties of Irregular Aerosol Particles for Radiative Transfer Simulations. *J. Atmos. Sci.* **2021**, *78*, 2089–2111. [CrossRef]

41. Hadamard, J. *Lectures on Cauchy's Problems in Linear Partial Differential Equations*; Yale Univ. Press: New Haven, CT, USA, 1923.
42. Romanov, P.; O'Neill, N.T.; Royer, A.; McArthur, B.L.J. Simultaneous Retrieval of Aerosol Refractive Index and Particle Size Distribution from Ground-Based Measurements of Direct and Scattered Solar Radiation. *Appl. Opt.* **1999**, *38*, 7305–7320. [CrossRef] [PubMed]
43. Pérez-Ramírez, D.; Whiteman, D.N.; Veselovskii, I.; Kolgotin, A.; Korenskiy, M.; Alados-Arboledas, L. Effects of Systematic and Random Errors on the Retrieval of Particle Microphysical Properties from Multiwavelength Lidar Measurements Using Inversion with Regularization. *Atmos. Meas. Tech.* **2013**, *6*, 3039–3054. [CrossRef]
44. Tesche, M.; Ansmann, A.; Müller, D.; Althausen, D.; Engelmann, R.; Freudenthaler, V.; Groß, S. Vertically Resolved Separation of Dust and Smoke over Cape Verde Using Multiwavelength Raman and Polarization Lidars during Saharan Mineral Dust Experiment 2008. *J. Geophys. Res. Atmos.* **2009**, *114*. [CrossRef]
45. Burton, S.P.; Ferrare, R.A.; Hostetler, C.A.; Hair, J.W.; Rogers, R.R.; Obland, M.D.; Butler, C.F.; Cook, A.L.; Harper, D.B.; Froyd, K.D. Aerosol Classification Using Airborne High Spectral Resolution Lidar Measurements—Methodology and Examples. *Atmos. Meas. Tech.* **2012**, *5*, 73–98. [CrossRef]
46. Nicolae, D.; Vasilescu, J.; Talianu, C.; Binietoglou, I.; Nicolae, V.; Andrei, S.; Antonescu, B. A Neural Network Aerosol-Typing Algorithm Based on Lidar Data. *Atmos. Chem. Phys.* **2018**, *18*, 14511–14537. [CrossRef]
47. Veselovskii, I.; Hu, Q.; Goloub, P.; Podvin, T.; Barchunov, B.; Korenskii, M. Combining Mie–Raman and Fluorescence Observations: A Step Forward in Aerosol Classification with Lidar Technology. *Atmos. Meas. Tech.* **2022**, *15*, 4881–4900. [CrossRef]
48. Di Biagio, C.; Formenti, P.; Balkanski, Y.; Caponi, L.; Cazaunau, M.; Pangui, E.; Journet, E.; Nowak, S.; Andreae, M.O.; Kandler, K.; et al. Complex Refractive Indices and Single-Scattering Albedo of Global Dust Aerosols in the Shortwave Spectrum and Relationship to Size and Iron Content. *Atmos. Chem. Phys.* **2019**, *19*, 15503–15531. [CrossRef]
49. Dubovik, O.; Holben, B.; Eck, T.F.; Smirnov, A.; Kaufman, Y.J.; King, M.D.; Tanré, D.; Slutsker, I. Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations. *J. Atmos. Sci.* **2002**, *59*, 590–608. [CrossRef]
50. Fisher, R.A.; Russell, E.J. On the Mathematical Foundations of Theoretical Statistics. *Philos. Trans. R. Soc. London. Ser. A Contain. Pap. A Math. Or Phys. Character* **1922**, *222*, 309–368. [CrossRef]
51. Tarantola, A. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*; Elsevier Sci: New York, NY, USA, 1987.
52. Rodgers, C.D. *Inverse Methods for Atmospheric Sounding: Theory and Practice*; Atmospheric, Oceanic and Planetary Physics; World Scientific Publishing Co. Pte. Ltd.: Singapore, 2000; Volume 2.
53. Gavin, H.P. *The Levenberg-Marquardt Algorithm for Nonlinear Least Squares Curve-Fitting Problems*; Department of Civil and Environmental Engineering, Duke University: Durham, NC, USA, 2020.
54. Davies, C.N. Size Distribution of Atmospheric Particles. *J. Aerosol Sci.* **1974**, *5*, 293–300. [CrossRef]
55. Whitby, K.T. THE PHYSICAL CHARACTERISTICS OF SULFUR AEROSOLS. In *Sulfur in the Atmosphere*; Husar, R.B., Lodge, J.P., Moore, D.J., Eds.; Pergamon: Oxford, UK, 1978; pp. 135–159, ISBN 978-0-08-022932-4.
56. Ott, W.R. A Physical Explanation of the Lognormality of Pollutant Concentrations. *J. Air Waste Manag. Assoc.* **1990**, *40*, 1378–1383. [CrossRef]
57. Remer, L.A.; Kaufman, Y.J. Dynamic Aerosol Model: Urban/Industrial Aerosol. *J. Geophys. Res. Atmos.* **1998**, *103*, 13859–13871. [CrossRef]
58. Tanré, D.; Remer, L.A.; Kaufman, Y.J.; Mattoo, S.; Hobbs, P.V.; Livingston, J.M.; Russell, P.B.; Smirnov, A. Retrieval of Aerosol Optical Thickness and Size Distribution over Ocean from the MODIS Airborne Simulator during TARFOX. *J. Geophys. Res. Atmos.* **1999**, *104*, 2261–2278. [CrossRef]
59. Zhang, X. *Matrix Analysis and Applications*; Wang, Y., Ed.; Tsinghua University Press: Beijing, China, 2013.
60. Omar, A.; Won, J.-G.; Winker, D.; Yoon, S.; Dubovik, O.; Mccormick, M. Development of Global Aerosol Models Using Cluster Analysis of Aerosol Robotic Network (AERONET) Measurements. *J. Geophys. Res* **2005**, *110*, 10–14. [CrossRef]
61. Müller, D.; Ansmann, A.; Mattis, I.; Tesche, M.; Wandinger, U.; Althausen, D.; Pisani, G. Aerosol-Type-Dependent Lidar Ratios Observed with Raman Lidar. *J. Geophys. Res. Atmos.* **2007**, *112*. [CrossRef]
62. Müller, D.; Veselovskii, I.; Kolgotin, A.; Tesche, M.; Ansmann, A.; Dubovik, O. Vertical Profiles of Pure Dust and Mixed Smoke–Dust Plumes Inferred from Inversion of Multiwavelength Raman/Polarization Lidar Data and Comparison to AERONET Retrievals and in Situ Observations. *Appl. Opt.* **2013**, *52*, 3178–3202. [CrossRef]
63. Hu, Q.; Wang, H.; Goloub, P.; Li, Z.; Veselovskii, I.; Podvin, T.; Li, K.; Korenskiy, M. The Characterization of Taklamakan Dust Properties Using a Multiwavelength Raman Polarization Lidar in Kashi, China. *Atmos. Chem. Phys.* **2020**, *20*, 13817–13834. [CrossRef]
64. Veselovskii, I.; Goloub, P.; Podvin, T.; Bovchaliuk, V.; Derimian, Y.; Augustin, P.; Fourmentin, M.; Tanre, D.; Korenskiy, M.; Whiteman, D.N.; et al. Retrieval of Optical and Physical Properties of African Dust from Multiwavelength Raman Lidar Measurements during the SHADOW Campaign in Senegal. *Atmos. Chem. Phys.* **2016**, *16*, 7013–7028. [CrossRef]
65. Hu, Q.; Goloub, P.; Veselovskii, I.; Podvin, T. The Characterization of Long-Range Transported North American Biomass Burning Plumes: What Can a Multi-Wavelength Mie–Raman-Polarization-Fluorescence Lidar Provide? *Atmos. Chem. Phys.* **2022**, *22*, 5399–5414. [CrossRef]
66. Guyon, P.; Boucher, O.; Graham, B.; Beck, J.; Mayol-Bracero, O.L.; Roberts, G.C.; Maenhaut, W.; Artaxo, P.; Andreae, M.O. Refractive Index of Aerosol Particles over the Amazon Tropical Forest during LBA-EUSTACH 1999. *J. Aerosol Sci.* **2003**, *34*, 883–907. [CrossRef]

67. Hoyningen-Huene, W.; von Schmidt, T.; Schienbein, S.; Kee, C.A.; Tick, L.J. Climate-Relevant Aerosol Parameters of South-East-Asian Forest Fire Haze. *Atmos. Environ.* **1999**, *33*, 3183–3190. [CrossRef]
68. Yamasoe, M.A.; Kaufman, Y.J.; Dubovik, O.; Remer, L.A.; Holben, B.N.; Artaxo, P. Retrieval of the Real Part of the Refractive Index of Smoke Particles from Sun/Sky Measurements during SCAR-B. *J. Geophys. Res. Atmos.* **1998**, *103*, 31893–31902. [CrossRef]
69. Schuster, G.L.; Dubovik, O.; Holben, B.N. Angstrom Exponent and Bimodal Aerosol Size Distributions. *J. Geophys. Res. Atmos.* **2006**, *111*. [CrossRef]
70. Veselovskii, I.; Hu, Q.; Goloub, P.; Podvin, T.; Choël, M.; Visez, N.; Korenskiy, M. Mie–Raman–Fluorescence Lidar Observations of Aerosols during Pollen Season in the North of France. *Atmos. Meas. Tech.* **2021**, *14*, 4773–4786. [CrossRef]
71. Ducos, F.; Hu, Q.; Popovici, I. AUSTRAL User's Guide—AUtomated Server for the TReatment of Atmospheric Lidars. 2022. Available online: https://www.researchgate.net/publication/362155848_AUSTRAL_User's_Guide_AUtomated_Server_for_the_TReatment_of_Atmospheric_Lidars (accessed on 30 August 2022).