https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Assessing the limits of lidar aerosol inversions using a probabilistic machine learning approach

Matthew Hayman[1]

[1]Earth Observing Laboratory, NSF National Center for Atmospheric Research, Boulder, CO, USA

Correspondence: Matthew Hayman (mhayman@ucar.edu)

Abstract. Aerosol-cloud interactions represent a major source of uncertainty in climate models and addressing this requires accurate, vertically-resolved observations of aerosol microphysical properties such as effective radius and number concentration. While Raman and High Spectral Resolution lidars provide vital bulk optical measurements, fundamentally they do not measure these quantities directly; retrieving microphysical properties from these observations is a mathematically ill-posed and non-unique inverse problem. Existing retrieval algorithms—including traditional physics-based inversions, semi-empirical parametrization methods, and recent empirical machine learning approaches—typically force a deterministic, single-point solution. This masks fundamental uncertainties and relies on opaque priors. In this study, we introduce a machine learning framework that embraces the inherent non-uniqueness of the lidar inversions. Trained on a dataset of *in situ* aerosol size distributions coupled with Mie theory, our model predicts the Probability Density Function (PDF) of possible aerosol microphysical states rather than a single estimate. We leverage this framework to evaluate the theoretical capabilities and error sensitivities of four distinct lidar architectures, ranging from advanced multi-wavelength Raman configurations ($6\beta+2\alpha$, $3\beta+2\alpha$) to spaceborne-relevant systems ($1\beta+1\alpha$ at 355 nm, 3β). Under idealized, low-noise conditions, advanced multi-wavelength systems best constrain the microphysical solution space. However, under realistic observational uncertainties, the accuracy of these complex architectures degrades, whereas simpler configurations with direct extinction measurements (e.g., $1\beta+1\alpha$) prove significantly more robust. Ultimately, this best-case analysis establishes that lidar has the potential to constrain effective radius to approximately a factor of 2 and concentration to an order of magnitude, demonstrating that probabilistic retrievals can play an important role in providing mathematically transparent observational constraints for global climate models.

1 Introduction

Aerosol-cloud interactions remain one of the most significant sources of uncertainty in global climate models (Patel et al., 2024). Reducing this uncertainty requires vertically resolved profiles of fundamental aerosol microphysical properties—specifically, aerosol number concentration and size (effective radius)—which dictate the ability of particles to act as Cloud Condensation Nuclei (CCN) (Lenhardt et al., 2025). While advanced lidar systems, such as High Spectral Resolution Lidars (HSRL) and Raman lidar, provide essential vertically resolved measurements of aerosol optical properties (backscatter and extinction coefficients), relating these optical observables back to microphysical states requires solving Fredholm integral equations (Veselovskii et al., 2002; Müller et al., 2001). This represents a mathematically ill-posed inverse problem character-

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

ized by a non-unique solution space, making it exceptionally difficult to disentangle particle size from number concentration (Qing et al., 1989; Shifrin and Zolotov, 2003).

Historically, this microphysical inverse problem has been tackled using various physics-based numerical approaches, including Tikhonov regularization (Veselovskii et al., 2002) and lookup tables (LUTs) (Patel et al., 2024). Alternatively, to bypass complex numerical inversions, semi-empirical parameterization methods like the POlarization LIdar PHOtometer Networking (POLIPHON) technique have been widely adopted. POLIPHON separates aerosol components (e.g., dust and non-dust) using depolarization measurements and applies static, aerosol-type-dependent conversion factors—derived from AERONET sun photometer datasets—to map lidar extinction directly to microphysical properties like CCN (Ansmann et al., 2026). While these methods apply necessary mathematical constraints or empirical parameterizations to stabilize the retrieval, they share a fundamental limitation: they ultimately force the algorithm to yield a single, deterministic "best-fit" solution. By outputting only a single estimate, these traditional algorithms and static conversions fail to acknowledge the range of possible physical solutions, masking the true non-uniqueness of the optical inversion problem where entirely different combinations of aerosol size and concentration can yield identical lidar signatures.

Recently, purely data-driven linear regression (Lenhardt et al., 2023) and machine learning (ML) (Redemann and Gao, 2024) paradigms have been proposed to bypass the rigid *a priori* assumptions and mathematical instabilities of traditional physics-based retrievals. Most notably, Redemann and Gao utilize Fully Connected Neural Networks trained directly on collocated airborne lidar and *in situ* measurements to predict higher-level aerosol properties. However, this approach ultimately outputs a single "most likely" estimate that may not accurately reflect the physical state of the observed aerosol layer. Relying entirely on strictly collocated remote and *in situ* measurements also presents practical challenges: training data is constrained by the availability of coordinated flights, resulting in sparse datasets that may under-sample the true variability of aerosol states. Furthermore, it introduces the persistent physical reality that remote and *in situ* instruments rarely measure the exact same air volumes (Redemann and Gao, 2024). Finally, fundamentally, no algorithmic framework can generate missing physical information. Because lidar observations often lack the intrinsic information content required for a unique, deterministic inversion of aerosol properties, ML-based solutions are forced to compensate. When deterministic solutions are impossible, these models naturally default to the statistical distributions of their training data, effectively relying on opaque, implicit priors and hidden assumptions to bridge the information gap.

In contrast to approaches that force deterministic outputs and rely on scarce collocated data, this paper introduces a machine learning framework designed to explicitly embrace the theoretical non-uniqueness of lidar observations while leveraging large training datasets. In essence, rather than trying to find a single answer, this method seeks to define the range of possible answers. Rather than attempting to collocate remote and *in situ* measurements, our approach utilizes datasets of high-resolution *in situ* aerosol size distributions. By combining these distributions with Mie theory to directly forward-calculate the corresponding lidar observables, a more globally representative, physically consistent training dataset is generated comprising millions of observation pairs.

Trained on this massive dataset, this model does not attempt to force a single estimated output. Instead, it is designed to output the full Probability Density Function (PDF) of possible aerosol number concentration and effective radius solutions for

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

a given set of lidar observations (backscatter and extinction at various wavelengths). By predicting the complete distribution of viable microphysical states, this approach respects the fundamental physics of the lidar equations while providing a robust and transparent quantification of retrieval uncertainty.

This study then leverages a machine learning framework to evaluate four distinct observational configurations. Two of these architectures are aligned with instruments used for aerosol retrievals in prior works. The other two are of particular relevance to current and proposed satellite missions.

First, the $6\beta + 2\alpha$ case is analyzed, which utilizes backscatter coefficients at six wavelengths (355, 400, 532, 710, 800, and 1064 nm) and extinction coefficients at 355 and 532 nm. This configuration, described by Veselovskii et al. (2002), represents the "best-case" information content achieved by advanced Raman systems, though it remains a mathematically ill-posed inversion problem when attempting to retrieve full size distributions (Veselovskii et al., 2002). Second, the $3\beta + 2\alpha$ configuration is considered—the standard for contemporary multiwavelength Raman and High Spectral Resolution Lidar (HSRL) systems—as discussed in the case study by Müller et al. (2001).

Recognizing the immediate need for robust retrievals from existing satellite missions, the $1\beta + 1\alpha$ at 355 nm case is also evaluated. This configuration is specifically designed to assess the potential of the European Space Agency (ESA) EarthCARE mission, which employs the ATLID (Atmospheric LIDar) system (Illingworth et al., 2015; Donovan et al., 2024). Finally, a 3β-only case (355, 532 and 1064 nm) is analyzed to evaluate the retrieval potential of proposed NASA mission concepts, such as Fleet for the Atmosphere Linking Commercial Observations with NASA (FALCON)-Lidar, intended as follow-on to the CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) mission. This 3β configuration is particularly relevant for assessing how backscatter-only data from next-generation spaceborne systems can be used to estimate a PDF of microphysical solutions without the benefit of independent extinction measurements.

In testing these four scenarios, the goal is to quantify uncertainty, assess existing best-case data product performance and enable potential design tradeoffs for future lidar development.

1.1 Lidar Instrumentation

This work mostly focuses on retrievals obtained through high spectral resolution lidar (HSRL) or Raman lidar platforms. These instrument architectures have designs that allow well constrained estimates of bulk backscatter and extinction optical properties of an atmospheric volume without resorting to considerable assumptions about the volume under interrogation. Backscatter lidar inversions using the Fernald (Fernald, 1984), Klett (Klett, 1981, 1985) or related inversions rely on assumptions of the atmospheric column and will thus tend to produce biased volumetric retrievals of both variables driven by the bias in the assumed lidar ratios. This is because the extinction and backscatter terms are intractably entangled for each range gate (in effect one measurement with two unknowns). These errors are further compounded by the fact that backscatter lidar cannot directly characterize a number of the instrument terms that affect the retrieval. For example, both Klett and Fernald state that they assume geometric overlap correction is applied to the lidar signals as justification for ommitting the term from their inversion derivations. However there is no robust way to directly measure geometric overlap and errors in this correction will propagate into the inversion. HSRL (Fiocco et al., 1971; Shimizu et al., 1983; She et al., 1992; Piironen and Eloranta, 1994; Eloranta,

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

2005; Hair et al., 2008; Hayman and Spuler, 2017; Burton et al., 2018) and Raman lidar (Cooney et al., 1969; Melfi, 1972; Ansmann et al., 1990, 1992; Whiteman et al., 1992; Whiteman, 2003; Müller et al., 1999) use different optical techniques to obtain quantitative measurements of bulk aerosol backscatter and extinction. In both cases, an additional channel only measures molecular backscatter which acts as a calibration target for each range bin. Because the molecular backscatter and extinction coefficient can be predicted with relatively low error using assumed pressure and temperature profiles, this provides a simple mechanism to convert the backscatter ratio (total backscatter to molecular backscatter) to the volume backscatter coefficient. Within the estimate of backscatter coefficient, a calibration for the differential overlap (the ratio of overlap functions) between the combined (standard backscatter) and molecular channels is needed. HSRL can measure this instrument property directly by either tuning the laser off the aerosol blocking filter, or by removing the optic entirely. Raman lidar typically cannot directly measure differential overlap and must infer it (typically using an assumed clear air profile), which contributes an additional source of error to the retrievals.

Extinction retrievals for either instrument architecture relies on comparing the molecular lidar signal to that expected based on the temperature and pressure profile. Here both techniques suffer from being unable to directly measure the geometric overlap function of the molecular channel which can contribute additional, difficult to trace, error to the retrieval. Various techniques are used to infer the geometric overlap function (e.g. using relatively clear air scenes when available, wide-field-of-view telescopes with rapidly converging overlap functions, very careful instrument alignment with active beam steering) however the estimate and assumptions of the overlap function will likely contribute additional (often not well quantified) error to the extinction retrievals. This means even for high performance HSRL and Raman lidar, extinction measurements of aerosols typically contain relatively high fractions of error which is difficult to quantify. This is particularly true close to the instrument.

Errors due to shot noise are most frequently quantified and reported in lidar data products. However errors can also result from signal processing efforts to suppress shot noise. In addition to imposing direct biases in the data, they may limit the ability to compare extinction and backscatter directly as one product may be smoothed more than the other. Retrievals can also rely on the higher signal-to-noise backscatter estimate to impart the structure onto extinction estimates. Instead of estimating extinction directly, one estimates the lidar ratio (extinction to backscatter ratio) which is assumed to be more slowly varying. This produces better looking products, and has been used to generally produce superior lidar retrievals (Marais et al., 2016; Burton et al., 2025) but it's important to recognize that it also introduces covariance between the backscatter and extinction terms. The analysis presented here assumes independent errors between input variables and these factors may need more careful consideration if the proposed algorithm were deployed operationally.

Finally, it should be noted that error estimates in lidar retrievals can be understated by lidar operators. Reported errors may not consider all possible sources or may implicitly assume exact knowledge of some difficult to know property, particularly avoiding those that are not explicitly called out in retrieval algorithms. The lidar community at large is also constantly discovering new potential error sources. Part of the challenge associated with these lidar retrievals is that they are extremely difficult to directly validate. Performance is often evaluated using path integrated quantities and bootstrapping approaches (evaluating one HSRL/Raman data product against other HSRL/Raman lidar). Finally, bulk validation performance metrics are not necessarily representative of individual observations. A common effect in single value performance metrics is "performance by dilution",

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

where high retrieval performance in relatively common and mundane scenes will tend to dominate bulk metrics and may not apply well to more complex scenes.

We should note that these critiques are not necessarily strictly unique to lidar systems. Many remote sensors leverage assumptions to obtain highly desirable scientific data products without any mechanism to fully test the validity of those assumptions or propagate their potential errors. The focus on uncertainties of data from lidar is largely a product of our own expertise and the problem considered in this work. The omission of other sensor types is a product of the scope of this work, and should not be taken to imply that lidar are inferior to other remote sensors for aerosol observation.

2 The Lidar Inversion Impasse

Inverting aerosol microphysical properties from lidar observations represents an inherently ill-posed problem where the degrees of freedom in the solution space typically exceed the number of independent measurements (Veselovskii et al., 2002). Because lidar measurements are volume-integrated quantities, disentangling particle number density from particle size creates a fundamental impasse in the inversion problem. This theoretical fact, is further verified by analysis performed in observational campaigns (Lenhardt et al., 2025).

The relationship between backscatter (β) and extinction (α) coefficients and the underlying particle properties is described by a set of integral equations (Müller et al., 2001). Physically, the volume backscatter coefficient at a given wavelength (λ) is related to the particle size distribution $n_d(x)$ as follows:

$$\beta(\lambda) = \int_0^\infty \sigma_\pi(x,\lambda) n_d(x) dx \tag{1}$$

where $\sigma_\pi(x,\lambda)$ is the particle backscatter cross-section, which depends on both wavelength and particle properties x (typically reduced to particle radius r, but potentially including shape and complex refractive index). Similarly, the measured extinction is given by:

$$\alpha(\lambda) = \int_0^\infty \sigma_e(x,\lambda) n_d(x) dx \tag{2}$$

where $\sigma_e(x,\lambda)$ is the extinction cross-section. The core challenge is that the distribution $n_d(x)$ is theoretically continuous, meaning infinite combinations of size and concentration can satisfy these equations (Patel et al., 2024). Even in a simplified monodisperse population where $n_d(x) = n_d\delta(x - x_0)$, retrieving a unique concentration and size requires that β and α have sufficiently different sensitivities to particle size to overcome observational uncertainties. In essence, inversions of particle properties from lidar data are difficult because the combination of many small particles and a few large particles can produce the same bulk optical properties.

In order to obtain more independent information, additional wavelengths have been added to lidar systems (Veselovskii et al., 2002; Müller et al., 2001). The hope is that the *independent* information from these wavelengths is sufficient to overcome

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

uncertainties in the observations and further constrain the particle radius and number concentration retrievals. However robustly determining the impact on the constraints provided by additional wavelengths is difficult and assessments of their impact are limited.

3 Methods

In contrast to inverse methods, forward calculation of backscatter and extinction is deterministic and straightforward. Large datasets of in situ observations of aerosol size distributions can be used to calculate both the desired lidar retrieved properties and the volume optical properties that would be retrieved by advanced lidar instruments such as HSRL and Raman lidar. This provides a mechanism to close the loop on potential lidar inversions of aerosol properties, with actual observed cases, without having to resort to cumbersome validation experiments. It also means that the lidar observations themselves are idealized, and results should be interpreted as best case scenarios where additional degrees of freedom (e.g. non-spherical particles) and observational uncertainties will tend to reduce the performance of the inversion.

In this work in-situ observations of fine mode aerosols from the Ultra High Sensitivity Aerosol Spectrometer (UHSAS) is used during the series of HIAPER Pole-to-Pole Observation (HIPPO) campaigns (NSF NCAR Earth Observing Laboratory, 2017, 2019a, b, c) to estimate the range of naturally occurring lidar observations of aerosols and their associated microphysical properties. HIPPO flights represent 320 flight hours of data with over 1 million individual spectra. The UHSAS provides radius resolved concentrations of aerosols ranging from 0.0275 μm to 0.5 μm in 100 logarithmically spaced bins. From these observations, Mie theory can be used to calculate the backscatter and extinction cross sections of each particle size for a given wavelength and the resulting associated volume backscatter and extinction coefficients can be directly calculated using Eq. (1) and (2) respectively. It should be noted that the UHSAS does not measure the particle refractive index, so for the purpose of this work, a range of lidar parameters is calculated for real refractive indices between 1.3 and 1.5 and imaginary refractive indices between 0 and 0.06. Because these are indiscriminately sampled in the training data, these variables are effectively treated as uniformly distributed unknowns. Note that this represents some break in consistency between the UHSAS observations and how they are forward modeled. The UHSAS particle sizing method typically assumes a refractive index to relate the particle scattering amplitude to a physical radius. Here we accept the radius size provided by the processed instrument data, but then impart a variety of possible refractive indices. While this imparts some inconsistency, our aim with the UHSAS size distributions is to leverage realistic cases. Given that we are leveraging extensive ensembles of these distributions, it is unlikely to drastically alter the results of the study.

The desired retrieved parameters are also calculated directly from the UHSAS observations. Particle concentration is calculated by summing the concentration of each size bin, and the effective radius is obtained from the ratio of third and second moments of the particle radius

$$\bar{r} = \frac{\sum_i n_d[i] r[i]^3}{\sum_i n_d[i] r[i]^2} \tag{3}$$

where $\bar{r}$ is the effective radius and i is the ith bin in the measured UHSAS histogram.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

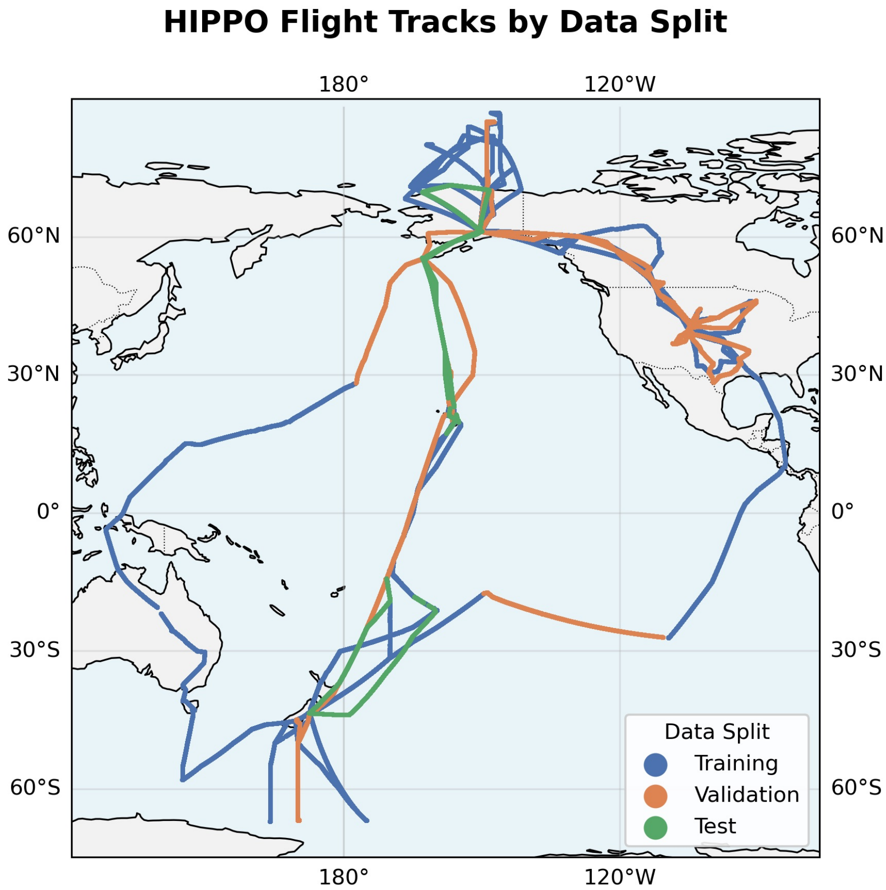

Figure 1. Map of flight tracks for data used in training, testing and validation of the PDF estimation methods employed here. Map generated using Cartopy (Met Office, 2010–2026) Base map features (coastlines, land, ocean, and country borders) are courtesy of Natural Earth (public domain) (Kelso and Patterson, 2026).

Data used for this analysis is obtained from HIPPO 1, and 3-5 (during HIPPO 2 the UHSAS was not operating correctly) (NSF NCAR Earth Observing Laboratory, 2017, 2019a, b, c). These observations extend over a large range of latitudes, to help assess the generalization of the retrievals across geographic regions (see Figure 1), but we should emphasize that most of the observations were over or near the Pacific Ocean so the general distribution of aerosol characteristics considered in this work may be smaller and more constrained than what exists across the entire globe and at different times of the year. However observations were not restricted to the marine boundary layer, so the aerosol types are not restricted to sea salt. This work will tend to *underestimate* uncertainties in retrievals if the global distribution of aerosol properties is significantly under-sampled in the HIPPO campaigns.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

3.1 Holdout Data

In order to perform hyperparameter optimization and evaluate the performance of PDF estimates, we holdout some of the UHSAS data. The holdout validation data is used to tune and optimize the hyper parameters of the PDF estimation models to avoid over fitting or over simplified models. An additional test dataset is then subsequently used to evaluate the performance of the models.

Splitting holdout data from training data is a non-trivial problem for many machine learning applications. The common practice of randomly sampling from the full dataset assumes that data points are uncorrelated. However given that datasets used here are temporally aligned and the atmosphere has relatively high correlation of features in time and space, randomly sampling observations is not an appropriate approach. Holdout data separated using a random sampling approach will tend to result in overfitting. It will also overestimate the model performance when applied to test data even though the solution will not generalize well to cases that are not explicitly sampled in the analysis. In an attempt to obtain a greater degree of independence between training, validation and test data, we holdout entire flights from the HIPPO data for validation and testing.

The purpose of using holdout data is for tuning and evaluating the proposed neural network (NN) solution. For our later assessment of lidar retrieval uncertainties in aerosol size and concentration, the full dataset (training, validation and test) is used.

3.2 Histogram PDF Estimation

As a performance baseline, we can directly estimate the PDF for input (calculated lidar backscatter and extinction) data and labels (concentration and effective radius from measured spectra) by using a multi-dimensional histogram approach. In principle, we could take a single set of lidar observations and ask, how many other microphysical parameters correspond to this same set of observed lidar signals? It is unlikely, however, any other data points would *exactly* match any one set of lidar observations. Under this histogram technique, a region (hyper volume) is defined over which the input variables are treated as identical ($\beta \pm \epsilon_\beta, \alpha \pm \epsilon_\alpha$). All labels (effective radius and concentration) corresponding to inputs falling into that hyper volume are then used to form a 2D histogram of particle concentration and effective radius where the output histogram approximates the solution PDF. In essence, any set of observed size distributions within our defined range that yield identical calculated lidar backscatter and extinction are collected.

While this approach enables a direct calculation of the spread of non-unique solutions corresponding to the input data, it suffers from some practical limitations. First, we have to set the hyper volume regions for grouping inputs such that any input data contained in the region are treated as "identical". This results in a prescribed input resolution (ϵ) which may not be appropriate for the particular input case. If the input resolution ϵ is too large, the output PDF will be unnecessarily smeared out. If the input resolution is too small, there will be large gaps in inputs where there is insufficient data to produce useful output PDF. Second, we have to select an output histogram resolution. If this is chosen too small, the histogram will be noisy and it may not be possible to evaluate the estimated PDF at all possible data points (particularly when using independent test data). If chosen too large, the histogram will have overly coarse resolution and inaccurately represent the features of the actual PDF.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Notably, we would tend to treat histograms as piecewise constant over the bins, and thus bin sizes that are large will tend to poorly represent PDF changes over the interval.

3.3 NN PDF Estimation

A machine learning approach has been developed to estimate a continuous and smooth PDF of aerosol number density and effective radius for a given set of (input) lidar observations. The PDF is parameterized as an exponential with Chebyshev polynomial arguments which ensures that the function is positive, continuous and smooth. In order to address sparseness in the data, a neural net is trained to interpolate these polynomial coefficients. Through this approach, the architecture "learns" to predict the PDF of aerosol number density and effective radius for a given set of lidar observations.

In order to train the neural net, we employ maximum likelihood estimation by using the negative log-likelihood of the predicted PDF as a loss function. The neural net is able to predict the best possible PDF that represents the input data by minimizing this loss.

The exponential polynomial parameterized PDF is written

$$f(\bar{r}, n_d) = \frac{\exp\left[\mathbf{P}(\bar{r}, n_d) \cdot g\right]}{\int_0^\infty \int_0^\infty \exp\left[\mathbf{P}(\bar{r}, n_d) \cdot g)\right] d\bar{r} dn_d} \tag{4}$$

where $f(\bar{r}, n_d)$ is the joint PDF of effective radius $\bar{r}$ and number density n_d, $\mathbf{P}(\bar{r}, n_d)$ is a vector of the Chebyshev polynomial terms with respect to the two variables $\bar{r}$ and n_d, and g are the polynomial coefficients. The denominator is a normalization term for the PDF so that $f(\bar{r}, n_d)$ integrates to 1.

To obtain the PDF from lidar observations, we aim to estimate the polynomial coefficients g as a function of lidar observations $\mathbf{x}$ such that the coefficients are now described $g(\mathbf{x})$. The goal is to obtain $f(\bar{r}, n_d; \mathbf{x})$ (now a function of inputs $\mathbf{x}$) for a set of lidar observations. In doing this, we effectively seek to interpolate the polynomial coefficients between observations while generating a smooth fit to the sparse data. This implicitly assumes that the polynomial coefficients themselves are smoothly varying such that they can be interpolated between observed data points.

In this work, $g(\mathbf{x})$ is approximated using a multilayer dense neural network (NN) which is trained within a maximum likelihood framework leveraging the PDF definition. The estimated PDF is modified from (4) as

$$f(\bar{r}, n_d; \mathbf{x}) = \frac{\exp\left[\mathbf{P}(\bar{r}, n_d) \cdot g(\mathbf{x})\right]}{\int_0^\infty \int_0^\infty \exp\left[\mathbf{P}(\bar{r}, n_d) \cdot g(\mathbf{x})\right] d\bar{r} dn_d} \tag{5}$$

Note that the integral in the denominator is a normalization term, but it remains important for this estimation problem. Because there is no explicit solution for the integral of exponential polynomials beyond order 2, this is evaluated numerically in the training process, which is not ideal but necessary.

In choosing the basis functions for the estimated PDF, there are a number of tradeoffs. By using an exponential polynomial, we drive the solution toward a smooth function which can have a large range of values, is always positive, and the log-likelihood can be directly evaluated without loss of numerical precision. The drawback is that the integral cannot be explicitly evaluated.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

The loss function we aim to minimize in training $g(\mathbf{x})$ is the negative log-likelihood of the PDF and is therefore

$$\mathcal{L}(\bar{r}, n_d; \mathbf{x}) = -\ln f(\bar{r}, n_d) = \ln \left\{ \int\limits_0^\infty \int\limits_0^\infty \exp\left[\mathbf{P}(\bar{r}, n_d) \cdot g(\mathbf{x})\right] d\bar{r}\, dn_d \right\} - \mathbf{P}(\bar{r}, n_d)) \cdot g(\mathbf{x}) \tag{6}$$

Thus for a given set of lidar inputs $\mathbf{x}$ (volume backscatter and extinction coefficients for any number of wavelengths) the NN parameters in $g(\mathbf{x})$ are adjusted through gradient descent to minimize $\mathcal{L}(\bar{r}, n_d)$ with corresponding training targets $\bar{r}$ and n_d. Through this process, we are able to estimate the joint PDF of effective radius and number concentration for a set of lidar observations. This avoids having to set a tolerance (ϵ) as in the histogram method which inherently also affects the spread of the PDF. This NN approach also enables recovery of a smooth, continuous PDF even when data points are sparse. The NN itself is dependent on the number and types of inputs, and must thus be trained and hyperparameter optimized independently for each input configuration (combinations of extinction, backscatter and wavelengths).

To handle the considerable dynamic range in input and output variables, all actual interfacing variables are logarithmic quantities. As is typical in machine learning applications, these logarithmic inputs ($\mathbf{x}$) and labels (n_d and $\bar{r}$) are also normalized for training and evaluation such that all the training data is bounded between 0 and 1. This normalization operation is set based on the training data and the same operation is used for validation and test data.

We should note that evidential NN (Amini et al., 2020; Schreck et al., 2023) were not used here because in the existing formulation they are not able to account for covariances in the output variables, which are significant in this problem. In addition they assume the output PDF is well approximated by a normal distribution which corresponds to an exponential polynomial of order 2. However hyper parameter optimization always resulted in polynomial orders much higher than this, suggesting that the assumption of a normally distributed PDF is not an ideal approximation.

We hyperparameter optimized the NN using the ECHO package developed at NSF-NCAR and built on Optuna. This was used to determine optimal number of layers in the dense network, the number of nodes in each layer, polynomial orders for both estimated variables, learning rate, batch size and threshold for the scheduler used to reduce the learning rate when the validation loss plateaus.

Table 1 shows the most relevant parameters obtained through hyper-parameter optimization for each lidar architecture considered here. We should note that we do not believe these hyper-parameters should be copied verbatim for use in operational algorithms, but they do provide some key insights into the level of complexity supported by the data. Most notably, the polynomial orders are well beyond second order in both retrieved parameters, which indicates that Gaussian PDFs would be over-simplifications for the solutions considered here.

Once the NN is trained, the execution in inference mode is very fast. This makes it trivial to numerically investigate the effects of uncertainties on the input parameters through an ensemble analysis. The PDF of solutions resulting from errors in the lidar observations can be estimated using

$$\hat{f}(\bar{r}, n_d, \mathbf{x}) = \frac{1}{N} \sum_{n=1}^{N} f(\bar{r}, n_d, \mathbf{x} + \Delta_n) \tag{7}$$

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Table 1. Optimized neural network hyperparameters for each evaluated lidar architecture.

Model	Layers	Nodes per Layer	n_d Polynomial Order	$\bar{r}$ Polynomial Order
$6\beta+2\alpha$	6	3384	7	23
$3\beta+2\alpha$	6	2062	23	38
355 nm	8	1858	26	39
3β	4	3438	16	32

where Δ_n is a vector of random numbers corresponding to uncertainty in $\mathbf{x}$. Initially we will use this method to more easily compare outputs from the NN and the histogram method where Δ_n is a uniform distribution corresponding to the hyper-volume tolerance ϵ. Later we will evaluate the impact of uncertainty in the lidar observations on the spread in the solution space where Δ_n will be normally distributed. While not required, for purposes of analysis presented here, all random numbers in Δ_n are statistically independent.

3.4 PDF Results

In this section, we baseline the NN model's performance against a direct histogram approach. It is important to clarify that the objective of this analysis is strictly to evaluate how accurately each method represents the underlying data; at this point, it is not an assessment of the performance of any particular lidar configuration. Rather, our primary goal is to determine whether the NN can overcome the inherent limitations of the histogram approach without introducing significant additional error.

To ensure a fair comparison and avoid biasing the results in favor of the NN, we optimized the histogram configurations prior to testing. We determined the ideal number of input and output bins by computing the histograms using training data and calculating the Negative Log-Likelihood (NLL) of the resulting solutions against the validation data. The histogram configuration that yielded the lowest validation NLL was then selected as the benchmark for evaluating the NN solution on the independent test dataset.

A consideration in this evaluation is how to handle empty regions where the histogram contains no data points (which can occur due to a lack of either corresponding input or output data). For our analysis, we filter out these empty cases but evaluate performance using the sum—rather than the mean—of the NLL across all valid validation or test data points (validation data used to optimize bins, while test data used to benchmark against the NN). Utilizing the sum operation allows us to implicitly penalize the model for missing data points without causing the NLL to evaluate to infinity.

The optimal histogram bin sizes and ϵ for each input configuration, as determined by the lowest validation NLL, are detailed in Table 2. Note that the inputs and outputs are logarithmic quantities and scaled between 0 and 1 based on training data. The value of ϵ is reported on this same scale and the output bins widths thus span this input range. These optimized histograms subsequently serve as the baseline for evaluating the NN model with the test data.

Similar to how the histogram and NN models were optimized using validation NLL, the most objective way to compare their performance is by calculating the NLL on an independent test dataset. We evaluated the NLL of both solutions against

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Table 2. Histogram input tolerance and number of output histogram bins for each model.

Model	ϵ	Output Bins
$6\beta + 2\alpha$	0.019	43
$3\beta + 2\alpha$	0.009	43
355 nm	0.005	79
3β	0.007	58

a holdout test set consisting of over 3 million points (33 flight hours consisting of more than 100,000 aerosol spectra with 28 combinations of real and imaginary refractive index). To ensure a fair comparison, we filtered out cases where the histogram did not evaluate to a finite number. Consequently, the reported histogram NLL represents the sum of these filtered, valid results. For a comprehensive view, we report the NN NLL evaluated on both this filtered subset and the entire unfiltered dataset, alongside the total number of filtered points. These results are detailed in Table 3.

As Table 3 shows, the NN consistently performs as well as, or better than, the histogram approach. Notably, the NN's overall score improves when the unfiltered points are included. This indicates that it provides meaningful solutions even in cases where the histogram fails, likely by interpolating from training data across wider gaps than the discrete histogram bins permit. For the multi-wavelength lidar data, the NN outperforms the histogram even when restricted to the filtered cases. Conversely, performance is nearly identical between the two approaches for the single-wavelength 355 nm observations (when only considering filtered cases).

This discrepancy highlights a classic trade-off: as input dimensions increase, data is spread across larger (higher dimensional) hyper-volumes, causing the data for any particular case to become sparse. While any model requires more training data to combat sparsity in higher dimensions, the NN handles this challenge much more robustly than a direct histogram approach, by smoothly interpolating through data gaps. With the lower-dimensional single input data, the histogram suffers much less from low point density, bringing the two solutions into closer alignment. Nevertheless, the NN's ability to evaluate test cases that the histogram cannot finitely process demonstrates its practical benefit.

Table 3. Test Negative Log-Likelihood (NLL) analysis across inputs.

Inputs	Histogram NLL	NN NLL (filtered)	NN NLL (unfiltered)	Filtered Points
$6\beta + 2\alpha$	-1.27×10^7	-1.67×10^7	-1.76×10^7	207,332 (6.31%)
$3\beta + 2\alpha$	-1.29×10^7	-1.48×10^7	-1.58×10^7	260,602 (7.94%)
355 nm	-1.13×10^7	-1.13×10^7	-1.15×10^7	76,926 (2.34%)
3β	-1.19×10^7	-1.22×10^7	-1.26×10^7	158,044 (4.81%)

Beyond the aggregate NLL analysis, we present two example cases in Figures 2 and 3 to provide a more intuitive, visual comparison. The top row of each figure displays the predicted NN solution PDF for a specific, error-free input data point. Because the histogram approach inherently bins—and therefore spreads out—accepted inputs, we generated an "ensemble

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

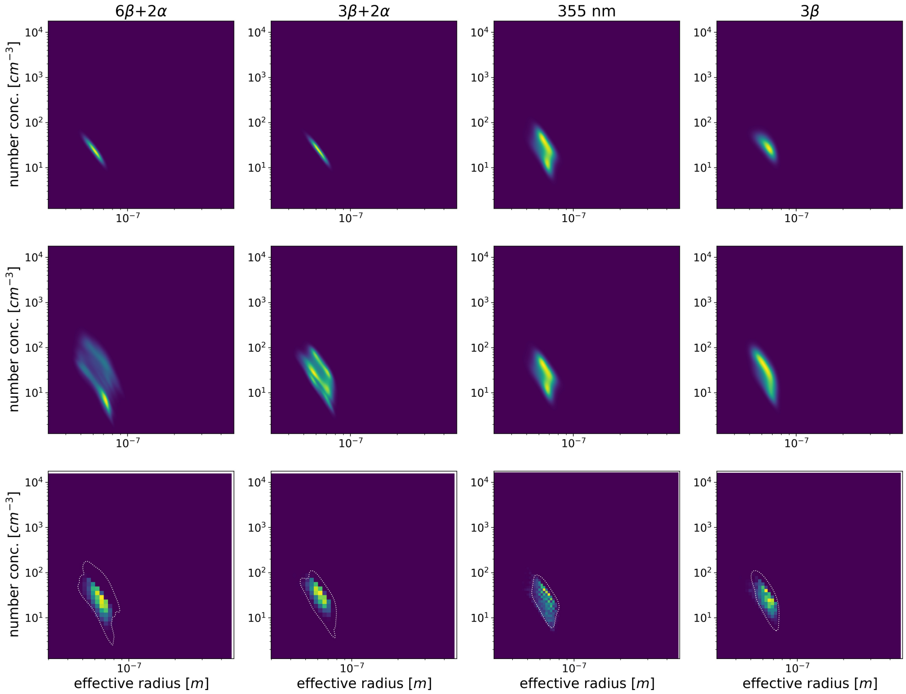

Figure 2. Example output PDFs for each input configuration. Results are shown for the NN (top row), the NN for an ensemble input covering the histogram input bin width (middle row) and the histogram (bottom row). In the histogram figures, the contour shows the 10% boundary of the NN ensemble case for comparison.

NN" PDF for a more direct comparison. To do this, we evaluated the NN multiple times while adding uniform random noise to the inputs corresponding to the histogram bin widths. The resulting ensemble PDF, calculated using Eq. (7), is shown in the middle row. Finally, the bottom row displays the directly calculated histogram output, overlaid with a white contour representing the 10% of peak boundary of the ensemble NN (middle row) for reference.

When provided with error-free data (top rows), the NN tends to produce more tightly constrained PDFs. As expected, applying the ensemble analysis to match histogram binning broadens the output space (with the exception of the 355 nm-only NN), resulting in a significantly less constrained result. This broadening highlights a key reason the NN outperforms the histogram in the aggregate NLL analysis: the finite, prescribed bin sizes of the histogram cause unnecessary smearing in the output PDF. The ensemble NN PDFs closely resemble the directly calculated histogram PDFs. Because the histogram calculation is straightforward and transparent, it is more reasonable to implicitly trust its general structural accuracy. The fact that the NN mimics this structure (when accounting for bin width) provides intuitive confidence in the trained NN's output, complementing the robust, albeit less intuitive, NLL metric.

While the ensemble NN and histogram outputs share the same general features, occasional discrepancies arise in peak values and exact probability densities. These differences stem from the respective limitations of each approach. For instance,

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

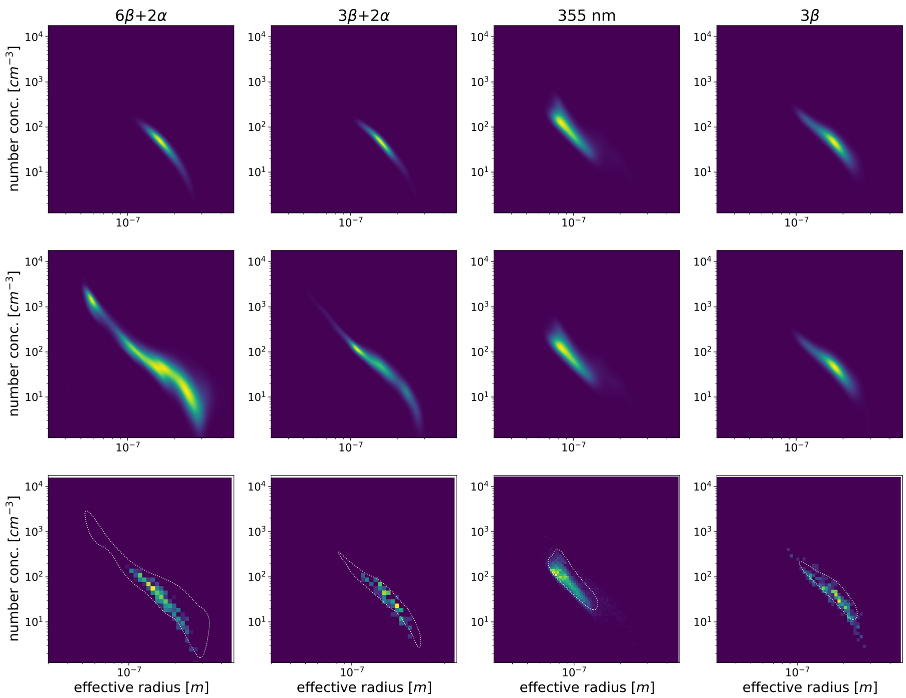

Figure 3. Example output PDFs for each input configuration. Results are shown for the NN (top row), the NN for an ensemble input covering the histogram input bin width (middle row) and the histogram (bottom row). In the histogram figures, the contour shows the 10% boundary of the NN ensemble case for comparison.

data sparsity limits the histogram's ability to resolve the PDF in lower-probability regions, meaning the bounds of the NN ensemble output are generally broader than the non-zero clusters of the histogram. Conversely, the NN may exhibit overfitting or excessive noise sensitivity (see the Impact of Lidar Uncertainty section below), imposing complex features not strictly reflective of the underlying data. When weighing these discrepancies, the aggregate results in Table 3 remain the critical baseline. Ultimately, the direct NN output—not the ensemble output—is the actual product of interest, and the NLL analysis consistently demonstrates that it matches or exceeds the performance of the direct histogram calculation.

4 Evaluation of Lidar Architectures

Building on the neural network (NN) solution described in the previous sections, we now analyze the potential of lidar observations to constrain estimates of particle effective radius and number concentration. Because we are no longer evaluating the NN model performance, we expand our analysis data to include all training, validation, and test datasets.

First, we establish that incorporating lidar data successfully constrains the solution space of aerosol size and concentration. Figure 4 shows a histogram of all effective radius and concentration data points for the full dataset. The contours in each plot represent the 10% of peak boundaries of the NN-predicted PDFs for three example cases across the four lidar architectures

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

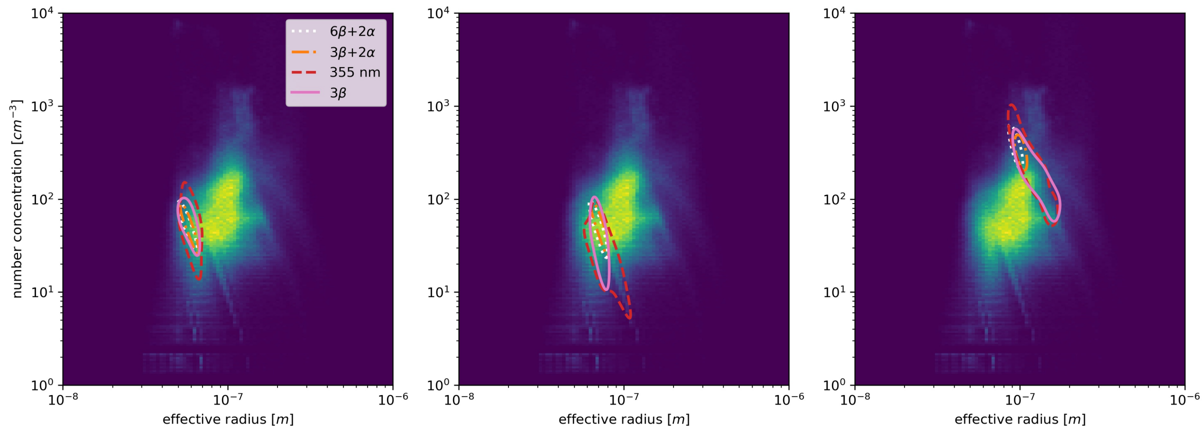

Figure 4. Histograms of particle effective radius and concentration for the entire dataset. The contours show the 10% peak height boundaries for the NN outputs of three example cases across the four evaluated lidar architectures.

considered. It is clear that the lidar data significantly reduces the set of possible aerosol properties compared to the global distribution. However, the principal axes of these reduced solution spaces do not align with the axes of the target properties. Consequently, the resulting sets of possible properties can still span a significant range of values in both particle effective radius and concentration, even though the overall area of the PDF is drastically reduced by the lidar constraint.

4.1 Performance and Sensitivity to Observational Uncertainty

To quantify retrieval performance, we analyze the ultimate uncertainty in the aerosol properties based on the 90% bounds of the resulting PDF, expressing it as a fractional spread of possible solutions. Specifically, we locate the 5% and 95% crossing points of the cumulative distribution function (CDF) and define the ratio of the upper to lower bound as the factor of spread.

To evaluate how uncertainty in the observed lidar parameters affects this spread, we perform an ensemble calculation using Eq. (7), where Δ_n is normally-distributed, independent for all lidar variables, and has a variance dictated by percent error. We consider observation errors of 0%, 2%, 4%, 9%, 13%, and 22%.

For context, the baseline spread of the entire unconstrained dataset is a factor of 61 in concentration and 3.6 in effective radius. The distributions of the spread in the NN solutions under varying errors are shown in Figure 5. At low observation errors, both the $6\beta+2\alpha$ and $3\beta+2\alpha$ architectures constrain the number density spread between a factor of 2 and 10, and the effective radius to better than a factor of 1.5. However, as observation errors increase, the distribution of the spread shifts to higher values. The $3\beta+2\alpha$ configuration is particularly sensitive; at a 13% input error, a significant number of cases produce a PDF spread greater than the baseline spread of all possible solutions (a factor fo 61).

Conversely, the 355 nm single-wavelength retrieval utilizing both backscatter and extinction is more robust, changing very little as observation errors increase. The 3β retrieval performs similarly to the 355 nm configuration at low errors, but exhibits more uncertainty sensitivity. Figure 6 visualizes these error sensitivity dynamics by plotting the boundary containing 90%

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

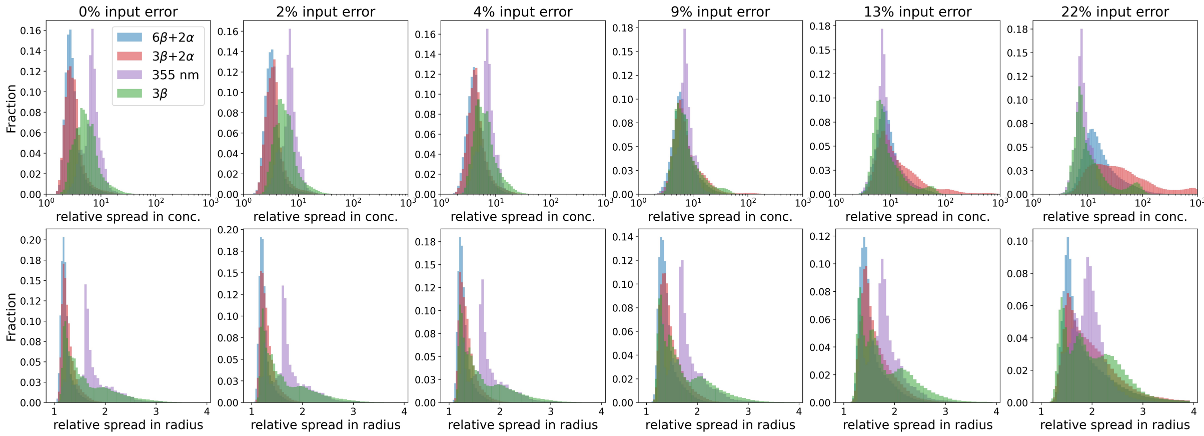

Figure 5. Histograms detailing the factor of spread in NN PDF outputs across the dataset. The limits are evaluated under varying percentages of input error to establish the impact of observational uncertainty on retrieval performance.

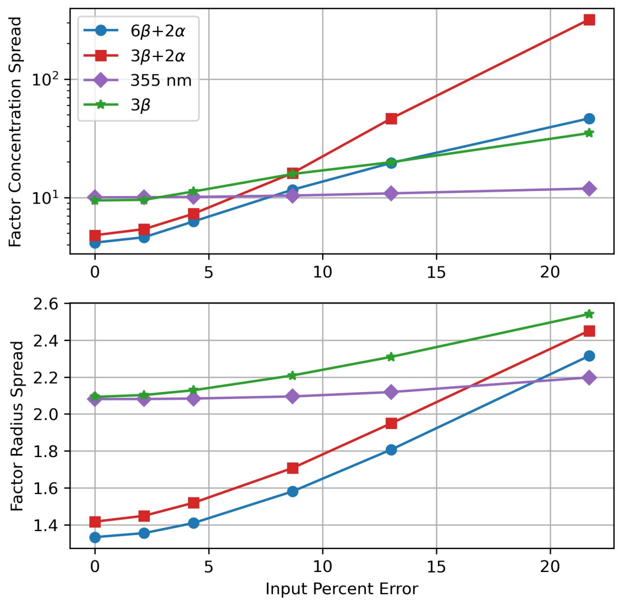

Figure 6. Plots illustrating the 90% containment of spread in retrievals as a function of percent input error for varying lidar architectures.

of all cases from the histograms in Figure 5. It is evident that while multi-wavelength retrievals containing extinction data provide superior constraints at low uncertainty, they likely rely on minute differences in input variables. When lidar uncertainty increases, retrieval accuracy degrades sharply. The single-wavelength observation naturally produces a wider, more uncertain constraint, but benefits from being less sensitive to observational uncertainty.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

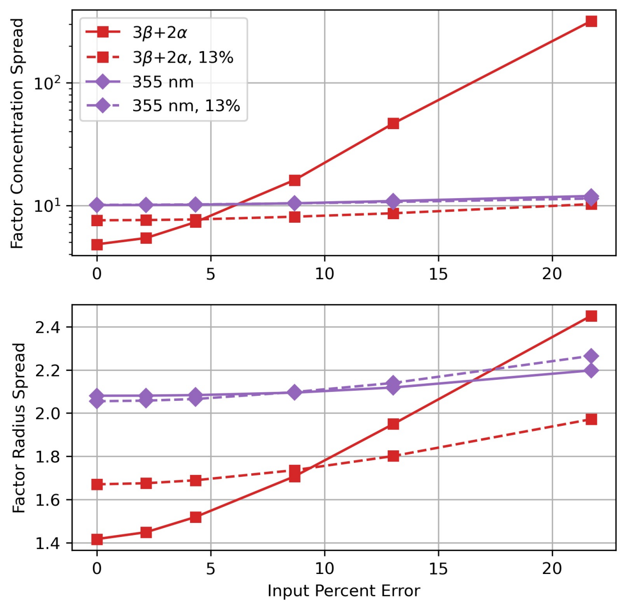

Figure 7. Plots showing 90% containment of spread in retrievals as a function of percent input error. Two wavelength configurations are compared, trained both without error (solid lines) and with 13% injected input training error (dashed lines).

4.2 Mitigating Uncertainty

The degradation of multi-wavelength retrieval spreading under higher observation uncertainty conditions can be mitigated by accounting for the instrument uncertainty during the NN training process. By injecting random noise reflective of real-world instrument uncertainties into the lidar inputs during training, the NN avoids over-reliance on signal differences that cannot be realistically captured by the instrument.

To demonstrate this, we trained a NN for the error sensitive $3\beta + 2\alpha$ architecture alongside the 355 nm ($1\beta + 1\alpha$) control, introducing a random 13% statistically independent, normally distributed noise to the lidar inputs during training. The resulting 90% containment curves for these additional cases are shown as dashed lines in Figure 7. While the 355 nm baseline remains relatively unaffected, the noise-trained $3\beta + 2\alpha$ output becomes significantly less sensitive to input uncertainty at higher input error levels. This robustness comes at the cost of higher uncertainty in the low-error regimes. Because operational lidar uncertainty is often scene-dependent, variable, and analysis is limited in scope such that it tends to be underestimated, utilizing a noise-injected training approach (e.g., 13%) seems appropriate for operational algorithms. Ideally, future training processes would encapsulate the specific, likely error levels of each individual data product rather than applying a blanket error value. An operational approach would also need to account for covariances between input data products.

Finally, a necessary caveat to this analysis is the assumption of independent, uncorrelated, and normally distributed errors. In practice, estimating minimum errors in multi-wavelength lidar retrievals involves highly noise-sensitive extinction measurements that rely on estimates of the lidar overlap function. Furthermore, errors are often correlated; for example, if extinction

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

estimates are derived by multiplying an estimated lidar ratio by the observed backscatter coefficient, the extinction errors will strongly correlate with backscatter errors.

Efforts to reduce noise in extinction estimates also introduce varying levels of structural assumptions. For example, some approaches assume a constant lidar ratio over coarse resolutions to produce low random noise, which can result in attractive but inaccurate images of extinction that reduce the fidelity of a 1:1 mapping between backscatter and extinction structures. Other forward-modeling approaches utilizing regularization may simultaneously estimate lidar ratios and backscatter coefficients, but integrating across heterogeneous structures can easily produce biases due to the nonlinear relationship between the estimated variables and observed signals. These systematic uncertainties must be carefully evaluated prior to the operationalization of any retrieval algorithm.

5 Discussion

The analysis of these four observational configurations highlights a critical trade-off between information content and noise sensitivity. Notably, the three-wavelength backscatter-only (3β) architecture is less performative than a single-wavelength backscatter and extinction ($1\beta+1\alpha$) configuration. This underscores the inherent advantages of HSRL and Raman lidar systems over elastic backscatter lidars. The direct measurement of extinction provides independent information that backscatter alone cannot supply. This advantage is further accentuated when considering that elastic backscatter lidars often carry observational uncertainties due to the assumptions used in their volume backscatter retrievals, which are not addressed through advanced statistical noise suppression.

While multi-wavelength lidar observations encompassing both backscatter and extinction theoretically demonstrate superior constraints on aerosol properties, this superiority holds only under conditions of low observational uncertainty. This presents a key caveat: increasing instrument complexity to capture additional wavelengths inevitably introduces new sources of systematic and random error. Consequently, evaluating these advanced architectures purely on a "low uncertainty" basis may not reflect operational realities. When realistic observational errors are factored in, many of the theoretical benefits of multi-wavelength systems are significantly reduced.

The inevitable counterargument to the limitations of lidar retrievals is that researchers can constrain the solution space using circumstantial data. Techniques such as POLIPHON (Ansmann et al., 2026) advocate for utilizing model back-trajectories, inferences about aerosol composition, or spatially constrained datasets to refine retrievals and eliminate statistical outliers.

While practically useful, this approach introduces a core epistemological question: at what point does a highly constrained, prior-dependent data product cease to be an objective "observation"? In the scientific method, observations are intended to stand apart from theory, confronting hypotheses with raw, independent reality. When retrieval algorithms heavily integrate prior knowledge and models, they risk merging the observations with the very principles they are meant to test. Because the layers of parameterization within these algorithms rarely propagate their uncertainties transparently, this practice risks generating self-validating data products, inadvertently steering the community toward false scientific dogma.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Rather than forcing a single deterministic "best guess" through rigid assumptions, providing a Probability Density Function (PDF) better encapsulates how lidar observations actually constrain possible aerosol properties. While a PDF may initially appear more cumbersome for scientific analysis than a single value, it is a necessary step for rectifying discrepancies across observational systems and models. Transparently reporting the full range of physical possibilities and only then applying additional constraints enables more robust, mathematically sound comparisons between global climate models and observational data and helps keep assumptions and their impacts on scientific conclusions more visible.

Finally, it is important to note that the analysis presented here serves as a "best-case" baseline. The framework assumes fine-mode, spherical aerosols. Introducing additional degrees of freedom—such as coarse-mode aerosols, asphericity, and highly variable complex refractive indices—will alter these dynamics, most likely introducing further uncertainty into the retrievals.

6 Conclusions

This study demonstrates that utilizing machine learning to predict a continuous Probability Density Function (PDF) of aerosol states offers a robust alternative to deterministic lidar inversions. By evaluating four distinct lidar architectures through forward modeling of high-resolution *in situ* data, we establish baseline expectations for what modern lidar systems can practically achieve when disentangling particle size from number concentration.

A general rule of thumb derived from this best-case analysis suggests HSRL and Raman lidar observations can constrain aerosol effective radius to within a factor of 2, and number concentration to within an order of magnitude. While multi-wavelength systems offer superior theoretical constraints, their heightened sensitivity to input noise means that simpler architectures—such as a $1\beta + 1\alpha$ HSRL system—often provide more robust, stable constraints under realistic observational uncertainties.

Ultimately, the choice of retrieval methodology must reflect its intended application. For operational forecasting where real-time utility is paramount, the integration of heavy assumptions and models to yield a specific data product may be acceptable, as the need for a definitive answer supersedes strict analytical rigor. However, for foundational climate science and model evaluation, observations must remain independent of the theories they seek to validate. By embracing the fundamental non-uniqueness of the lidar inversion problem and outputting probabilistic distributions, we can provide the climate modeling community with truly independent, mathematically transparent constraints on aerosol-cloud interactions.

Code and data availability. Data used in training and analysis of models and the trained, hyperparameter optimized NN models are available in Matthew Hayman (2026). Software used for training and analysis is publicly available at https://github.com/NCAR/aerosol-scattering .

Author contributions. M.H. performed the analysis and wrote the manuscript for this work.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

Competing interests. The authors have no competing interests to declare.

Acknowledgements. This material is based upon work supported by the NSF National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under Cooperative Agreement No. 1852977. We would like to acknowledge computing support from the Casper system (https://ncar.pub/casper) provided by the NSF National Center for Atmospheric Research (NCAR), sponsored by the National Science Foundation. The author would like to thank Scott Spuler and Robert Stillwell for their suggestions and review of this work. Gemini 3.1 Pro was used to assist with proofreading, language editing, and organization. It was also used to write and define the dense neural network classes for hyperparameter modularity. All content was reviewed, verified, and edited by the author, who takes full responsibility for the manuscript and its figures.

https://doi.org/10.5194/egusphere-2026-5320
Preprint. Discussion started: 17 September 2026

References

Amini, A., Schwarting, W., Soleimany, A., and Rus, D.: Deep evidential regression, Advances in Neural Information Processing Systems, 33, 20 004–20 016, https://doi.org/10.48550/arXiv.1910.02600, 2020.

Ansmann, A., Riebesell, M., and Weitkamp, C.: Measurement of atmospheric aerosol extinction profiles with a Raman lidar, Optics Letters, 15, 746, https://doi.org/10.1364/ol.15.000746, 1990.

Ansmann, A., Wandinger, U., Riebesell, M., Weitkamp, C., and Michaelis, W.: Independent measurement of extinction and backscatter profiles in cirrus clouds by using a combined Raman elastic-backscatter lidar, Applied Optics, 31, 7113, https://doi.org/10.1364/ao.31.007113, 1992.

Ansmann, A., Hofer, J., Mamouri, R.-E., Haarig, M., Baars, H., and Wandinger, U.: Aerosol optical-to-microphysical conversion factors for lidars and ceilometers from extended AERONET data analyses: POLIPHON update, Atmospheric Measurement Techniques, 19, 3801–3830, https://doi.org/10.5194/amt-19-3801-2026, 2026.

Burton, S. P., Hostetler, C. A., Cook, A. L., Hair, J. W., Seaman, S. T., Scola, S., Harper, D. B., Smith, J. A., Fenn, M. A., Ferrare, R. A., Saide, P. E., Chemyakin, E. V., and Müller, D.: Calibration of a high spectral resolution lidar using a Michelson interferometer, with data examples from ORACLES, Applied Optics, 57, 6061–6075, https://doi.org/10.1364/AO.57.006061, 2018.

Burton, S. P., Hair, J. W., Hostetler, C. A., Fenn, M. A., Smith, J. A., and Ferrare, R. A.: Aerosol extinction and backscatter Optimal Estimation retrieval for High Spectral Resolution Lidar, Atmospheric Measurement Techniques, 18, 6527–6543, https://doi.org/10.5194/amt-18-6527-2025, 2025.

Cooney, J. A., Orr, J., and Tomasetti, C.: Measurements separating the gaseous and aerosol components of laser atmospheric backscatter, Nature, 224, 1098–1099, https://doi.org/10.1038/2241098a0, 1969.

Donovan, D. P., van Zadelhoff, G.-J., and Wang, P.: The EarthCARE lidar cloud and aerosol profile processor (A-PRO): the A-AER, A-EBD, A-TC, and A-ICE products, Atmospheric Measurement Techniques, 17, 5301–5340, https://doi.org/10.5194/amt-17-5301-2024, 2024.

Eloranta, E. W.: High Spectral Resolution Lidar, in: Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, edited by Weitkamp, C., chap. 5, Springer, https://doi.org/10.1007/3-540-07743-X_18, 2005.

Fernald, F. G.: Analysis of atmospheric lidar observations: some comments, Applied Optics, 23, 652, https://doi.org/10.1364/ao.23.000652, 1984.

Fiocco, G., Benedetti-Michelangeli, G., Maischberger, K., and Madonna, E.: Measurement of Temperature and Aerosol to Molecule Ratio in the Troposphere by Optical Radar, Nature Physical Science, 229, 78–79, https://doi.org/10.1038/physci229078a0, 1971.

Hair, J. W., Hostetler, C. A., Cook, A. L., Harper, D. B., Ferrare, R. A., Mack, T. L., Welch, W., Izquierdo, L. R., and Hovis, F. E.: Airborne High Spectral Resolution Lidar for profiling aerosol optical properties, Applied Optics, 47, 6734–6752, https://doi.org/10.1364/AO.47.006734, 2008.

Hayman, M. and Spuler, S.: Demonstration of a diode-laser-based high spectral resolution lidar (HSRL) for quantitative profiling of clouds and aerosols, Optics Express, 25, A1096–A1110, https://doi.org/10.1364/OE.25.0A1096, 2017.

Illingworth, A. J., Barker, H. W., Beljaars, A., Ceccaldi, M., Chepfer, H., Clerbaux, N., Cole, J., Delanoë, J., Domenech, C., Donovan, D. P., Fukuda, S., Hirakata, M., Hogan, R. J., Huenerbein, A., Kollias, P., Kubota, T., Nakajima, T., Nakajima, T. Y., Nishizawa, T., Ohno, Y., Okamoto, H., Oki, R., Sato, K., Satoh, M., Shephard, M. W., Velázquez-Blázquez, A., Wandinger, U., Wehr, T., and van Zadelhoff, G.-J.: The EarthCARE Satellite: The next step forward in global measurements of clouds, aerosols, precipitation and radiation, Bulletin of the American Meteorological Society, 96, 1311–1332, https://doi.org/10.1175/BAMS-D-12-00227.1, 2015.

Kelso, N. V. and Patterson, T.: Natural Earth: Free vector and raster map data, https://naturalearthdata.com, 2026.

Klett, J. D.: Stable analytical inversion solution for processing lidar returns, Applied Optics, 20, 211, https://doi.org/10.1364/ao.20.000211, 1981.

Klett, J. D.: Lidar inversion with variable backscatter/extinction ratios, Applied Optics, 24, 1638, https://doi.org/10.1364/ao.24.001638, 1985.

Lenhardt, E., Gao, L., Hostetler, C. A., Ferrare, R. A., Burton, S. P., Moore, R. H., Ziemba, L. D., Crosbie, E., Sorooshian, A., Soloff, C., and Redemann, J.: Aerosol effective radius governs the relationship between cloud condensation nuclei (CCN) concentration and aerosol backscatter, Atmospheric Chemistry and Physics, 25, 13 747–13 768, https://doi.org/10.5194/acp-25-13747-2025, 2025.

Lenhardt, E. D., Gao, L., Redemann, J., Xu, F., Burton, S. P., Cairns, B., Chang, I., Ferrare, R. A., Hostetler, C. A., Saide, P. E., Howes, C., Shinozuka, Y., Stamnes, S., Kacarab, M., Dobracki, A., Wong, J., Freitag, S., and Nenes, A.: Use of lidar aerosol extinction and backscatter coefficients to estimate cloud condensation nuclei (CCN) concentrations in the southeast Atlantic, Atmospheric Measurement Techniques, 16, 2037–2054, https://doi.org/10.5194/amt-16-2037-2023, 2023.

Marais, W. J., Holz, R. E., Hu, Y. H., Kuehn, R. E., Eloranta, E. E., and Willett, R. M.: Approach to simultaneously denoise and invert backscatter and extinction from photon-limited atmospheric lidar observations, Applied Optics, 55, 8316–8334, https://doi.org/10.1364/AO.55.008316, 2016.

Matthew Hayman: Aerosol Microphysical and Lidar Properties Estimated from HIPPO Projects, https://doi.org/10.5281/zenodo.22049113, accessed 31 Aug 2026, 2026.

Melfi, S. H.: Remote measurements of the atmosphere using Raman scattering, Applied Optics, 11, 1605–1610, https://doi.org/10.1364/AO.11.001605, 1972.

Met Office: Cartopy: a cartographic python library with a matplotlib interface, Exeter, Devon, https://scitools.org.uk, 2010–2026.

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

Müller, D., Wandinger, U., Althausen, D., and Fiebig, M.: Comprehensive particle characterization from three-wavelength Raman-lidar observations: case study, Applied Optics, 40, 4863–4869, https://doi.org/10.1364/AO.40.004863, 2001.

NSF NCAR Earth Observing Laboratory: HIPPO: NCAR GV (HIAPER) Low Rate (LRT - 1 sps) Navigation, State Parameter, and Microphysics Flight-Level Data, Version 5.0, https://doi.org/10.5065/D6X928N0, accessed 21 Aug 2025, 2017.

NSF NCAR Earth Observing Laboratory: HIPPO-3: Low Rate (LRT - 1 sps) Navigation, State Parameter, and Microphysics Flight-Level Data, Version 5.0, https://doi.org/10.5065/D6QF8R6R, accessed 21 Aug 2025, 2019a.

NSF NCAR Earth Observing Laboratory: HIPPO-4: Low Rate (LRT - 1 sps) Navigation, State Parameter, and Microphysics Flight-Level Data, Version 3.0, https://doi.org/10.5065/D6V40SK6, accessed 21 Aug 2025, 2019b.

NSF NCAR Earth Observing Laboratory: HIPPO-5: Low Rate (LRT - 1 sps) Navigation, State Parameter, and Microphysics Flight-Level Data, Version 3.0, https://doi.org/10.5065/D6CZ35HX, accessed 21 Aug 2025, 2019c.

Patel, P. N., Jiang, J. H., Gautam, R., Gadhavi, H., Kalashnikova, O., Garay, M. J., Gao, L., Xu, F., and Omar, A.: A remote sensing algorithm for vertically resolved cloud condensation nuclei number concentrations from airborne and spaceborne lidar observations, Atmospheric Chemistry and Physics, 24, 2861–2883, https://doi.org/10.5194/acp-24-2861-2024, 2024.

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

Qing, P., Nakane, H., Sasano, Y., and Kitamura, S.: Numerical simulation of the retrieval of aerosol size distribution from multiwavelength laser radar measurements, Applied Optics, 28, 5259–5265, https://doi.org/10.1364/AO.28.005259, 1989.

Redemann, J. and Gao, L.: A machine learning paradigm for necessary observations to reduce uncertainties in aerosol climate forcing, Nature Communications, https://doi.org/10.1038/s41467-024-52747-y, 2024.

Schreck, J. S., II, D. J. G., Becker, C., Chapman, W. E., Elmore, K., Gantos, G., Kim, E., Kimpara, D., Martin, T., Molina, M. J., Pryzbylo, V. M., Radford, J., Saavedra, B., Willson, J., and Wirz, C.: Evidential Deep Learning: Enhancing Predictive Uncertainty Estimation for Earth System Science Applications, arXiv:2309.13207, https://doi.org/10.48550/arXiv.2309.13207, 2023.

She, C. Y., Alvarez, R. J., Caldwell, L. M., and Krueger, D. A.: High-spectral-resolution Rayleigh-Mie lidar measurement of aerosol and atmospheric profiles, Optics Letters, 17, 541–543, https://doi.org/10.1364/OL.17.000541, 1992.

Shifrin, K. S. and Zolotov, I. G.: The Use of Direct Observations over the Aerosol Particle Size Distribution for Inverting Lidar Data, Journal of Atmospheric and Oceanic Technology, 20, 1411–1420, https://doi.org/10.1175/1520-0426(2003)020<1411:TUODOO>2.0.CO;2, 2003.

Shimizu, H., Lee, S. A., and She, C. Y.: High spectral resolution lidar system with atomic blocking filters for measuring atmospheric parameters, Applied Optics, 22, 1373–1381, https://doi.org/10.1364/AO.22.001373, 1983.

Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Wandinger, U., and Whiteman, D.: Inversion with regularization for the retrieval of tropospheric aerosol parameters from multiwavelength lidar sounding, Applied Optics, 41, 3685–3699, https://doi.org/10.1364/AO.41.003685, 2002.

Whiteman, D. N.: Examination of the traditional Raman lidar technique. I. evaluating the temperature-dependent lidar equation, Applied Optics, 42, 2571–2592, https://doi.org/10.1364/AO.42.002571, 2003.

Whiteman, D. N., Melfi, S. H., and Ferrare, R. A.: Raman lidar system for measurement of water vapor and aerosols in the Earth's atmosphere, Applied Optics, 31, 3068–3082, https://doi.org/10.1364/AO.31.003068, 1992.