https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

Reconstructing the Full-Physics Model with Machine Learning for Aerosol Composition Retrieval

Denghui Ji[1], Xiaoyu Sun[1], Christoph Ritter[2], and Justus Notholt[1]

[1]Institute of Environmental Physics, University of Bremen, Otto-Hahn-Allee 1, 28359 Bremen, Germany
[2]Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, Telegrafenberg A43, 14473 Potsdam, Germany

Correspondence: Xiaoyu Sun (xiaoyu_sun@iup.physik.uni-bremen.de)

Abstract. Accurate aerosol composition retrievals support radiative forcing assessment, source attribution, air quality analysis, and improved modeling of aerosol–cloud–radiation interactions. Aerosol retrievals based solely on visible-wavelength aerosol optical depth (AOD) observations provide limited spectral sensitivity, which may be insufficient to reliably distinguish among aerosol types with similar optical properties. In this study, we present a new retrieval framework that combines multi-wavelength AOD observations from both the visible and infrared spectrum, enhancing aerosol type discrimination. A neural network forward model trained on simulations from the Model for Optical Properties of Aerosols and Clouds (MOPSMAP), which relates aerosol optical properties to spectral AOD, is embedded in an optimal estimation method (OEM) to retrieve aerosol composition. This machine learning-based forward model achieves computational efficiency without making compromises in accuracy. The neural network forward model achieves a mean R^2 of 0.99 with root-mean-square error below 0.01. The retrieval resolves up to four independent aerosol components, with degrees of freedom for signal about 3.75. In the total retrieval uncertainty, the forward model contributes less than 10%, confirming its robustness. We apply this hybrid method to ground-based observations, including data from the Aerosol Robotic Network (AERONET) and Fourier Transform Infrared spectrometer (FTIR) measurements. The retrieved aerosol compositions are consistent with physical expectations and validated through backward trajectory analysis. Furthermore, we successfully apply this method to satellite AOD observations and demonstrate its potential for global aerosol composition retrievals. The full development of a global dataset will be further addressed in future work.

1 Introduction

Aerosols play an important role in the climate system by influencing the Earth's radiation budget (Kuniyal and Guleria, 2019; Haywood, 2021), cloud microphysics (Mauritsen et al., 2011; Gong et al., 2023), and air quality (Garrett and Zhao, 2006). Depending on their properties, aerosols can either cool the Earth's surface by reflecting incoming solar radiation (Charlson and Wigley, 1994; Chang et al., 2022), or warm the atmosphere by absorbing sunlight (Weinbruch et al., 2012; Bond et al.,

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

2013; Breider et al., 2014; Groot Zwaaftink et al., 2016; Kodros, 2018). While the net global effect of aerosols is cooling, their climatic impact varies significantly with aerosol type, spatial distribution, and environmental conditions (Kaufman et al., 2002; Satheesh and Moorthy, 2005). For example, strongly scattering aerosols such as sulfate and sea salt typically have a cooling effect in lower-latitude regions. However, in the Arctic, sea salt aerosols can undergo hygroscopic growth under Arctic humidity, which enhances their infrared radiative properties (Ji et al., 2025) and potentially contributes to the longwave cloud radiative effects (Gong et al., 2023). These complexities highlight the importance of accurately observing aerosol microphysical and optical properties in different environmental conditions to better quantify the impact of aerosols on climate and improve the performance of climate models.

Aerosol optical depth (AOD) is a fundamental parameter used to describe the column-integrated extinction of solar radiation due to aerosols. AOD retrievals can be obtained through both active and passive remote sensing techniques. Active remote sensing methods, such as lidar, provide vertically resolved aerosol properties and have been widely used on both ground-based and satellite platforms (Jin et al., 2020; Floutsi et al., 2023). For instance, the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) mission provides detailed aerosol vertical distributions, offering crucial insights into aerosol transport and layering (Winker et al., 2007; Liu et al., 2009). Passive remote sensing, on the other hand, relies on the measurement of scattered and absorbed radiation and includes both satellite-based and ground-based instruments. Satellite sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) retrieve AOD on global scales using multi-spectral radiance measurements (Levy et al., 2007), while ground-based networks like the AErosol RObotic NETwork (AERONET) provide high-accuracy AOD measurements at multiple wavelengths through sun photometry (Holben et al., 1998; Giles et al., 2019).

Despite the abundance of AOD observations, retrieving aerosol composition from remote sensing remains challenging. Recent studies have introduced new methods for retrieving aerosol composition (Li et al., 2019; Ji et al., 2023). In particular, Fourier Transform Infrared spectrometer (FTIR) has been successfully employed to extract aerosol component information from infrared emission spectra (Ji et al., 2023). This method provides valuable insights into aerosol microphysical and chemical properties. The incorporation of infrared spectral information into aerosol retrieval algorithms offers a promising method for improving the accuracy of aerosol composition estimation. Barreto et al. (2020) and Alvárez et al. (2023) have established a detailed observation framework that combines AERONET and FTIR measurements to obtain aerosol AOD spectra spanning both visible and infrared wavelengths. Despite the availability of such comprehensive spectral observations, no existing retrieval algorithm has been developed to infer aerosol composition based on joint visible–infrared AOD data. This study aims to fill that gap.

In aerosol remote sensing, radiative transfer models and Mie scattering calculations are fundamental to developing a full-physics retrieval algorithm. For example, MODIS aerosol retrievals use look-up tables based on radiative transfer simulations (Levy et al., 2007), while AERONET applies a detailed multi-wavelength approach to observe aerosol size distribution and refractive index (Giles et al., 2019). However, the complex dependence of aerosol optical properties on size distribution, composition, relative humidity, and multiple scattering introduces strong nonlinearity into the aerosol retrievals, making traditional retrievals computationally intensive and challenging to optimize. To address these challenges, machine learning (ML) methods

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

have emerged as promising alternatives, offering the potential to approximate the nonlinear mappings between aerosol properties and observations more efficiently while retaining the underlying physical constraints learned from full-physics simulations.

In recent years, machine learning techniques have been widely explored to enhance remote sensing retrievals, offering substantial improvements in efficiency and data assimilation (Cobb et al., 2019; Himes et al., 2020; Doicu et al., 2021; Tian and Shi, 2022; Li et al., 2023). However, ML models are often criticized for their lack of physical interpretability, functioning as "black-box" algorithms without explicit ties to underlying atmospheric physics. Despite these limitations, some studies have demonstrated the potential of ML to replace specific components of physical models (Himes et al., 2020). For example, a hybrid radiative transfer and transfer learning framework is proposed to retrieve aerosol optical depth and fine-mode fraction from multi-spectral geostationary satellite data (Tang et al., 2025). Additionally, neural network–based retrieval approaches using TROPOMI O_2 A-band spectra have been developed for aerosol parameter inference (Rao et al., 2022), and radiative transfer emulators have been integrated into TROPOMI aerosol layer height algorithms (Nanda et al., 2019). These studies support the feasibility and growing applicability of machine learning in aerosol remote sensing.

The ultimate goal of this study is to develop an algorithm to retrieve global aerosol compositions from AOD observations at visible and infrared wavelengths. Traditionally, this retrieval relies on constructing a relationship between AOD and aerosol composition using full-physics models, such as the Model of Optical Properties of Aerosols and Clouds (MOPSMAP, (Gasteiger and Wiegner, 2018)). In this study, MOPSMAP is used to generate a training dataset, and a ML model is trained to capture the mapping between AOD and aerosol composition. The trained ML model then serves as a forward model, replacing the traditional physical model in the inversion process. This approach can approximate the full-physics forward model with a faster, data-driven algorithm that can be applied globally.

In Section 2, we describe aerosol datasets used in this study, including both ground-based measurements and satellite observations. We present the construction steps of the ML database, the training process, and how it is integrated into the retrieval algorithm in Sect. 3. Section 4 presents the results, followed by a discussion on the implications and limitations of the proposed approach.

2 Data

2.1 Multi-band AOD Measurements from AERONET and FTIR

In this study, ground-based measurements are conducted in Ny-Ålesund (11.5° E, 78.9° N), including a sun photometer (AERONET) and a Fourier Transform Infrared spectrometer. The FTIR system (Notholt et al., 1995) is a Bruker 120HR instrument operated as part of the Network for the Detection of Atmospheric Composition Change (NDACC). FTIR leads high resolution, 0.0035 cm^{-1}, spectra in infrared. Barreto et al. (2020) and Alvárez et al. (2023) provide a detailed methodology, the Langley calibration method, for measuring AOD using the infrared spectrum from FTIR. Following their approach, this study derives aerosol AOD observations in the infrared spectral range, including 1020.90, 1238.25, 1558.25, 2133.40, 2192.00, and 2314.20 nm.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

The Aerosol Robotic Network (AERONET) is a global ground-based remote sensing network, providing long-term, high-quality AOD measurements across multiple wavelengths (Floutsi et al., 2023). Standard AERONET sun photometers retrieve AOD at 340, 380, 440, 500, 675, 870, 1020, and 1640 nm, covering the ultraviolet (UV) to shortwave infrared (SWIR) range. These multi-wavelength observations enable the derivation of key aerosol properties, including Ångström Exponent, Fine/-Coarse Mode Fraction, and Single Scattering Albedo (SSA), which are widely used for satellite validation, climate model evaluation, and aerosol trend analysis.

In Ny-Ålesund, the selected wavelengths start from 440 nm, as shorter wavelengths (340 and 380 nm) are not available. In summary, based on the combined AERONET and FTIR observations, the aerosol optical depth (AOD) retrievals in this study are performed at the following wavelengths: 440, 550, 675, 870, 1020, 1558, and 2192 nm.

2.2 Multi-band AOD Measurements from Satellite

The Visible Infrared Imaging Radiometer Suite (VIIRS) aboard the Suomi National Polar-orbiting Partnership (SNPP) satellite provides NASA's standard Level-3 monthly Deep Blue aerosol products. These products offer global, gridded measurements of AOD over land and ocean on a $1^\circ \times 1^\circ$ grid. The Deep Blue algorithm, originally applied to the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and the Moderate Resolution Imaging Spectroradiometer (MODIS), has been adapted for VIIRS to ensure continuity in aerosol data records.

The monthly aggregated product (AERDB_D3_VIIRS_SNPP) is derived from the Version-2.0 daily level 3 gridded products (Sayer et al., 2018). It calculates arithmetic mean values from daily data to produce monthly statistics. This dataset has been available since March 1, 2012. Satellite-derived AOD from sensors such as VIIRS and MODIS covers wavelengths at 490, 550, 670, 865, 1240, 1610, and 2250 nm over ocean. While the ground-based AOD measurements, e.g. AERONET and FTIR, cover different wavelengths. In order to facilitate the broad applicability of the machine learning database (given in the next section) to potential future observational datasets, the satellite AOD measurements are interpolated at varying wavelengths employing the Ångström exponent (Ångström, 1929) to align with the spectral bands of the ground-based AOD measurements:

$$\tau(\lambda_{\text{new}}) = \tau(\lambda_{\text{sat}}) \left(\frac{\lambda_{\text{new}}}{\lambda_{\text{sat}}} \right)^{-\alpha} \tag{1}$$

where $\tau(\lambda_{\text{new}})$ is the interpolated AOD at wavelength (440, 550, 675, 870, 1020, 1558, 2192 nm); $\tau(\lambda_{\text{sat}})$ is the AOD at a known wavelength measured by satellites; α is the Ångström exponent, typically calculated using two known AOD values. This approach integrates visible, near-infrared, and infrared observations, providing a continuous aerosol spectral dataset suitable for both current and future remote sensing applications.

3 Method

To retrieve aerosol composition from multi-wavelength AOD measurements, this study develops a hybrid retrieval framework that integrates machine learning with optimal estimation. As shown by the black module labeled "Emulation Forward Model" in Fig. 1, we first reconstruct the physical forward model (MOPSMAP) using a neural network trained on a large synthetic

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

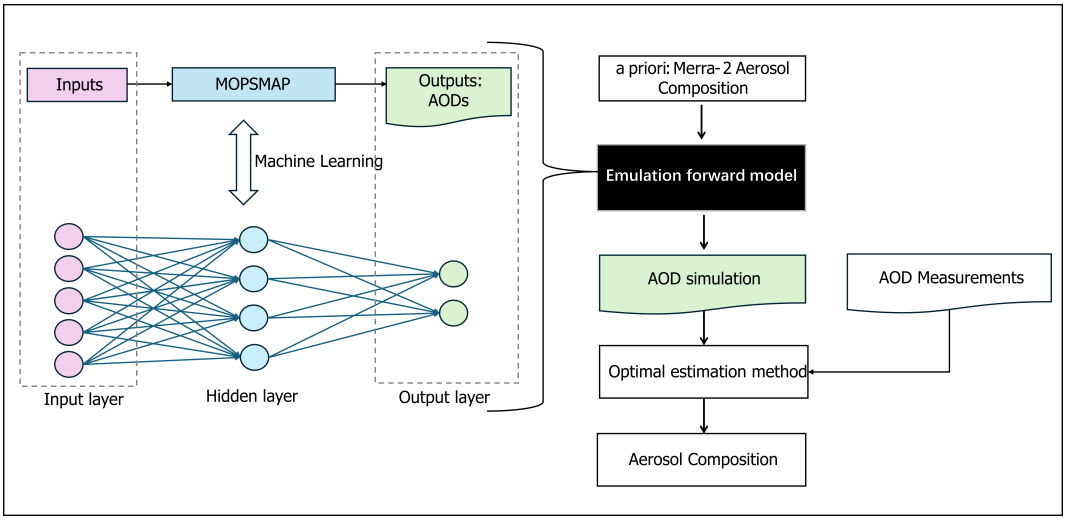

Figure 1. Schematic workflow of the aerosol composition retrieval framework developed in this study. MOPSMAP is used to generate a synthetic aerosol optical property database by varying aerosol composition and microphysical parameters. A neural network model is trained on this database to simulate AOD spectra. During retrieval, multi-band AOD observations and prior composition information (e.g., from MERRA-2 (Gelaro et al., 2017)) are used within an optimal estimation method to derive the final aerosol component fractions.

database of aerosol optical simulations. The input parameters include aerosol component fractions, effective radius, single scattering albedo, and relative humidity, and the outputs are the corresponding AOD spectra across multiple wavelengths. This machine learning model effectively emulates the nonlinear mapping of the physical model while greatly reducing computational cost. Once trained, the neural network replaces the traditional radiative transfer model as the forward model.

In the retrieval step, this machine-learned forward model is embedded into an optimal estimation algorithm to iteratively infer aerosol composition from observed AOD spectra. The retrieval is constrained by both a prior knowledge (e.g., from reanalysis or climatology dataset) and measurement uncertainties. This hybrid approach ensures physical consistency, enhances computational efficiency, and enables flexible application to both ground-based and satellite observations.

To implement the proposed aerosol composition retrieval framework, we follow a structured approach consisting of three main steps. These are outlined as follows and will be described in detail in the following sections:

1. **Synthetic Dataset Generation:** A large AOD dataset is generated using MOPSMAP by varying aerosol component fractions, as well as four physically-constrained parameters: single scattering albedo (SSA), asymmetry factor (AF), effective radius (Reff), and relative humidity (RH). .

2. **Machine Learning Forward Model:** A neural network is trained to emulate the MOPSMAP forward model, mapping input parameters to multi-wavelength AOD spectra.

3. **Retrieval via Optimal Estimation:** The ML-based forward model is integrated into an optimal estimation framework to retrieve aerosol composition from observed AOD.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

3.1 Aerosol Optical Database Simulation with MOPSMAP

The Model for Optical Properties of Aerosols and Clouds (MOPSMAP, (Gasteiger and Wiegner, 2018)) is a numerical tool designed to compute aerosol and cloud optical properties based on Lorenz-Mie theory and the T-matrix method. It supports a wide range of aerosol compositions, including sulfates, sea salt, black carbon, mineral dust, and organic aerosols, with flexible size distributions (e.g., log-normal, gamma) and shape assumptions (spherical and non-spherical particles). Covering a broad spectral range from ultraviolet (UV) to thermal infrared (IR), MOPSMAP provides key optical parameters such as extinction, scattering, and absorption coefficients, single scattering albedo (SSA), asymmetry parameter, and phase functions. These outputs are widely used in radiative transfer simulations, climate models, and remote sensing applications.

To construct a comprehensive dataset for training a machine learning model, we use the MOPSMAP for aerosol optical property simulations. We consider five primary aerosol composition, similar to MERRA-2 (The Modern-Era Retrospective analysis for Research and Applications, Version 2) reanalysis data (Gelaro et al., 2017): sea salt, sulfate, black carbon, dust, and insoluble aerosols. The size distributions of dry aerosols follow a log-normal distribution, with sea salt and dust ranging from 0.01 to 0.5 μm, while sulfate, black carbon, and insoluble aerosols range from 0.01 to 0.1 μm. Varying proportions of these five aerosol classes are randomly assigned, with their total constrained to sum to 100%. This process is repeated 10,000 times, with each time a randomly varying proportion of five aerosols, to produce a dataset that covered a wide range of aerosol mixtures commonly found in the atmosphere. Hygroscopic growth significantly alters aerosol optical properties, particularly in the infrared spectrum. We incorporate this effect by varying the relative humidity (RH) between 55% and 95% in the simulations. The optical properties of the aerosol mixtures are then computed under these conditions.

In summary, as given in Tab.1, using MOPSMAP, we simulate 10,000 cases, obtaining a set of aerosol optical properties across multiple wavelengths, including: Extinction coefficient at visible and infrared wavelengths (440, 550, 675, 870, 1020, 1558, 2192 nm), Single scattering albedo (SSA), Asymmetry factor (Af) and Effective radius (Reff). Given the complexity of the full-physics model MOPSMAP, a machine learning (ML) approach is introduced to simplify the retrieval process while maintaining accuracy. Instead of directly running MOPSMAP simulations for each retrieval, a neural network model is trained on the dataset generated by MOPSMAP, effectively learning the relationship between aerosol microphysical properties and their optical characteristics. The script used to generate this database is available in Section data and code.

3.2 Neural Network Model Training

Directly using MOPSMAP in the retrieval is challenging, likely due to the high dimensionality of its input parameters and the strong nonlinearity in the model. Therefore, to replace MOPSMAP as a forward model, we develop a machine learning framework that learns the relationship between aerosol composition and its optical properties. The trained model allows for rapid calculations in retrieval applications. All detailed procedures and comments are available in the accompanying code repository (see "code data availability"). Below, we provide a brief summary of the training workflow.

We adopt a fully connected feed-forward neural network with two hidden layers. The architecture is defined as:

$$\text{Input}\ (9) \rightarrow \text{FC}(32) \rightarrow \text{ReLU} \rightarrow \text{FC}(32) \rightarrow \text{ReLU} \rightarrow \text{FC}(9) \rightarrow \text{Output (AOD)} \tag{2}$$

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

Table 1. Overview of input parameters and simulated outputs used in the MOPSMAP-based aerosol optical property database.

Category	Parameter	Description
Input Parameters	Aerosol Composition	Fractions of sea salt, sulfate, black carbon, dust, and insoluble; sum to 100%
	Size Distribution	Log-normal; sea salt and dust: 0.01–0.5 µm; others: 0.01–0.1 µm (dry mode)
	Relative Humidity (RH)	Varied from 55% to 95%, affecting wet particle size
	Particle Shape	spherical
Output Variables	AOD Spectrum	Extinction AOD at 440, 550, 675, 870, 1020, 1558, and 2192 nm
	Single Scattering Albedo (SSA)	at 440 nm for training
	Asymmetry Factor (AF)	at 440 nm for training
	Effective Radius (Reff)	Wet effective radius derived from size and RH

Here, FC(n) denotes a fully connected layer with n neurons. Each hidden layer is followed by a rectified linear unit (ReLU) activation function, defined as:

$$\text{ReLU}(z) = \max(0, z) \tag{3}$$

ReLU introduces non-linearity into the model, is computationally efficient, and helps mitigate the vanishing gradient problem, enabling effective training of deep networks.

The input features consist of 5 aerosol component fractions (sea salt, sulfate, black carbon, dust, and insoluble), along with 4 physically-constrained parameters: SSA, AF, Reff, and RH. The output labels are the corresponding AODs at eight wavelengths: 440, 500, 550, 675, 870, 1020, 1558, and 2192 nm.

To ensure generalization and avoid overfitting, the dataset is split as:

$$\text{Train} : \text{Validation} : \text{Test} = 70\% : 15\% : 15\% \tag{4}$$

The model is trained using the Adam optimizer with a learning rate of 10^{-3} and batch size of 64. Performance is evaluated using the mean squared error (MSE) as the loss function:

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{N}(Y_{\text{true}}^{(i)} - Y_{\text{pred}}^{(i)})^2 \tag{5}$$

Model accuracy is assessed with the root mean square error (RMSE) and coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum(Y_{\text{true}} - Y_{\text{pred}})^2}{\sum(Y_{\text{true}} - \bar{Y})^2} \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum(Y_{\text{true}} - Y_{\text{pred}})^2} \tag{7}$$

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

By using this machine learning-based forward model, we achieve a computationally efficient alternative to MOPSMAP, making it feasible for large-scale aerosol composition retrievals from both ground-based and satellite measurements. This approach not only reduces computational cost but also preserves the essential physical relationships governing aerosol optical properties, enabling large-scale and physically consistent aerosol composition retrievals in the subsequent optimal estimation framework.

3.3 Aerosol Composition Retrieval Using Optimal Estimation

To retrieve aerosol composition from multi-wavelength AOD measurements, we apply the optimal estimation method (OEM) (Rodgers, 2000). The key idea is to iteratively adjust the aerosol composition vector until the simulated AOD spectrum matches the observed one, under physical constraints provided by prior knowledge and measurement uncertainty.

In traditional full-physics approaches, such as using MOPSMAP directly, the state vector $\mathbf{x}$ may include high-dimensional microphysical properties like aerosol size distribution, number concentration, and refractive index for each aerosol component:

$$\mathbf{x} = [n_1(r), N_1, \ldots, n_5(r), N_5]^T \tag{8}$$

where n_i and N_i are the aerosol size distribution and number concentration of the five aerosol types: sea salt, sulfate, soot, dust, and insoluble aerosols. $n_i(r)$ denotes the log-normal size distribution for component i, defined as:

$$n_i(r) = \frac{N_i}{\sqrt{2\pi}, \ln\sigma, r} \exp\left[-\frac{1}{2}\left(\frac{\ln r - \ln r_{\text{mod}}}{\ln\sigma}\right)^2\right] \tag{9}$$

r_{mod} and σ are internal parameters of the log-normal distribution of the aerosol. The MOPSMAP captures physical details in aerosol optical properties but could result in ill-posed inverse problems. In practice, we have also implemented a full-physics optimal estimation algorithm based on MOPSMAP directly; however, due to poor retrieval convergence and unstable performance, this approach is not further considered in the current study.

However, if we have reconstructed MOPSMAP using machine learning, the trained model can greatly simplify the input parameters and can guarantee the accuracy of the output aerosol AOD simulation, then the state vector can be reduced from nearly ten dimensions to five dimensions:

$$\mathbf{x} = [N_1, N_2, N_3, N_4, N_5]^T \tag{10}$$

ML makes a clever connection between all input parameters to have only 5 parameters, which is easier to converge. Thus, the nonlinearity of the inversion process can be reduced, and the accuracy and speed of the inversion can be improved. This model approximates the forward mapping from aerosol composition to AOD as:

$$\mathbf{y} = f(\mathbf{x}; \boldsymbol{\theta}) \tag{11}$$

where $\mathbf{x} \in \mathbb{R}^5$ represents the aerosol component fractions (sea salt, sulfate, BC, dust, and insoluble), and $\boldsymbol{\theta} = \{\text{SSA}, \text{AF}, \text{Reff}, \text{RH}\}$ are fixed auxiliary parameters that encode environmental and optical conditions. The model f is learned from a large MOPSMAP-generated dataset and replaces the computationally intensive radiative transfer step.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

The OEM retrieves $\mathbf{x}$ by minimizing a cost function that balances fidelity to the observed AOD spectrum, $\mathbf{y}_{\text{obs}}$, and deviation from a prior estimate $\mathbf{x}_a$:

$$J(\mathbf{x}) = (\mathbf{y}_{\text{obs}} - f(\mathbf{x};\boldsymbol{\theta}))^T \mathbf{S}y^{-1} (\mathbf{y}_{\text{obs}} - f(\mathbf{x};\boldsymbol{\theta})) + (\mathbf{x} - \mathbf{x}_a)^T \mathbf{S}_a^{-1} (\mathbf{x} - \mathbf{x}_a) \tag{12}$$

Here, $\mathbf{S}_a$ is the a prior covariance matrix, and $\mathbf{S}_y$ is the observation error covariance derived from AOD measurement uncertainty.

The state vector is updated iteratively using the Gauss-Newton method. The Jacobian matrix $\mathbf{K}$, representing the sensitivity of AOD to changes in aerosol components, is numerically computed via finite differences:

$$\mathbf{K} = \frac{\partial f(\mathbf{x};\boldsymbol{\theta})}{\partial \mathbf{x}} \tag{13}$$

The gain matrix $\mathbf{G}$ and update equation are:

$$\mathbf{G} = (\mathbf{K}^T \mathbf{S}_y^{-1} \mathbf{K} + \mathbf{S}_a^{-1})^{-1} \mathbf{K}^T \mathbf{S}y^{-1} \tag{14}$$

$$\mathbf{x}_{n+1} = \mathbf{x}_n + \mathbf{G}(\mathbf{y}_{\text{obs}} - f(\mathbf{x}_n;\boldsymbol{\theta})) \tag{15}$$

where n is the iteration index.

This hybrid retrieval framework reduces computational cost and avoids non-convergence issues common in full-physics OEMs, while maintaining physical realism through the machine-learned forward operator and inclusion of environmental parameters as constraints.

3.4 Uncertainty Analysis

As we mentioned before, a comprehensive virtual database is constructed, covering a wide range of aerosol compositions. To quantitatively assess retrieval uncertainty, we avoid relying solely on limited ground-based observations, which may not be representative. Instead, we randomly select 1,500 cases from our original database of 10,000 synthetic database, and use the corresponding AOD spectrum as the synthetic AOD observation. For each selected sample, we perturb the aerosol component by 10% (acted as a priori), followed by normalization. Subsequently, these 1,500 cases are processed through the full retrieval procedure. This experimental configuration facilitates a systematic and controlled evaluation of the retrieval algorithm under diverse aerosol scenarios, thereby supporting a robust assessment of error characteristics and retrieval performance.

The posterior error covariance matrix $\mathbf{S}_{\text{post}}$ for each sample is given by:

$$\mathbf{S}_{\text{post}} = \left(\mathbf{K}^{\mathrm{T}} (\mathbf{S}_y + \mathbf{S}_f)^{-1} \mathbf{K} + \mathbf{S}_a^{-1}\right)^{-1} \tag{16}$$

where $\mathbf{K}$ is the Jacobian matrix, estimated numerically by finite perturbations;$\mathbf{S}_y$ is the measurement error covariance matrix; $\mathbf{S}_f$ is the forward model error covariance matrix, estimated as the squared residuals between simulated and predicted AOD spectra, $\mathbf{S}_a$ is the prior error covariance matrix. The total retrieval uncertainty for each aerosol type is decomposed into three components: observation error contribution, a prior error contribution, and forward model error contribution. Finally, the result averaged over the 1500 cases allows us to quantify the dominant sources of uncertainty for each aerosol component.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

To further assess the information content of the retrieval system, we calculate the Averaging Kernel (AVK) matrix $\mathbf{A}$, defined as:

$$\mathbf{A} = \frac{\partial \mathbf{x}_{\text{retrieved}}}{\partial \mathbf{x}_{\text{true}}} = \mathbf{GK} \tag{17}$$

The diagonal elements of the AVK matrix indicate the degree to which each aerosol component is constrained by the observations. The trace of the AVK matrix gives the Degrees of Freedom for Signal (DoF):

$$\text{DoF} = \text{trace}(\mathbf{A}) \tag{18}$$

indicating how much independent information is effectively retrieved from the measurement. This highlights the potential and limitation of our retrieval algorithm in distinguishing aerosol types under realistic error assumptions.

4 Results

4.1 Characterization of the Synthetic Aerosol Optical Database

Figure 2 shows the mean normalized AOD spectra simulated by MOPSMAP for five pure aerosol types: sea salt, sulfate, black carbon (BC), dust, and insoluble aerosols. Each case corresponds to an idealized scenario in which a single aerosol type dominates (100% composition), allowing a clear examination of spectral distinctions. The properties on the aerosol classes, e.g. complex refractive indices of different aerosol components, are based on the OPAC (Optical Properties of Aerosols and Clouds), which provides standard optical properties for atmospheric aerosols under diverse environmental conditions (Hess et al., 1998).

All spectra are normalized to an AOD of 0.1 at 440 nm. However, different aerosol types exhibit distinct spectral shapes, particularly in the infrared range. Sea salt shows the flattest spectral curve, maintaining relatively high AOD values across infrared wavelengths (e.g., 1.5–2.2µm), consistent with its coarse-mode size distribution and strong infrared extinction. In contrast, sulfate exhibits the steepest decline in AOD with wavelength, indicative of its fine-mode nature and low absorption.

Although BC, dust, and insoluble aerosols display very similar behavior in the visible range (440–870 nm), their differences become more distinguishable in the infrared. Dust retains slightly higher AOD values beyond 1.5µm due to its scattering efficiency at longer wavelengths, while BC and insoluble aerosols separate further in the 1.5 - 2 µm range. These distinctions demonstrate the added value of infrared spectral information for differentiating aerosol types with similar visible characteristics. This confirms the potential of combining visible and infrared AOD in aerosol composition retrieval.

To better understand the diversity and coverage of the training dataset, we visualize the distributions of auxiliary parameters derived from the full synthetic database (all 10000 cases). Figure 3 shows the distributions of three key physical parameters used as auxiliary inputs in the machine learning model: single scattering albedo (SSA), asymmetry factor (AF), and effective radius (Reff). These parameters are evaluated at 440 nm. The SSA histogram reveals a strong right-skewed distribution with a mean value of 0.94, suggesting most aerosols in the database are weakly absorbing. The AF histogram, centered around 0.70, reflects the forward-scattering nature of the aerosol mixtures. The Reff distribution peaks near 0.25 µm and spans from 0.1

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

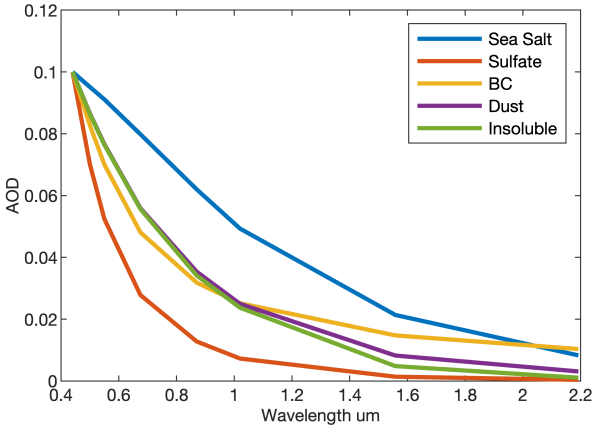

Figure 2. Normalized AOD spectra for five pure aerosol components simulated using MOPSMAP. Each spectrum assumes a single dominant aerosol type (100% composition) with AOD normalized at 440 nm.

to 0.7 μm, consistent with a mix of fine- and coarse-mode particles. These histograms demonstrate that the training database encompasses a wide range of realistic aerosol conditions.

4.2 Neural Network Trained Model vs. MOPSMAP

The machine learning model trained to replace MOPSMAP shows high accuracy in predicting AOD at multiple wavelengths. Figure 4(a-g) present a near-perfect agreement between predicted and original AOD, with R^2 values consistently above 0.99. The best performance is observed at 1.02 μm (R^2 = 0.9964), while all wavelengths exhibit minimal deviation from the 1:1 line, indicating reliable predictions of trained model.

The residual distribution in Fig.4(h) is centered around zero, confirming that prediction errors are symmetrically distributed with no systematic bias. The histogram shows that the majority of residuals remain within ± 0.05, further validating the model's precision.The performance metrics in the table highlight the robustness of the machine learning model. With a mean R^2 of 0.9927 and an RMSE of 0.0072, the model effectively captures the optical properties of aerosols. The low training, validation, and test losses (about 10^{-5}) suggest strong generalization ability, minimizing the risk of overfitting.

Overall, these results confirm that the machine learning model successfully replicates the MOPSMAP simulations, offering an efficient and accurate alternative for forward modeling in aerosol retrieval.

4.3 Retrieval Uncertainty Analysis

As we mentioned in Sec.3.4, to understand the source of retrieval uncertainty, we decompose the total posterior variance into contributions from prior, observation, and forward model errors. To better understand the relative importance of different uncertainty sources in the retrieval, we present their average contributions in Table 2, with further discussion below.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

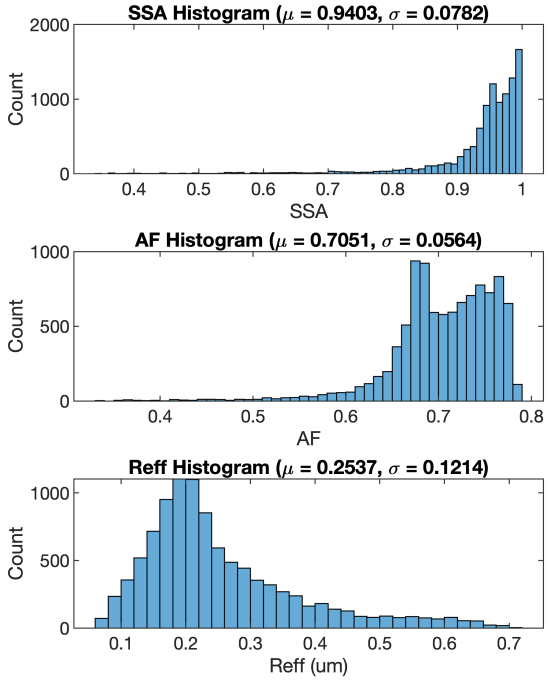

Figure 3. Histograms of key physical parameters in the training dataset: (top) Single scattering albedo (SSA), (middle) Asymmetry factor (AF), and (bottom) Effective radius (Reff), all evaluated at 440 nm. The mean and standard deviation of each parameter are indicated.

Table 2 summarizes the normalized contributions to the total retrieval uncertainty from a prior, observation, and forward model errors for each aerosol component, based on the 1500-case ensemble introduced in Section 3.4. The results highlight distinct sensitivities across aerosol types. For sea salt and sulfate, the a prior and observation contribute comparably (e.g., 53.8% vs. 39.4% for sea salt), indicating that these components are well constrained by the AOD spectral information. In contrast, black carbon retrieval is heavily dependent on a prior assumptions, with 93.4% of the uncertainty attributed to the a priori, reflecting its relatively weak spectral signature in the AOD spectrum. Dust and insoluble aerosols fall in between, with both a prior and observational constraints playing meaningful roles.

Importantly, the contribution from the forward model error remains below 10% for all aerosol types, confirming the stability and reliability of the machine-learning-based forward model used in this study. These findings underscore the benefits of combining physically consistent training datasets with efficient retrieval algorithms, enabling robust composition inference while keeping model-induced uncertainty low.

To quantify the information content of the retrieval, we compute the averaging kernel matrix $\mathbf{A}$. The diagonal elements of $\mathbf{A}$ reflect the sensitivity of each retrieved parameter to the observations. A value close to 1 indicates strong observational constraint, while values near 0 suggest the solution is mainly determined by the a prior. The averaged averaging kernel matrix

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

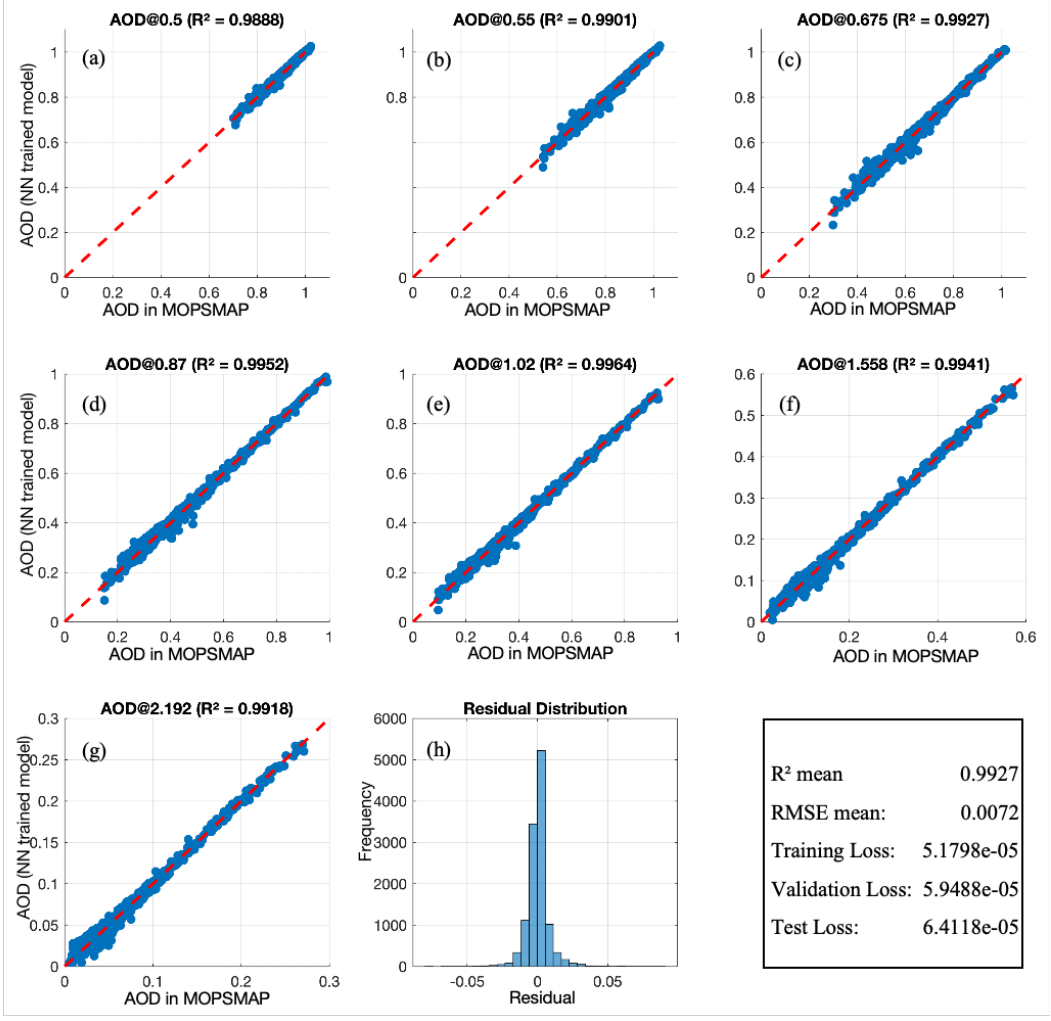

Figure 4. Performance evaluation of the machine learning model replacing MOPSMAP as the forward model. Figures (a-g) show scatter plots comparing the predicted AOD from machine learning model with the AOD from MOPSMAP at different wavelengths (0.5, 0.55, 0.675, 0.87, 1.02, 1.558, and 2.192 μm). The red dashed line represents the 1:1 reference line. Figure (h) displays the residual distribution of the predicted AOD values. The table summarizes key performance metrics, including the mean R^2, RMSE, and loss values for training, validation, and testing.

obtained from 1500 synthetic test cases is:

$$\mathbf{A} = \begin{bmatrix} 0.8462 & -0.0007 & 0.1459 & -0.0057 & -0.0264 \\ -0.0007 & 0.8906 & 0.1762 & -0.0503 & -0.0298 \\ 0.1459 & 0.1762 & 0.2671 & 0.2338 & 0.1492 \\ -0.0057 & -0.0503 & 0.2338 & 0.8564 & 0.0022 \\ -0.0264 & -0.0298 & 0.1492 & 0.0022 & 0.8893 \end{bmatrix}$$

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

Table 2. Normalized contributions (%) to the retrieval uncertainty for each aerosol component from prior, observation, and forward model error sources.

Component	a Prior (%)	Observation (%)	Model (%)
Sea Salt	53.8	39.4	6.8
Sulfate	57.6	36.0	6.4
Black Carbon	93.4	5.4	1.2
Dust	61.2	33.1	5.7
Insoluble	49.5	43.5	7.0

The diagonal values indicate high sensitivity for sea salt ($A_{11} = 0.85$), sulfate ($A_{22} = 0.89$), dust ($A_{44} = 0.86$), and insoluble aerosols ($A_{55} = 0.89$), while black carbon is less constrained ($A_{33} = 0.27$). The total degrees of freedom for signal (DoF), given by $\mathrm{trace}(\mathbf{A})$, is 3.75, indicating that approximately 4 independent parameters can be resolved from the measurement.

4.4 Aerosol Composition Retrieval from Ground-Based Observations

Figure 5 presents the retrieved aerosol composition and corresponding AOD spectral fit at Ny-Ålesund on 21 April 2020. The retrieval results indicate that sea salt, sulfate, and black carbon aerosols dominate during this aerosol event. Specifically, sea salt constitutes the largest fraction (~43%), followed by black carbon (~33%) and sulfate (~24%). Dust and insoluble aerosols contribute minimally ($< 1\%$). The observed AOD spectrum is well constructed by the forward model (Fig. 5b), with residual differences typically below 0.005 (Fig. 5c).

To further assess the potential source regions of the retrieved aerosols, a 120-hour backward trajectory analysis is conducted using the HYSPLIT model (Fig. 5d). Based on the HYSPLIT back-trajectory analysis, the air masses (below 1500 m) are mainly originated from the ocean. Specifically, both 500 m and 1500 m trajectories indicate that, two days earlier (on 19 April), vertical lifting of air masses from the open ocean region between Canada and Greenland likely introduced sea salt aerosols into the lower troposphere, subsequently reaching Ny-Ålesund. Sea salt has been released in the atmosphere in the lowest 500 m between northeast Greenland on the last day prior to advection towards Ny-Ålesund. This transport pattern supports the presence of sea salt in the retrieved result. The upper-level trajectory (around 3000 m altitude) originates near the US-Canada border, suggesting sulfate and black carbon aerosols transported over longer distances (approximately five days) from anthropogenic sources in North America. These trajectories support the retrieved aerosol composition, confirming sea salt dominance from lower-altitude oceanic pathways and sulfate and black carbon from long-range transport at higher altitudes. Overall, this retrieval approach, integrating machine learning and optimal estimation, successfully captures aerosol composition with high accuracy and consistency between observed and modeled AOD spectra.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

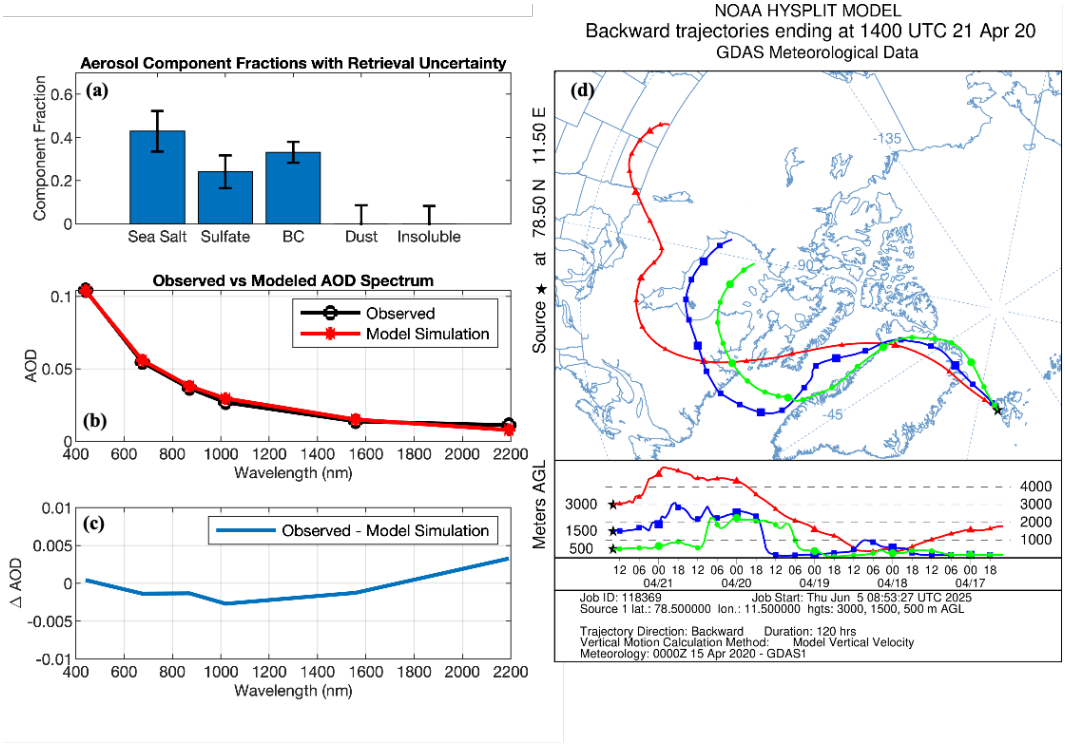

Figure 5. Aerosol composition retrieval from ground-based observations at Ny-Ålesund on 21 April 2020. (a) Retrieved aerosol component fractions (sea salt, sulfate, black carbon, dust, and insoluble aerosols) with uncertainties. (b) Observed versus modeled AOD spectra. (c) Residual differences between observed and simulated AOD. (d) 120-hour backward trajectories arriving at Ny-Ålesund (78.9°N, 11.9°E) at altitudes of 500 m, 1500 m, and 3000 m above ground level, computed using the NOAA HYSPLIT model (Stein et al., 2015).

4.5 Aerosol Composition from Satellite Observations

Based on the previous results using ground-based observations, we have demonstrated that machine learning models can yield reasonably accurate retrievals. However, for satellite-based applications, the available observations are currently limited to the AOD spectrum, without corresponding measurements of additional physical parameters such as SSA, asymmetry factor, effective radius, and relative humidity. To address this limitation, we consider two possible strategies:

(1) supplementing the satellite AOD data with physical parameters retrieved from other satellite products to form a comprehensive input for the retrieval algorithm;

(2) treating the physical constraints $\boldsymbol{\theta}$ in the forward model $F(\mathbf{x};\boldsymbol{\theta})$ as part of the state vector, thereby incorporating them directly into the retrieval.

The first approach allows for more accurate spectral simulation by incorporating multiple observational constraints, leading to improved physical consistency. However, it requires access to diverse datasets, which increases complexity. Moreover, satellite-based products for these additional physical constraints are not yet available as a consolidated database. In contrast, the

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

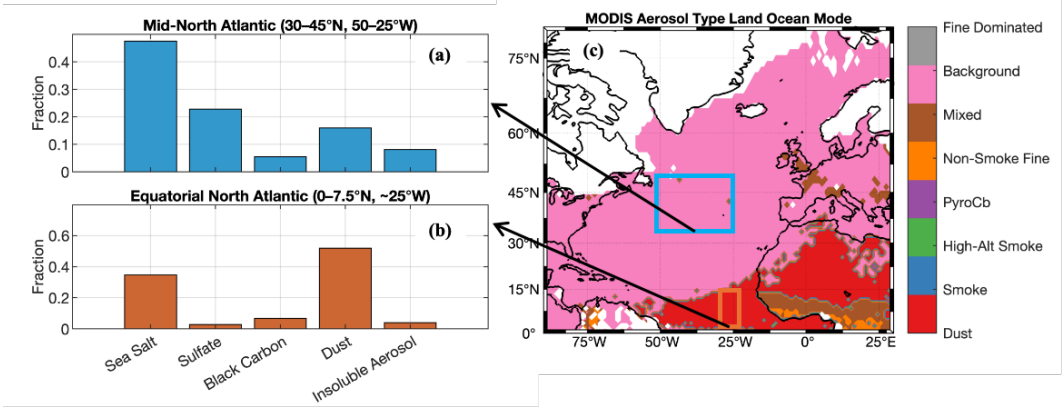

Figure 6. (a) and (b): Retrieved aerosol composition for two MODIS-labeled regions: "Background" (30–45°N, 50–25°W) and "Dust" (0–7.5°N, ∼25°W). Each bar shows the mean aerosol component fraction with labeled regions. (c) MODIS aerosol classification map.

second approach relies solely on AOD spectral observations, offering a easier solution. Given the complexity of developing a multi-source parameter dataset, this study adopts the second strategy as a preliminary attempt to extend our retrieval framework to satellite observations.

In March 2022, a Saharan dust outbreak transported a large amount of mineral dust across North Atlantic and into the Arctic. This event serves as one of the key reasons why March 2022 is selected as the case study of dust enhancement for demonstrating the aerosol retrieval results in this study. To evaluate the capability of our retrieval method in distinguishing aerosol types over the ocean, we apply the inversion algorithm to MODIS AOD spectral data in March 2022 and focus on the North Atlantic region. The MODIS aerosol classification product (Fig. 6c) provides a reference aerosol classification information, which label most of the mid-latitude North Atlantic as "Background" and the tropical eastern North Atlantic as "Dust". Based on this map, we selected two representative regions for further analysis:

- **Mid-North Atlantic (30–45°N, 50–25°W):** Representing the "Background" class.
- **Equatorial North Atlantic (0–7.5°N, ∼25°W):** Representing the "Dust" class.

The retrieved aerosol composition in these two regions is shown in Fig. 6a–b. Over the mid-latitude North Atlantic (panel a), sea salt dominates with a fraction exceeding 45%, followed by sulfate and a smaller contribution from dust and insoluble aerosols. This composition is consistent with clean marine air masses influenced by westerly flow. In contrast, the equatorial North Atlantic (panel b), located downwind of West Africa, shows a strong dust signal, with dust fractions exceeding 50% and reduced sea salt contributions. This suggests the MODIS "Dust" label is physically supported by our component-resolved retrieval. These results demonstrate the potential of our method to capture the aerosol type over the ocean. Moreover, this approach provides physically interpretable aerosol composition that can complement or refine existing satellite classification products.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

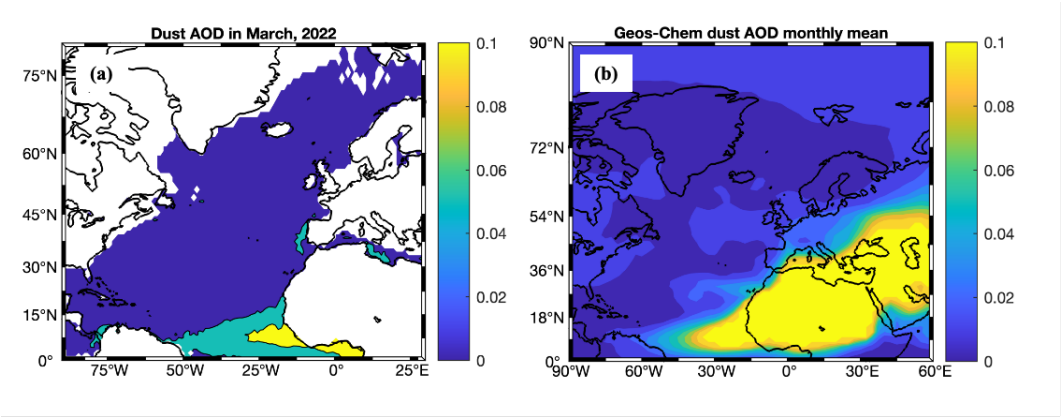

Figure 7. Comparison of dust AOD distributions in March 2022. (a): Retrieved dust AOD based on monthly MODIS AOD measurement; (b): GEOS-Chem model simulated monthly mean dust AOD. The retrieved dust AOD is derived by applying the retrieved dust composition fraction to the total MODIS AOD at 550 nm, thereby providing a dust AOD product.

In addition, we evaluate the spatial distribution of retrieved dust AOD during March 2022 (Fig.7a) and compare it with GEOS-Chem model simulations (Fig.7b, model setups are given in Appendix.A). Both datasets show a consistent dust plume stretching from the Sahara across the tropical Atlantic, with reduced influence over the mid-latitude Atlantic. The similarity in spatial distribution pattern between our retrieved dust AOD and model results confirms that the retrieved dust distribution is physically realistic and consistent with large-scale transport patterns.

5 Conclusions

This study shows the feasibility of integrating machine learning with physically based aerosol modeling to retrieve aerosol composition from multi-wavelength AOD observations. By using a neural network trained on a comprehensive database generated with MOPSMAP, we successfully emulate the aerosol spectral features with aerosol composition and optical properties. The resulting algorithm is easier to converge and efficient, suitable for application across diverse platforms, including ground-based FTIR and AERONET observations as well as satellite-based MODIS and VIIRS data. In addition, this retrieval method is faster (about 5 - 10 times) than traditional full-physics retrieval method, making it a promising tool for large-scale aerosol monitoring. The degrees of freedom for signal (DoF) analysis confirms the robustness of the retrieval framework. The diagonal elements of the averaging kernel matrix show strong observational constraints for sea salt, sulfate, dust, and insoluble aerosols ($A_{11} = 0.85$, $A_{22} = 0.89$, $A_{44} = 0.86$, $A_{55} = 0.89$), while black carbon is less constrained ($A_{33} = 0.27$), highlighting its stronger dependence on the prior. The total DoF of 3.75 suggests that approximately four independent aerosol parameters can be reliably retrieved from the multi-wavelength AOD observations.

However, on satellite platforms, infrared observations over land are still missing due to the strong and variable influence of surface emissivity (Li et al., 2020), which limits the global (land) applicability of this method. Improvements in infrared

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

AOD retrieval, especially over land surfaces, would significantly enhance the accuracy and reliability of aerosol composition inversion. Due to the lack of reliable land-based infrared AOD spectra from satellite observations, this study highlights the importance of ground-based measurements. The combined use of AERONET and FTIR observations helps fill this critical gap, providing high-quality, multi-wavelength AOD data that enable physically consistent and information-rich aerosol composition retrievals.

In summary, the integration of machine learning and multi-band AOD observations presents a promising method for aerosol composition retrieval. Continued efforts to improve infrared AOD accuracy, expand physical realism in training data, and incorporate additional observational constraints such as lidar profiles will be essential for achieving reliable, global-scale aerosol monitoring.

Data availability. The MERRA-2 reanalysis data used as a priori information are available from NASA's GES DISC at https://disc.gsfc.nasa.gov/datasets/M2T1NXAER_5.12.4 (hourly) and https://disc.gsfc.nasa.gov/datasets/M2TMNXAER_5.12.4 (monthly mean). Ground-based aerosol optical depth (AOD) observations are obtained from AERONET (https://aeronet.gsfc.nasa.gov/). The VIIRS Deep Blue Aerosol monthly Level 3 product (AERDB_M3_VIIRS_SNPP, version 2) used in this study is publicly available via NASA LAADS DAAC at https://doi.org/10.5067/VIIRS/AERDB_M3_VIIRS_SNPP.002. The synthetic aerosol dataset used for model training and testing, as well as the code implementing the retrieval algorithm, are available upon request from the corresponding author. FTIR AOD observations are also available upon request from the corresponding author.

Appendix A: GEOS-Chem Model Setups

GEOS-Chem is a global 3D chemical transport model widely applied to simulate atmospheric trace gases and aerosol transport, chemistry, and deposition Bey et al. (2001). Driven by assimilated meteorological data from the Goddard Earth Observing System (GEOS), GEOS-Chem allows detailed studies of atmospheric composition and transport patterns. In this study, we use GEOS-Chem version 13.4.0, driven by MERRA-2 reanalysis data Gelaro et al. (2017), configured with a horizontal resolution of $2^\circ \times 2.5^\circ$ and 47 vertical levels. The simulations spanned from January 1, 2021, to May 31, 2022.

Author contributions. J.D. conceived the study, developed the retrieval framework, conducted the data analysis, and led the manuscript writing. X.S. contributed to algorithm development and data interpretation. C.R. and J.N. reviewed the manuscript and provided feedback. All authors contributed to the final version of the paper.

Competing interests. Justus Notholt is a member of the editorial board of Atmospheric Measurement Techniques. The authors declare that they have no other competing interests.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

Acknowledgements. We gratefully acknowledge the funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 268020496 – TRR 172, within the Transregional Collaborative Research Center "ArctiC Amplification: Climate Relevant Atmospheric and SurfaCe Processes, and Feedback Mechanisms (AC)3". We thank the senate of Bremen for partial funding of this work.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

References

Alvárez, O., Barreto, A., García, O. E., Hase, F., García, R. D., Gröbner, J., León-Luis, S. F., Sepúlveda, E., Carreño, V., Alcántara, A., Ramos, R., Almansa, A. F., Kazadzis, S., Taquet, N., Toledano, C., and Cuevas, E.: Aerosol properties derived from ground-based Fourier transform spectra within the COllaborative Carbon Column Observing Network, Atmospheric Measurement Techniques, 16, 4861–4884, https://doi.org/10.5194/amt-16-4861-2023, 2023.

Barreto, Á., García, O. E., Schneider, M., García, R. D., Hase, F., Sepúlveda, E., Almansa, A. F., Cuevas, E., and Blumenstock, T.: Spectral Aerosol Optical Depth Retrievals by Ground-Based Fourier Transform Infrared Spectrometry, Remote Sensing, 12, https://doi.org/10.3390/rs12193148, 2020.

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., Li, Q., Liu, H. Y., Mickley, L. J., and Schultz, M. G.: Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation, Journal of Geophysical Research: Atmospheres, 106, 23 073–23 095, https://doi.org/https://doi.org/10.1029/2001JD000807, 2001.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., et al.: Bounding the role of black carbon in the climate system: A scientific assessment, J. Geophys. Res.: Atmos., 118, 5380–5552, https://doi.org/10.1002/jgrd.50171, 2013.

Breider, T. J., Mickley, L. J., Jacob, D. J., Wang, Q., Fisher, J. A., Chang, R. Y.-W., and Alexander, B.: Annual distributions and sources of Arctic aerosol components, aerosol optical depth, and aerosol absorption, Journal of Geophysical Research: Atmospheres, 119, 4107–4124, https://doi.org/https://doi.org/10.1002/2013JD020996, 2014.

Chang, E., Yalamanchili, T., and Qiu, C.: Modelling effects of alkylamines on sea salt aerosols using the Extended Aerosols and Inorganics Model, Journal of Emerging Investigators, https://doi.org/10.59720/20-234, 2022.

Charlson, R. and Wigley, T.: Sulfate aerosol and climatic change, Scientific American, 270, 48–57, https://doi.org/10.1038/SCIENTIFICAMERICAN0294-48, 1994.

Cobb, A. D., Himes, M. D., Soboczenski, F., Zorzan, S., O'Beirne, M., Baydin, A. G., Gal, Y., Domagal-Goldman, S., Arney, G., and Angerhausen, D.: An Ensemble of Bayesian Neural Networks for Exoplanetary Atmospheric Retrieval, The Astronomical Journal, 158, https://doi.org/10.3847/1538-3881/ab2390, 2019.

Doicu, A., Doicu, A., Efremenko, D., Loyola, D., and Trautmann, T.: An Overview of Neural Network Methods for Predicting Uncertainty in Atmospheric Remote Sensing, Remote. Sens., 13, 5061, https://doi.org/10.3390/rs13245061, 2021.

Floutsi, A. A., Baars, H., Engelmann, R., Althausen, D., Ansmann, A., Bohlmann, S., Heese, B., Hofer, J., Kanitz, T., Haarig, M., Ohneiser, K., Radenz, M., Seifert, P., Skupin, A., Yin, Z., Abdullaev, S., Komppula, M., Filioglou, M., Giannakaki, E., Stachlewska, I., Janicka, L., Bortoli, D., Marinou, E., Amiridis, V., Gialitaki, A., Mamouri, R., Barja, B., and Wandinger, U.: DeLiAn – a growing collection of depolarization ratio, lidar ratio and Ångström exponent for different aerosol types and mixtures from ground-based lidar observations, Atmospheric Measurement Techniques, https://doi.org/10.5194/amt-16-2353-2023, 2023.

Garrett, T. J. and Zhao, C.: Increased Arctic cloud longwave emissivity associated with pollution from mid-latitudes, Nature, 440, 787–789, https://doi.org/10.1038/nature04636, 2006.

Gasteiger, J. and Wiegner, M.: MOPSMAP v1.0: a versatile tool for the modeling of aerosol optical properties, Geoscientific Model Development, 11, 2739–2762, https://doi.org/10.5194/gmd-11-2739-2018, 2018.

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., da Silva, A. M., Gu, W., Kim, G.-K., Koster, R.,

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

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), J. Clim., 30, 5419 – 5454, https://doi.org/10.1175/JCLI-D-16-0758.1, 2017.

Giles, D. M., Sinyuk, A., Sorokin, M. G., Schafer, J. S., Smirnov, A., Slutsker, I., Eck, T. F., Holben, B. N., Lewis, J. R., Campbell, J. R., Welton, E. J., Korkin, S. V., and Lyapustin, A. I.: Advancements in the Aerosol Robotic Network (AERONET) Version 3 database – automated near-real-time quality control algorithm with improved cloud screening for Sun photometer aerosol optical depth (AOD) measurements, Atmospheric Measurement Techniques, 12, 169–209, https://doi.org/10.5194/amt-12-169-2019, 2019.

Gong, X., Zhang, J., Croft, B., Yang, X., Frey, M. M., Bergner, N., Chang, R. Y.-W., Creamean, J. M., Kuang, C., Martin, R. V., et al.: Arctic warming by abundant fine sea salt aerosols from blowing snow, Nature Geoscience, 16, 768–774, https://doi.org/10.1038/s41561-023-01254-8, 2023.

Groot Zwaaftink, C. D., Grythe, H., Skov, H., and Stohl, A.: Substantial contribution of northern high-latitude sources to mineral dust in the Arctic, J. Geophys. Res.: Atmos., 121, 13,678–13,697, https://doi.org/https://doi.org/10.1002/2016JD025482, 2016.

Haywood, J.: Atmospheric aerosols and their role in climate change, Climate Change, https://doi.org/10.1016/B978-0-444-63524-2.00027-0, 2021.

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.

Himes, M. D., Harrington, J., Cobb, A. D., Baydin, A. G., Soboczenski, F., O'Beirne, M., Zorzan, S., Wright, D., Scheffer, Z., Domagal-Goldman, S., and Arney, G.: Accurate Machine-learning Atmospheric Retrieval via a Neural-network Surrogate Model for Radiative Transfer, The Planetary Science Journal, 3, https://doi.org/10.3847/PSJ/abe3fd, 2020.

Holben, B., Eck, T., Slutsker, I., Tanré, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., 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.

Ji, D., Palm, M., Ritter, C., Richter, P., Sun, X., Buschmann, M., and Notholt, J.: Ground-based remote sensing of aerosol properties using high-resolution infrared emission and lidar observations in the High Arctic., Atmospheric Measurement Techniques, 16, https://doi.org/10.5194/amt-16-1865-2023, 2023.

Ji, D., Palm, M., Buschmann, M., Ebell, K., Maturilli, M., Sun, X., and Notholt, J.: Hygroscopic aerosols amplify longwave downward radiation in the Arctic, Atmospheric Chemistry and Physics, 25, 3889–3904, https://doi.org/10.5194/acp-25-3889-2025, 2025.

Jin, Y., Nishizawa, T., Sugimoto, N., Ishii, S., Aoki, M., Sato, K., and Okamoto, H.: Development of a 355-nm high-spectral-resolution lidar using a scanning Michelson interferometer for aerosol profile measurement., Optics express, 28 16, 23 209–23 222, https://doi.org/10.1364/oe.390987, 2020.

Kaufman, Y., Tanré, D., and Boucher, O.: A satellite view of aerosols in the climate system, Nature, 419, 215–223, https://doi.org/10.1038/nature01091, 2002.

Kodros, J. K.: Size-resolved mixing state of black carbon in the Canadian high Arctic and implications for simulated direct radiative effect, Atmospheric Chemistry and Physics, https://doi.org/10.5194/acp-18-11345-2018, 2018.

Kuniyal, J. and Guleria, R.: The current state of aerosol-radiation interactions: A mini review, Journal of Aerosol Science, https://doi.org/10.1016/J.JAEROSCI.2018.12.010, 2019.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

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

Li, C., Li, J., Dubovik, O., Zeng, Z., and Yung, Y.: Impact of Aerosol Vertical Distribution on Aerosol Optical Depth Retrieval from Passive Satellite Sensors, Remote. Sens., 12, 1524, https://doi.org/10.3390/rs12091524, 2020.

Li, H., Gu, M., Zhang, C., Xie, M., Yang, T., and Hu, Y.: Retrieving Atmospheric Gas Profiles Using FY-3E/HIRAS-II Infrared Hyperspectral Data by Neural Network Approach, Remote. Sens., 15, 2931, https://doi.org/10.3390/rs15112931, 2023.

Li, Z., Xie, Y., Zhang, Y., Li, L., Xu, H., Li, K., and Li, D.: Advance in the remote sensing of atmospheric aerosol composition, National Remote Sensing Bulletin, https://doi.org/10.11834/jrs.20198185, 2019.

Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., Omar, A., Powell, K., Trepte, C., and Hostetler, C.: The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance, Journal of Atmospheric and Oceanic Technology, 26, 1198–1213, https://doi.org/10.1175/2009JTECHA1229.1, 2009.

Mauritsen, T., Sedlar, J., Tjernström, M., Leck, C., Martin, M., Shupe, M., Sjogren, S., Sierau, B., Persson, P., Brooks, I., et al.: An Arctic CCN-limited cloud-aerosol regime, Atmospheric Chemistry and Physics, 11, 165–173, https://doi.org/10.5194/acp-11-165-2011, 2011.

Nanda, S., de Graaf, M., Veefkind, J. P., ter Linden, M., Sneep, M., de Haan, J., and Levelt, P. F.: A neural network radiative transfer model approach applied to the Tropospheric Monitoring Instrument aerosol height algorithm, Atmospheric Measurement Techniques, 12, 6619–6634, https://doi.org/10.5194/amt-12-6619-2019, 2019.

Notholt, J., Meier, A., and Peil, S.: Total column densities of tropospheric and stratospheric trace gases in the undisturbed Arctic summer atmosphere, Journal of Atmospheric Chemistry, 20, 311–332, https://doi.org/10.1007/BF00694500, 1995.

Rao, L., Xu, J., Efremenko, D. S., Loyola, D. G., and Doicu, A.: Aerosol Parameters Retrieval From TROPOMI/S5P Using Physics-Based Neural Networks, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15, 6473–6484, https://doi.org/10.1109/JSTARS.2022.3196843, 2022.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding: Theory and Practice, World Scientific, Singapore, ISBN 978-981-02-2740-1, https://doi.org/10.1142/9789812813718, 2000.

Satheesh, S. and Moorthy, K.: Radiative effects of natural aerosols: A review, Atmospheric Environment, 39, 2089–2110, https://doi.org/10.1016/J.ATMOSENV.2004.12.029, 2005.

Sayer, A., Hsu, N. C., Lee, J., Bettenhausen, C., Kim, W., and Smirnov, A.: Satellite Ocean Aerosol Retrieval (SOAR) Algorithm Extension to S-NPP VIIRS as Part of the "Deep Blue" Aerosol Project, Journal of Geophysical Research: Atmospheres, 123, 380 – 400, https://doi.org/10.1002/2017JD027412, 2018.

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.

Tang, C., Shi, C., Letu, H., Yin, S., Nakajima, T., Sekiguchi, M., Xu, J., Zhao, M., Ma, R., and Wang, W.: Development of a hybrid algorithm for the simultaneous retrieval of aerosol optical thickness and fine-mode fraction from multispectral satellite observation combining radiative transfer and transfer learning approaches, Remote Sensing of Environment, 319, 114 619, https://doi.org/https://doi.org/10.1016/j.rse.2025.114619, 2025.

Tian, J. and Shi, J.: A High-Accuracy and Fast Retrieval Method of Atmospheric Parameters Based on Genetic-BP, IEEE Access, 10, 19 458–19 468, https://doi.org/10.1109/ACCESS.2022.3151868, 2022.

https://doi.org/10.5194/egusphere-2025-3289
Preprint. Discussion started: 15 July 2025

Weinbruch, S., Wiesemann, D., Ebert, M., Schütze, K., Kallenborn, R., and Ström, J.: Chemical composition and sources of aerosol particles at Zeppelin Mountain (Ny Ålesund, Svalbard): An electron microscopy study, Atmospheric Environment, 49, 142–150, https://doi.org/https://doi.org/10.1016/j.atmosenv.2011.12.008, 2012.

Winker, D., Hunt, W., and McGill, M.: Initial performance assessment of CALIOP, Geophysical Research Letters, 34, https://doi.org/10.1029/2007GL030135, 2007.

Ångström, A.: On the atmospheric transmission of sun radiation and on dust in the air, Geografiska Annaler, 11, 156–166, https://doi.org/10.1080/20014422.1929.11880498, 1929.