ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

PAUL SCHERRER INSTITUT
PSI

Invertible neural networks for retrieving aerosol properties

written by **David Degen**,

in scientific collaboration with **Dr. Romana Boiger**, **Dr. Martin B. Gysel**, **Alireza Moallemi**, **Dr. Robin L. Modini**

supervised by **Dr. Andreas Adelmann**

June 2021

Abstract

Knowing the microphysical properties of aerosols is important in various areas, from predicting the course of climate change to monitoring air quality in cities, industrial areas, and medical facilities. Traditionally, when these properties are probed using light scattering data, they are retrieved either using pre-computed lookup tables, which is fast but not accurate, or by physically-based iterative algorithms, such as GRASP-OPEN, which is accurate but not fast. During the last 40 years, machine learning algorithms have been applied to different variations of this problem, hoping to establish a method that is both accurate and fast.

The following report describes the application of an invertible neural network model to simulate multi-angle light scattering data, namely the scattering phase function and the polarized phase function, on one hand, and to solve the inverse problem of retrieving aerosol properties from multi-angle light scattering data on the other hand. It is found that this method delivers accurate results and might be practical for the retrieval of microphysical properties of aerosols from light scattering data as well as for the initialization of physically-based iterative algorithms.

Table of Contents

1 Introduction

Atmospheric aerosols are particles suspended in air of various chemical composition and size, ranging from a few nanometres up to 100 µm [10]. They can be solid, liquid or a mixture between these states of matter. Depending on their composition and properties, such as size, shape, surface, etc., they can have different effects on the Earth's energy budget: They can scatter incoming solar radiation leading to a decrease in surface temperature or they can absorb incoming solar radiation, thus, heating the atmosphere which, in turn, again decreases the surface temperature. Both these processes are referred to as the DIRECT effect of aerosols. Furthermore, aerosols can influence cloud formation as well as the hydrologic circle and thereby have an INDIRECT EFFECT on the radiative transfer between Earth and the surrounding. [7]

Aerosols originate from various sources, such as sea spray, plants, mineral dust and volcanoes, but also from combustion processes, such as forest fires, combustion of fossil fuels etc. [10]. The 1991 eruption of the Philippine volcano Mount Pinatubo, during which about 17 megatons of SO_2 were erupted into the atmosphere [19], is an impressive example of the impact aerosols can have on the Earth's climate. Over the course of that year, the aerosol cloud spread out until it had formed a global layer of sulfuric acid haze in the Earth's atmosphere. During subsequent years, global temperatures dropped by about 0.5 °C while ozone depletion substantially increased [19]. This exemplifies that aerosols can have a huge impact on the earth's climate and, consequently, continuous, efficient monitoring and classification of aerosols is essential to quantitatively assess the future of the Earth's climate as well as the human impact on climate change. But there are also numerous other areas, in which analyses of aerosols are desired, from the combat for clean air to reduce the approximate 9 million premature deaths worldwide that are annually caused by aerosols [15] through to the containment of pandemics.

To infer the microphysical properties of aerosols, different kinds of measurements are performed, such as satellite remote sensing [11], ground-based remote sensing, and in-situ measurements [7]. Most of these methods make use of the fact that a lot of information about the aerosols is contained in the angular dependence of the intensity and polarization of light scattered on the particles. Unfortunately, the retrieved parameters can be very sensitive to variations in the light scattering data, which renders the inverse problem ill-posed. Traditionally, the aerosol properties are retrieved from the scattering data either using pre-computed lookup tables, which is naturally fast but not accurate, or by physically-based iterative algorithms, such as GRASP-OPEN [5], which is accurate but not fast. Since Pluchino et al. (1980) [18], machine learning algorithms, i.e. algorithms using a nonlinear approximation of the multivariable function connecting the multi-angle light scattering data to the microphysical properties of the aerosol particles or vice versa, have been applied to various versions of the problem [3, 8, 9, 16, 21], with the hope to establish a third option that is both accurate and fast. A thorough discussion of the work done within the last 40 years would go beyond the scope of this report, therefore I refer the curious reader to the review paper of Berdnik and Loiko (2016) [4]. We must content ourselves with recognizing that neural networks are a promising tool for problems where a high speed of data processing is required without concessions regarding the quality of inference. Given that neural networks are universal approximators [12] that are amortized, they are particularly well suited: after an upfront computational cost at the simulation and training stage, the evaluation of new data points is cheap and efficient. It goes without saying that the recent advances in information technology have vastly extended the number of problems that can be tackled using artificial neural networks and, as a consequence, have contributed to the considerable momentum the field of simulation-based inference has gained.

The following report revolves around the application of an invertible neural network model conceived by Ardizzone et al. (2018) [1] to capture the interaction of light with aerosols. More precisely, the invertible neural network is used to simulate multi-angle light scattering data, concretely the scattering phase function and the polarized phase function, given a set of microphysical aerosol properties on one hand, and to solve the inverse problem of retrieving these microphysical properties from the scattering data on the other.

In Sec. 2, a short overview of the relevant microphysical properties of aerosols is given and the quantities commonly used to describe the scattering of light by aerosols are introduced. Subsequently, in Sec. 3, I give a brief description of the measurement device, thus, motivating the structure of the simulated data used to train the model, which is presented thereafter together with the applied preprocessing. Sec. 4 introduces the architecture of the invertible artificial neural network used to model the relation between the microphysical properties of aerosols and the light scattering data, followed by an overview of the domain space, i.e. the ranges

of values of the hyperparameters that we explored using random sampling. The performances of the best models we found in our investigation are presented in Sec. 5 and discussed in Sec. 6.

2 Aerosols

To understand the quantities that we are trying to relate using the invertible neural network, this section contains an introduction to some important microphysical properties of aerosols in Sec. 2.1 as well as the theoretical model commonly used to describe the scattering of light by a group of aerosol particles, given in Sec. 2.2.

2.1 Properties of aerosols

Aerosols, an abbreviation for *aero-solutions*, are suspensions of fine solid particles or liquid droplets in a gas, typically air. In order to be suspended for a sensible time, the particles must be small enough, conventionally between a few nanometres and up to 100 µm. [10]

The size of an aerosol is a very fundamental property. Depending on the measurement technique applied, there are various ways to define the size of an aerosol, such as the electrical mobility radius, stokes radius or the optical radius. The OPTICAL RADIUS, which we will use in this study and denote by r, is defined as the radius of a spherical particle delivering the same scattering data in a given optical measurement device detecting particles by their interaction with light. [10] Note that this generalizes the mathematical concept of a radius in the sense that it allows the assignment of radii to non-spherical particles, such as ellipsoids.

If measurements of atmospheric aerosol particle populations are taken, there are, unsurprisingly, always particles of various forms and sizes involved. This differs from the possibilities given in a laboratory, where the particle population can be controlled to be uniform up to a spread in particle diameter of a few per cent. Such a population of particles of roughly uniform size is referred to as MONODISPERSE; contrarily, populations consisting of particles of sizes distributed over a larger range are called POLYDISPERSE. Within these two categories, the particle populations can be described more specifically using a PARTICLE SIZE DISTRIBUTION FUNCTION (PSD). A simple method to obtain a particle size distribution function consists of taking a histogram of the number of particles in successive size intervals. With sufficiently fine intervals, the distribution would then become what is known as a differential size distribution. Another common approach is to use kernel density estimation with log-normal distribution functions. [10]

In this study, we consider a monodisperse size distribution function, represented by a log-normal distribution. To described this log-normal particle size distribution function, $n(r)$, the following three parameters are used: total volume concentration, V_{tot}, median optical radius, r_m, and geometric standard deviation, σ,

$$n(r) = \frac{V_{tot}}{\sqrt{2\pi}} \frac{1}{\sigma} \frac{1}{r} \exp\left[-\frac{(\ln r - \ln r_m)^2}{2\sigma^2} \right].$$

Another important property of an aerosol species is the REFRACTIVE INDEX. The refractive index of an aerosol indicates its ability to scatter and absorb light. It is commonly expressed in a complex form as $m_\lambda = n_\lambda + ik_\lambda$, where the real term, n_λ, indicates scattering and the imaginary term, k_λ, indicates absorption [20]. Note that both the real and the imaginary part of the refractive index depend on the wavelength of the interacting light, which we shall denote as λ.

2.2 The theoretical model of the scattering of light

Electromagnetic radiance can be described by a tuple of values known as the STOKES VECTOR

$$\boldsymbol{I} = \begin{pmatrix} I & Q & U & V \end{pmatrix}^T$$

with I denoting the total intensity of the beam, Q a measure of polarization along 0° and 90° degree in a plane of reference perpendicular to the direction of propagation, U a measure of polarization along 45° and 135° degree in said plane of reference and V a measure of the direction of angular rotation.

In order to describe a scattering event, it is common to relate the stokes vector of the incident light beam with the stokes vector describing the scattered light per angle using the SCATTERING MATRIX $\boldsymbol{P}(\theta, \varphi, \lambda; p)$,

i.e. $\boldsymbol{I}_{\text{scat}}(\theta, \varphi, \lambda; p) = \boldsymbol{P}(\theta, \varphi, \lambda; p)\boldsymbol{I}_{\text{in}}(\lambda)$ with φ denoting the azimuthal and θ the polar angle for an incident beam of light at wavelength λ focused along $\theta = 0$ on an ensemble of particles at the origin with microphysical properties p. Componentwise, this can be written as

$$\begin{pmatrix} I_{\text{scat}} \\ Q_{\text{scat}} \\ U_{\text{scat}} \\ V_{\text{scat}} \end{pmatrix} = K \begin{pmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \\ P_{41} & P_{42} & P_{43} & P_{44} \end{pmatrix} \begin{pmatrix} I_{\text{in}} \\ Q_{\text{in}} \\ U_{\text{in}} \\ V_{\text{in}} \end{pmatrix}$$

with K a constant depending on the measurement process that is of no concern for our computational investigation of the problem. For a spherically symmetric particle, the relation simplifies to $\boldsymbol{I}_{\text{scat}}(\theta, \lambda; p) = \boldsymbol{P}(\theta, \lambda; p)\boldsymbol{I}_{\text{in}}(\lambda)$, which in particular implies that $P_{12} = P_{21}$ [14]. Further, in the case of unpolarized incident light, $\boldsymbol{I}_{\text{in}} = \begin{pmatrix} I_{\text{in}} & 0 & 0 & 0 \end{pmatrix}^T$, and given that the PSI polar nephelometer merely measures the intensity, I_{scat}, and the degree of linear polarization along 0° and 90° degree in the plane of reference, Q_{scat}, the above system of equations reduces to

$$\begin{cases} I_{\text{scat}} &= KP_{11}(\theta, \lambda; p)I_{\text{in}} \quad \text{and} \\ Q_{\text{scat}} &= KP_{12}(\theta, \lambda; p)I_{\text{in}}. \end{cases}$$

Lastly we need to consider the the POLARIZED PHASE FUNCTION (PPF), a common tool in this field used to describe polarization which is defined as $P_{\text{ppf}}(\theta, \lambda; p) = -\frac{P_{12}}{P_{11}}$.

Measurement data obtained by the nephelometer at PSI can provide us with both the coefficient relating the incoming intensity with the intensity of the scattered light per angle, $P_{11}(\theta, \lambda; p)$, which is referred to as the SCATTERING PHASE FUNCTION, as well as with the polarized phase function, $P_{\text{ppf}}(\theta, \lambda; p)$, containing the degree of polarization for the scattered light per angle. Therefore, we obtain

$$P_{11}(\theta, \lambda; p) = \frac{1}{K}\frac{I_{\text{scat}}(\theta, \lambda; p)}{I_{\text{in}}(\theta, \lambda)} \qquad P_{\text{ppf}}(\theta, \lambda; p) = -\frac{P_{12}}{P_{11}} = -\frac{Q_{\text{scat}}(\theta, \lambda; p)}{I_{\text{scat}}(\theta, \lambda; p)}. \tag{1}$$

3 Synthetic data set and experimental outlook

To motivate the structure of the synthetic data set used to train the invertible model, we will start this section by considering the capabilities of the polar nephelometer that is currently being calibrated at PSI. Subsequently, in Sec. 3.2, we will examine properties of the simulated data set, thus, motivating the preprocessing procedure applied to the data presented in Sec. 3.3.

3.1 PSI polar nephelometer

The PSI polar nephelometer [14] is based on the Polarized Imaging Nephelometer [6], depicted in Figure 1, which was developed and built at the Laboratory for Aerosols, Clouds and Optics of the University of Maryland, Baltimore County. Like its predecessor, the PSI polar nephelometer is designed to measure both quantities given in Equation 1, the scattering phase functions, $P_{11}(\theta, \lambda; p)$, and polarized phase functions, $P_{\text{ppf}}(\theta, \lambda; p)$. Initially, the instrument will only be able to measure using light at the wavelength of 532 nm. Ultimately, however, it is meant to perform measurements at three different wavelengths, namely $\lambda \in \Lambda = \{450\,\text{nm}, 532\,\text{nm}, 630\,\text{nm}\}$. Accordingly, we will consider both cases: the case in which we only use scattering data originating from measurements using light at the wavelength of 532 nm, and the other case, in which we assume to have measurements of all of the three wavelengths at our disposal.

It is expected that P_{11} and P_{ppf} can both be measured at an angular resolution of approximately 1°. Due to limitations in the physical design of the device, measurements can not be performed at extreme forward and backward angles, beyond the so-called TRUNCATION ANGLES. We will consider truncation angles of 5° and 175°. In addition, due to the need to mitigate stray light effects, measurements also can not be performed over the range from 85° to 95°.

Concerning the uncertainty in the measurement data, we assume the measurement device to have normally distributed random measurement errors for simplicity. We take the values from Dolgos and Martins (2014) on

the precursor instrument to the PSI polar nephelometer as a reference, giving 2.8 % relative error in P_{11} and 0.046 absolute error in P_{12} for a significance level of $p = 0.95$ [6].

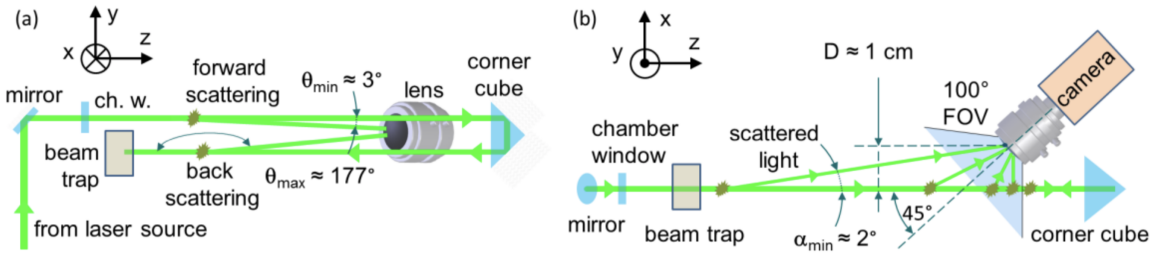

Figure 1: Schematic of the imaging arrangement employed in the scattering chamber of the Polarized Imaging Nephelometer [6].

3.2 Simulated data

The data used to train the invertible neural network was simulated using the Generalized Retrieval of Atmosphere and Surface Properties (GRASP) algorithm [5]. Both the microphysical aerosol properties as well as the light scattering data are contained in matrices, with each row corresponding to one sample and each column representing a physical quantity. More specifically, $\boldsymbol{X}_D$ contains the microphysical quantities V_{tot}, r_m, σ, n_λ, and k_λ, while $\boldsymbol{Y}_D$ contains the quantities $P_{11}(\theta_l, \lambda; x_i)$ and $P_{\text{ppf}}(\theta_l, \lambda; x_i)$; both for $\theta_l \in [5°, 85°] \cup [95°, 185°]$, $l \in \{1, \ldots, L = 160\}$, $\lambda \in \Lambda \subseteq \{450\,\text{nm}, 532\,\text{nm}, 630\,\text{nm}\}$ and x_i a configuration in parameter space $\boldsymbol{X}_D$ for each sample $i \in \{1, \ldots, N = 100\,000\}$. The unimodal data set contains 100 000 samples. Figure 2 shows the maximum, the mean and the minimum per angle of the design variables $\boldsymbol{Y}_D$.

The aerosol particle properties of interest, their units and the respective intervals from which they were sampled using Latin Hypercube sampling are given in Table 1. To prevent the sampling of unphysical combinations of microphysical aerosol properties, as for example $n_{450} = 1.4$ and $n_{532} = 1.6$ [14], the approximate constraint was applied that the real parts of the refractive indices are all equal, i.e. $n_{450} = n_{532} = n_{630}$. To elude this limitation of being unable to retrieve different real refractive indices for different wavelengths, one could use a prior probability suppressing unphysical parameter combinations when sampling the parameters needed for the simulation of the training data set. Determining such a prior of course already requires a certain knowledge about the manifold we are trying to investigate, making this seemingly to some kind of the chicken or the egg problem. It should, however, be possible to gradually approximate the prior by excluding combinations that are known to be unphysical and iteratively improving the prior using density estimations of microphysical aerosol properties obtained via physically-based iterative algorithms, such as GRASP-OPEN, from experimental scattering data.

Table 1: Microphysical particle properties, their units and the respective intervals from which they ware sampled using Latin Hypercube sampling. Explicitly, these are the total volume concentration, V_{tot}, the median optical radius, r_m, the geometric standard deviation, σ, and the real and imaginary part of the refractive index, n_λ and k_λ respectively.

Variable	Unit	Interval
V_{tot}	$(\mu\text{m}^3\text{cm}^{-3})$	$[1, 5000]$
r_m	µm	$[0.1, 2.5]$
σ	µm	$[1.4, 1.45]$
n_λ		$[1.33, 1.6]$
k_λ		$[0, 0.2]$

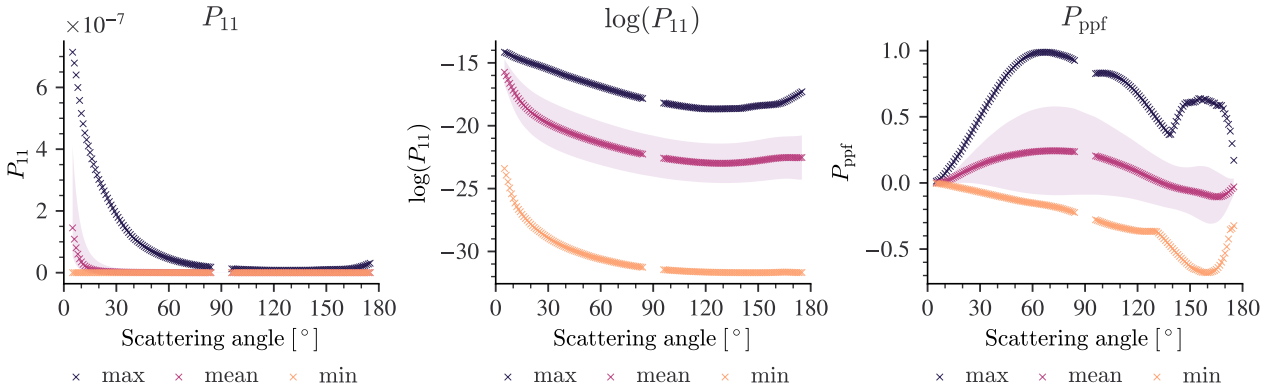

Figure 2: Maximum, mean, and minimum per angle of $P_{11}, \log(P_{11})$ and P_{ppf} for $\lambda = 532\,\text{nm}$. The filled area round the respective means represents the standard deviation per angle.

3.3 Preprocessing the data

As is usual in machine learning, we split the respective data set into a training data set and a test set with the common ratio of 80/20. During training, the model only sees the training data, which is in turn split into a training and a validation set at each epoch. Evaluating the model that scores best during training on the test set after training, gives the possibility to probe how the model performs on data that it has not seen yet, thus, delivering a measure of how well the model is expected to generalize.

To investigate the gain in information achievable by using the scattering data $\{\log(P_{11})(\boldsymbol{\theta}, \lambda; p), P_{\text{ppf}}(\boldsymbol{\theta}, \lambda; p)\}_{\lambda \in \Lambda}$ of different wavelengths simultaneously, we considered the Pearson product-momentum correlation depicted in Figure 3. Unsurprisingly, it turns out that the correlation between pairs of wavelengths close to each other, i.e. $\{450\,\text{nm}, 532\,\text{nm}\}$ and $\{532\,\text{nm}, 630\,\text{nm}\}$, is stronger than the correlation between $\{450\,\text{nm}, 630\,\text{nm}\}$. Considering this strong degree of correlation between the latter two wavelengths, it becomes apparent that the choice about the number of wavelengths used per measurement constitutes a trade-off between precision and efficiency. For a fast retrieval of microphysical aerosol properties from samples, it might be experimentally preferable to perform measurements using as few wavelengths as possible, i.e. only $\{532\,\text{nm}\}$ or $\{450\,\text{nm}, 630\,\text{nm}\}$, since this would reduce the amount of measurement data that has to be taken. For simulating the scattering data given some microphysical aerosol properties on the other hand, it is presumably preferable to be cable of simulating P_{11}, P_{ppf} for as many wavelengths as possible. The impossibility of satisfying both of these potential interests simultaneously with one network, is a fundamental limitation of this approach and for that matter probably most approaches deploying invertible neural networks to solve this inverse problem.

The strong correlations between pairs of wavelengths that are closer to each other might, however, be of theoretical interest, for they suggest that the dependence of the scattering and the polarized phase function on the wavelength could potentially be learnt conveniently by training a model that takes the wavelength of the incident light beam as an additional input and output parameter. This would allow not only to retrieve particle properties using an incident light beam of arbitrary wavelength within the given learnt range but also allow to simulate the scattering data as a continuous function of the wavelength. Realizing this would allow cheap explorations of the manifold $\{P_{11}, P_{\text{ppf}}\}$ enabling further theoretical studies, I conclude that this task goes beyond the scope of this project, but might be an interesting topic for future research.

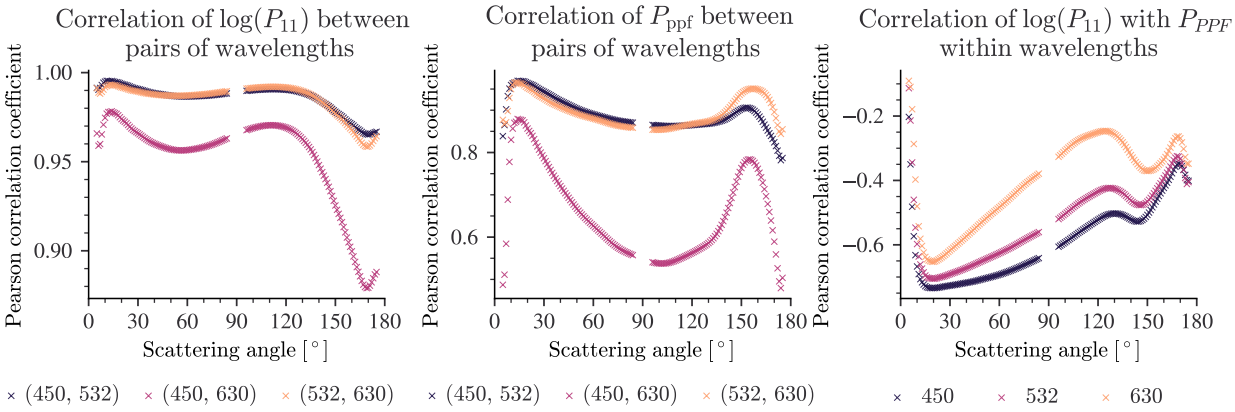

Figure 3: Pearson product-moment correlation coefficient of $\log(P_{11})$ and P_{ppf} per angle of measurement. In the first two figures, the correlation coefficient between different wavelengths is plotted against the angle of measurement, once for $\log(P_{11})$, once for P_{ppf}. In the third figure, the correlation coefficient between $\log(P_{11})$ and P_{ppf} is plotted against the angle of measurement within different wavelengths. The first two figures show that there is a strong correlation between pairs of wavelength that are neighbouring, namely between $\{450\,\text{nm}, 532\,\text{nm}\}$ and $\{532\,\text{nm}, 630\,\text{nm}\}$. The third figure shows that for larger wavelengths $\log(P_{11})$ and $\log(P_{\text{ppf}})$ are less correlated.

Figure 4 shows different transformations of the data that we considered. Ultimately, we found our best model by applying the logarithm to P_{11} and then re-scaling the features StandardScaler($\{\log(P_{11}), P_{\text{ppf}}\}$) [17] to a mean of 0 and a standard deviation of 1 for each angle. The application of the logarithm to P_{11} was motivated by the fact that the variance of $\log(P_{11})$ is higher than the variance of P_{11} which can be seen by comparing the first column of panels in Figure 4 with the second, as well as by comparing the first panel of Figure 2 with the second.

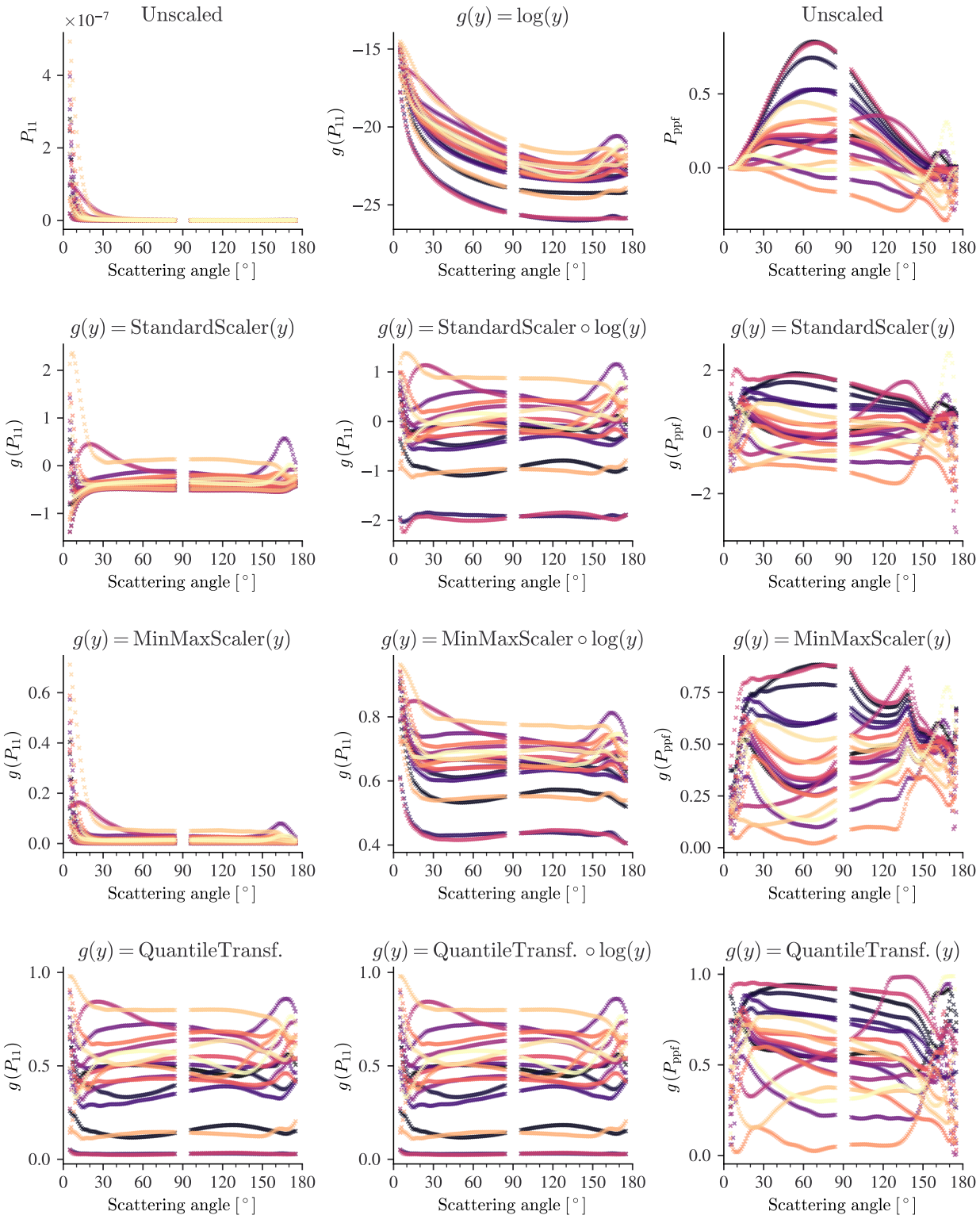

Figure 4: The effects of preprocessing on P_{11} and P_{ppf} for a series of particles for an incident light beam of wavelength 532 nm. The left column contains P_{11} processed by different scalers, the second column $\log(P_{11})$ processed by different scalers and the third column P_{ppf} transformed by different scalers. It becomes apparent that the application of the logarithm on P_{11} helps to better differentiate the curves for all scalers but the QuantileTransformer. The latter seems to spread the curves most uniformly over the interval $[0, 1]$.

4 The statistical model

This section starts with a description of the invertible neural network deployed to relate the microphysical properties of the aerosol particles with the light scattering data. Subsequently, in Sec. 4.2, I give details on the implementation as well as an overview of the hyperparameter space we probed to find the models presented in Sec. 5.

As introduced in Section 3.1, we will continue to denote the list of all sampled microphyiscal aerosol properties by $\boldsymbol{X}_D$; while we denote the list of all simulated light scattering data by $\boldsymbol{Y}_D$. Explicitly,

$$\boldsymbol{X}_D = \left\{\{V, r_m, \sigma, n_\lambda, k_\lambda\}_{\lambda\in\Lambda}(p_i)\right\}_{1\leq i\leq N} \qquad \boldsymbol{Y}_D = \left\{\{P_{11}(\theta_l, p_i, \lambda),\, P_{\mathrm{ppf}}(\theta_l, p_i, \lambda)\}_{\lambda\in\Lambda, 1\leq l\leq L}\right\}_{1\leq i\leq N},$$

Further, we will denote the smallest connected manifold with closure that contains the ranges of microphysical parameters from which we sampled $\boldsymbol{X}_D$ with $\boldsymbol{X}$ and the smallest connected manifold with closure that contains the ranges of the simulated values of $\boldsymbol{Y}_D$ with $\boldsymbol{Y}$. We will also call the mapping $f : \boldsymbol{X} \to \boldsymbol{Y}$ the FORWARD PROCESS, and the retrieval from aerosol properties given the light scattering data the corresponding INVERSE PROCESS.

In general, the inverse process is intractable and not well-defined, for the function describing the forward process must not be injective. Consequently, the inverse process should stochastically select a $\boldsymbol{X} \in f^{-1}(\boldsymbol{Y})$ according to the distribution of the complete posterior $p(\boldsymbol{X}|\boldsymbol{Y})$. To compensate for this loss of information inherent to the forward process', additional latent output variables z are introduced to capture the information about $x \in \boldsymbol{X}$ that is not contained in the corresponding light scattering data $y_x \in \boldsymbol{Y}$. During the inverse process, these latent variables are sampled from a standard multivariate normal distribution, i.e. $z \sim \mathcal{N}(0, \mathrm{id}_{d_z})$, with d_z denoting the dimension of the latent space $\boldsymbol{Z}$, which is another hyperparameter. In short, the invertible neural network learns to associate hidden parameter values x with distinct pairs of measurements and latent variables $[y, z]$. Ignoring the additional padding of both $\boldsymbol{X}$ and $\boldsymbol{Y}$ with zeros to further increase the internal flexibility of the model and to guarantee that input and output dimensions are the same, we thus have

$$\hat{f}_D : \boldsymbol{X} \subset \mathbb{R}^{3+2|\Lambda|} \to \boldsymbol{Y} \times \boldsymbol{Z} \subset \mathbb{R}^{2L|\Lambda|} \times \mathbb{R}^{d_z}; \qquad x \mapsto [y, z]$$

$$\hat{g}_D : \boldsymbol{Y} \times \boldsymbol{Z} \subset \mathbb{R}^{2L|\Lambda|} \times \mathbb{R}^{d_z} \to \boldsymbol{X} \subset \mathbb{R}^{3+2|\Lambda|}; \qquad [y, z] \mapsto x$$

such that $\hat{g} \circ \hat{f} = \mathrm{id}$.

4.1 The invertible neural network model

The following characterization and discussion of invertible neural networks, INNs for short, is closely following the paper of Ardizzone et al. (2018), which the eager reader is encouraged to read in order to get a full treatment of the subject. This being said, the class of invertible neural networks is characterized by three properties:

- The mapping from the input domain to the output domain is bijective so that an inverse exists.
- Both the forward and the inverse mapping are cheap, i.e. efficiently computable.
- The Jacobian of both mappings is tractable, allowing for the explicit computation of posterior probabilities.

The advantage of networks that are invertible by construction is the fact that training them on the forward process $\hat{f}_D : \boldsymbol{X} \to \boldsymbol{Y} \times \boldsymbol{Z}$ delivers the inverse $\hat{g}_D : \boldsymbol{Y} \times \boldsymbol{Z} \to \boldsymbol{X}$ for free simply by running them backwards; since both the forward process and the inverse process are performed by the same network with the same weights. Given that fully connected feedforward networks are not invertible, Ardizzone et al. (2018) suggested building blocks consisting of an Affine Coupling Block (ACB) followed by a permutation layer. As both ACBs and permutations are invertible, so is their composition, which we will refer to as a block, as well as any concatenations of such blocks, which in turn make up the model, as illustrated in Figure 5.

In the forward process of an ACB, the block's input vector u is split into two halves, u_1 and u_2, which are then both transformed by an affine function with coefficients $\exp(s_i)$ and t_i for $i \in \{1, 2\}$, using element-wise multiplication, $\odot$, and addition, $+$,

$$\mathbf{v}_1 = \mathbf{u}_1 \odot \exp\left(s_2\left(\mathbf{u}_2\right)\right) + t_2\left(\mathbf{u}_2\right), \qquad \mathbf{v}_2 = \mathbf{u}_2 \odot \exp\left(s_1\left(\mathbf{v}_1\right)\right) + t_1\left(\mathbf{v}_1\right)$$

For the corresponding inverse process, the output is split into two halves $\mathbf{v} = [\mathbf{v}_1, \mathbf{v}_2]$, and the inverse of the above relations is after simple algebraic transformations found to be

$$\mathbf{u}_2 = (\mathbf{v}_2 - t_1(\mathbf{v}_1)) \odot \exp(-s_1(\mathbf{v}_1)), \qquad \mathbf{u}_1 = (\mathbf{v}_1 - t_2(\mathbf{u}_2)) \odot \exp(-s_2(\mathbf{u}_2))$$

The coefficient functions s_i, t_i can be any function, as for example a fully dense neural network, and must not be invertible. The permutation layers are needed in order to shuffle the elements of the subsequent layer's input in a randomized but stationary way, leading to a different split of variables $u = [u_1, u_2]$ for every block enhancing the interaction among the variables and, thus, the expressiveness of the model.

To facilitate the learning of complex models for small dimensional parameter spaces $\boldsymbol{X}$, Ardizzone et al. (2018) [1] suggests to pad both the in- and output of the network with an equal number of zeros and, thus, enable the network's interior layers to embed the data into a representation space of higher dimension.

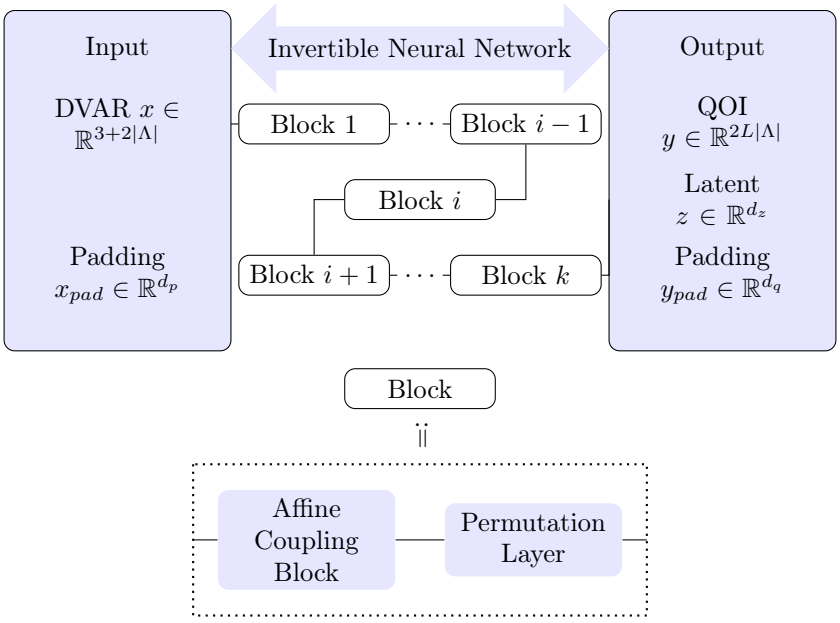

Figure 5: Schematic of the architecture of the invertible neural network. Each block consists of an affine coupling block followed by a permutation layer.

The loss function used to train the network is the composition of numerous loss functions:

$$\mathcal{L}_{inv} = w_x\mathcal{L}_x + w_y\mathcal{L}_y + w_z\mathcal{L}_z + w_r\mathcal{L}_r + w_p\mathcal{L}_p.$$

The first loss, $\mathcal{L}_x$, ensures that the sampled design variable distribution $\hat{p}_X(X)$ matches the one of the data set, $p_X(X)$. $\mathcal{L}_y$ guarantees that the relation between the microphysical particle properties and the light scattering data is captured by the model. Ardizzone et al. (2018) suggest using the mean squared error, which we did, i.e. $\mathcal{L}_y = \mathbb{E}[(y - \hat{f}_D(x_y))^2]$. $\mathcal{L}_z$ makes sure that the model samples latent vectors z from the desired normal distribution $p(z)$. Additionally, the function guarantees that the latent variables are sampled independently of the desired quantities of interest y, i.e. $p(z|y) = p(z)$, thus, preventing the redundant encoding of information. $\mathcal{L}_z$ is implemented by Maximum Mean Discrepancy, a kernel-based framework for comparing distributions only accessible through samples. $\mathcal{L}_r$ ensures the robustness of the model with respect to small perturbations in the measurement data, which we modeled using Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma_r)^N$. The loss is thus given by $\mathcal{L}_r = \mathbb{E}[(g(\hat{f}_y(x) + \epsilon) - x)^2]$. Lastly $\mathcal{L}_p$ makes sure that the amplitude of the noise fed into the model through the padding dimensions is low, i.e. $\mathcal{L}_p = \sum_{n \in \text{pad}} \|x_n\|^2 + \|y\|^2$, to assure that no information is encoded there.

4.2 Implementation and hyperparameters

We were able to adapt a Python implementation of the invertible neural network of Bellotti (2020) [2], who used it to solve an inverse problem involving the Argonne Wakefield Accelerator.

To find good hyperparameters, we used random search together with the Asynchronous Successive Halving Algorithm (ASHA) developed by [13]. ASHA is a scheduler that allows for a broad, parallel exploration of the hyperparameter space by early stopping models scoring worse than a certain fraction, in our case one fourth, of all the models that have been trained to the respective epoch of comparison. We compared the quality of the models with respect to mean absolute error of the forward process.

The ranges of the hyperparameter space that we investigated using random search is given in Table 2. As activation function for each neuron we used ReLUs. For optimization, we used Adam with default parameters, a constant learning rate of 5×10^{-5} and a batch size of 8.

Table 2: The part of the hyperparameter space that we investigated using random search.

Wavelengths	Depth	Width	Blocks	w_x	w_y	w_z	w_r	w_p	σ_r
$\Lambda = \{532\}$	$[2, 4]$	$[40, 90]$	[3, 4]	$[100, 200]$	350	$[180, 350]$	$[220, 350]$	$[0, 20]$	$[0.05, 0.3]$
$\Lambda = \{450, 630\}$	$[2, 4]$	$[80, 130]$	[3, 4]	$[100, 200]$	350	$[180, 350]$	$[220, 350]$	$[0, 20]$	$[0.05, 0.3]$
$\Lambda = \{450, 532, 630\}$	$[2, 4]$	$[140, 200]$	[3, 4]	$[100, 200]$	350	$[180, 300]$	$[220, 320]$	$[0, 20]$	$[0.05, 0.3]$

5 Results

As motivated in Sec. 3, we performed three hyperparameter scans on three different subsets of the simulated data corresponding to three different sets of wavelengths. The first model was only trained on data originating from one wavelength, namely 532 nm. Its performance is described in Sec. 5.1. The second model was trained on data originating from two wavelengths, namely 450 nm and 630 nm. Its performance is discussed in Sec. 5.2. The third model was trained on data originating from all of these three wavelengths. Details on its performance are given in Sec. 5.3.

Each of these subsections describing a model is divided in four parts: one describing the training history of the model, one committed to the performance of the forward process on the test set, one committed to the performance of the inverse process on the test set, and, lastly one containing a qualitative description of the predictions made by the best model.

To evaluate the performance of the models, we use the coefficient of determination, defined as 1 minus the residual sum of squares divided by the total sum of squares

$$R^2 = 1 - \frac{\sum_{i=1}^{M} (z_i - \hat{z}_i)^2}{\sum_{i=1}^{M} (z_i - \bar{z})^2}$$

for a given data set $\{(z_i, \hat{z}_i)\}_{1 \leq i \leq M}$ with $\hat{z}_i$ denoting the prediction of the respective quantity from the test data set, z_i, and $\bar{z} = \frac{1}{M} \sum_{i=1}^{M} z_i$ the mean of the respective quantity from the test data set of size $M = 20\,000$.

To compare the models having trained on data from different sets of wavelengths among each other, we use the adjusted coefficient of determination, defined as $R^2_{adj} = 1 - \frac{(1-R^2)(M-1)}{M-k-1}$ with k denoting the number of explanatory variables in the model and M again the size of the test data set. We further use the absolute error, $\Delta_{abs}\hat{z}_i = |z_i - \hat{z}_i|$, and the relative error in per cent, $\Delta_{rel}\hat{z}_i = \frac{|z_i - \hat{z}_i|}{z_i} \cdot 100$, as well as percentile scores of these values.

5.1 Model trained on one wavelength

The following section discusses models trained on the subset of the data corresponding to the wavelength $\Lambda = \{532\,\text{nm}\}$.

The best of the 64 models trained within the scope of this hyperparameter scan was chosen according to the highest coefficient of determination, R^2, scored on the validation data set for the inverse process at the last epoch of training. Figure 6 shows the R^2 value of the validation data plotted against the R^2 value of the training data for each model at the end of training. This representation suggests that the forward process is slightly overfitting, while the inverse processes would be expected to generalize perfectly. It should further be noted that this hyperparameter scan delivered numerous models that would potentially be more accurate surrogates for the forward process than the one chosen following our simple selection criterion.

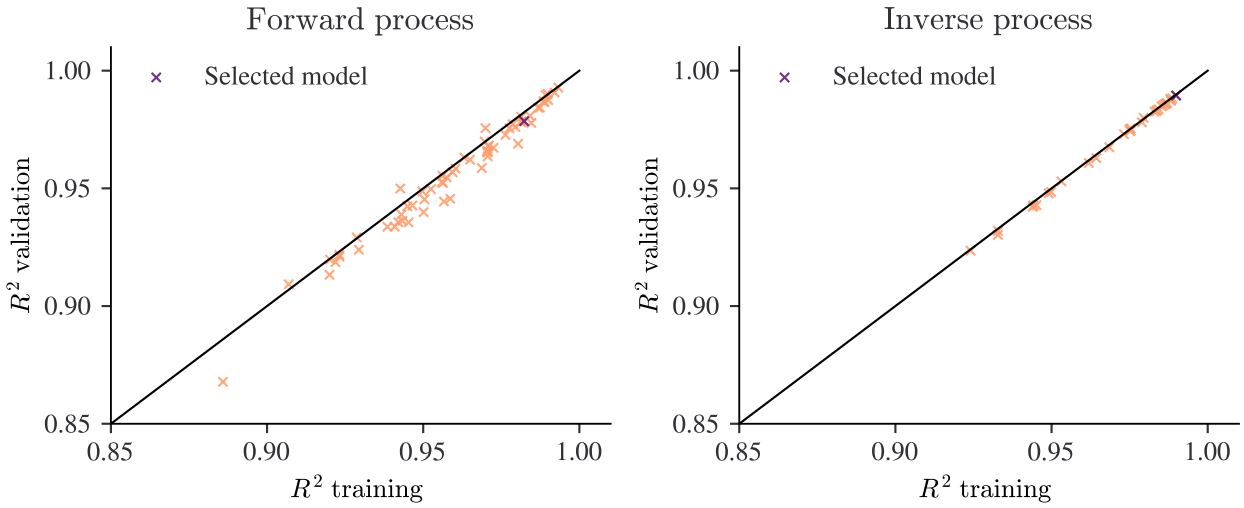

Figure 6: Results of the hyperparameter scan. The coefficient of determination, R^2, of the validation data is plotted against the R^2 value of the training data.

Table 3: Hyperparameters of the selected model.

Blocks	Depth	Width	Nominal dimension	w_p	w_r	w_x	w_y	w_z	σ_r
4	3	87	328	9	331	160	350	346	0.095

The hyperparameters of the selected model are given in Table 3. At the last epoch of training, the coefficients of determination of the forward process of this model scored $R^2_{train} = 0.982$ on the training data and $R^2_{val} = 0.979$ on the validation data. The coefficients of determination of the inverse process scored $R^2_{train} = 0.990$ on the training data and $R^2_{val} = 0.989$ on the validation. On one core of an Intel Xeon Gold 6152 processor, the average prediction time of the forward process of the selected model is 0.264 ms; the average prediction time of the inverse process 15.0 ms.

5.1.1 Training history

Figure 7 depicts the history of the coefficient of determination, R^2, of the forward and the inverse process. The maximum of R^2 over the 80 training epochs of the forward process is reached around epoch 27, the overall trend is ambiguous. Contrarily, the R^2 value of the inverse process is on the rise up until the last epoch. Figure 8 shows the evolution of the mean absolute error of both the forward and the inverse processes over 80 training epochs.

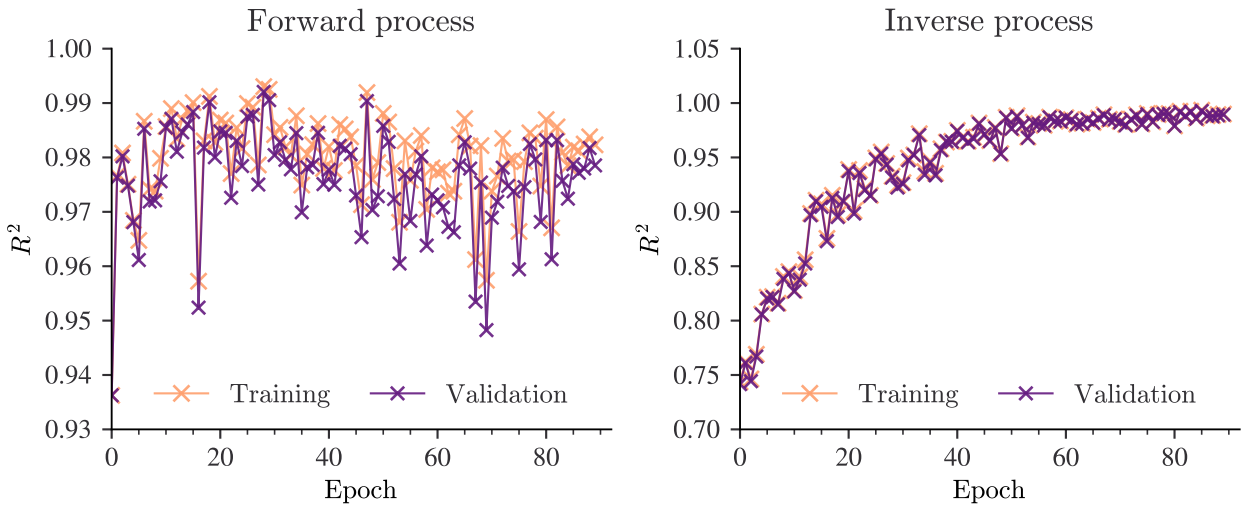

Figure 7: History of the coefficient of determination, R^2, of the forward and the inverse process.

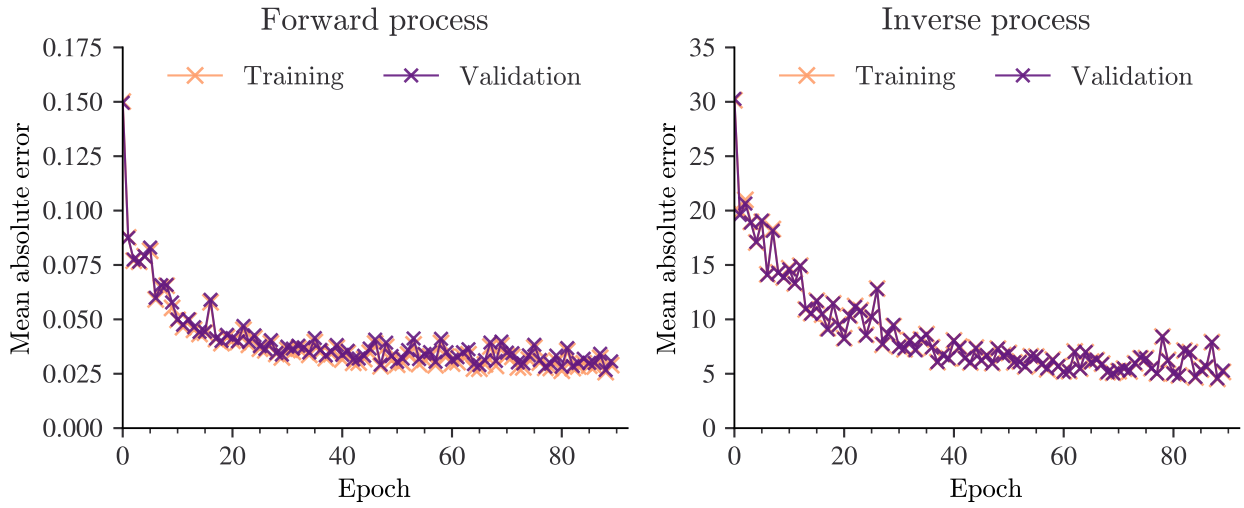

Figure 8: Evolution of the mean absolute error of both the forward and the inverse process over 80 training epochs.

5.1.2 Generalization of the forward process

The most compact assessment of the quality of the forward process is given in Table 4, which contains the maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel}\log(P_{11}(\theta))$ and of the absolute errors of the polarized phase function, $\Delta_{abs}P_{\text{ppf}}(\theta)$. A less compact but more nuanced representation of the accuracy is then given in Figure 9, which shows the 50th, 75th, 90th, 95th and 99th percentiles of $\Delta_{rel}\log(P_{11}(\theta))$ and $\Delta_{abs}P_{\text{ppf}}(\theta)$ as a function of the scattering angle. Note that besides broader trends in the errors across the angles, there are individual angles that significantly differ in accuracy from their respective neighbourhoods.

Figure 10 depicts the coefficient of determination, R^2, once for the logarithm of the scattering phase function, $\log(P_{11})$, and once for the polarized phase function, P_{ppf}, plotted against the scattering angles. Here can be observed that the angles with the lowest R^2 are close to 5° in the case of $\log(P_{11})$ and close to 175° in the case of P_{ppf}. Besides these broader trends across the angles, there are individual angles that significantly differ in accuracy from their respective neighbourhoods. The overall variation in R^2 is, however, merely of the order of 10^{-3}.

Table 4: The maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel}\log(P_{11}(\theta))$, on one hand and of the absolute errors of the polarized phase function, $\Delta_{abs}P_{\text{ppf}}(\theta)$, on the other. Both are rounded to three decimal places.

$\max_{\theta\in\Theta}$ of percentiles	$\Delta_{rel}\log(P_{11}(\theta))$ [%]	$\Delta_{abs}P_{\text{ppf}}(\theta)$
50%	0.145	0.005
75%	0.252	0.009
90%	0.380	0.013
95%	0.485	0.016
99%	1.108	0.035

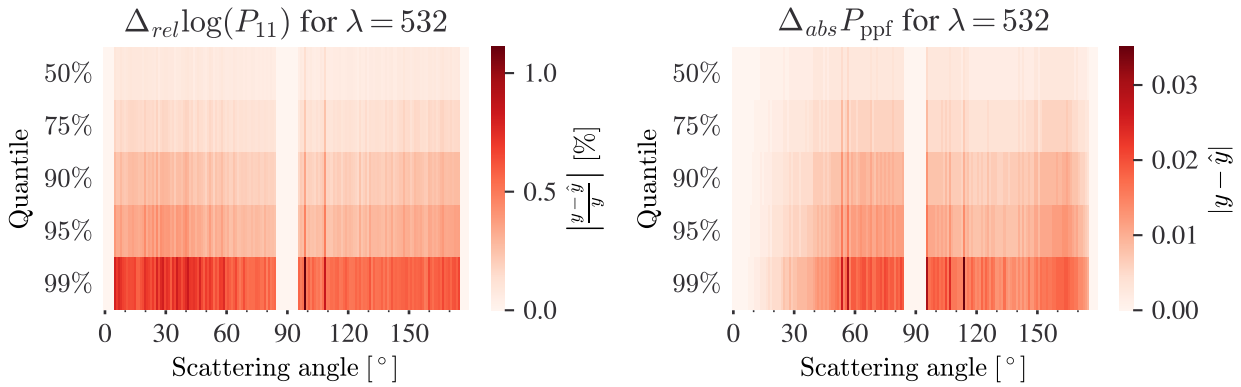

Figure 9: 50th, 75th, 90th, 95th and 99th percentiles of the relative error of the predicted logarithm of the scattering phase function, $\log(P_{11})$, and of the absolute error of the predicted polarized phase function, P_{ppf}, as a function of the scattering angle.

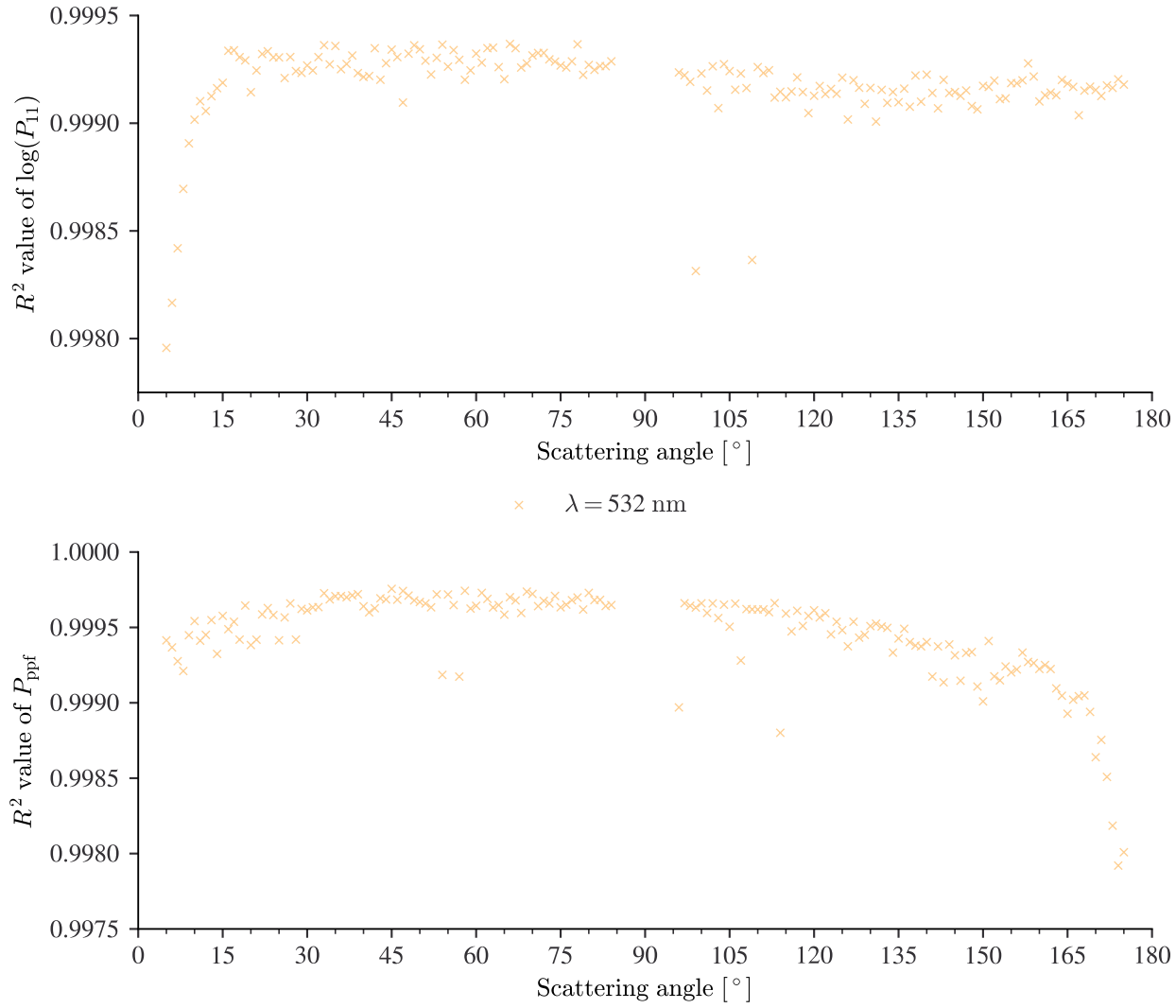

Figure 10: Coefficient of determination, R^2, once for $\log(P_{11})$ and once for P_{ppf} plotted against the scattering angles.

5.1.3 Generalization of the inverse process

To quantify the inaccuracy of the inverse process, Table 5 contains the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties. Given that some of these features vary over several orders of magnitude, it is difficult to compare the precision between features in this representation without consulting Table 1 listing the ranges of each of these features. For this reason there is an alternative representation given in Figure 11, where the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties are normalized by the respective range from which the parameters were sampled, i.e. $\Delta_{abs}\hat{y}_i/(\max_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i - \min_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i)$. According to this representation, the errors of the standard deviation σ of the log-normal function are significantly larger than for the other features. Also the R^2_{adj} value of σ is smaller than the R^2_{adj} values of the other microphysical particle properties given in Table 6.

Table 5: 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties rounded to three decimal places.

Percentiles of Δ_{abs}	n_{532}	k_{532}	V_{tot}	r_m	σ
50%	0.001	0.001	18.507	0.005	0.002
75%	0.002	0.001	34.576	0.009	0.003
90%	0.003	0.002	56.821	0.014	0.005
95%	0.003	0.002	74.291	0.018	0.006
99%	0.006	0.005	125.722	0.030	0.011

Figure 11: The absolute error of the retrieved microphysical aerosol particle properties, normalized by the respective range from which the parameters were sampled, i.e. $\Delta_{abs}\hat{y}_i/(\max_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i - \min_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i)$.

Table 6: R^2_{adj} of the retrieved microphysical aerosol particle properties rounded to two decimal places.

n_{532}	k_{532}	V_{tot}	r_m	σ
1.00	1.00	1.00	1.00	0.95

Figure 12 displays the microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. In this representation, samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity. The dispersion is the highest for the standard deviation σ of the unimodal log-normal size distribution function and the lowest for the median optical radius r_m.

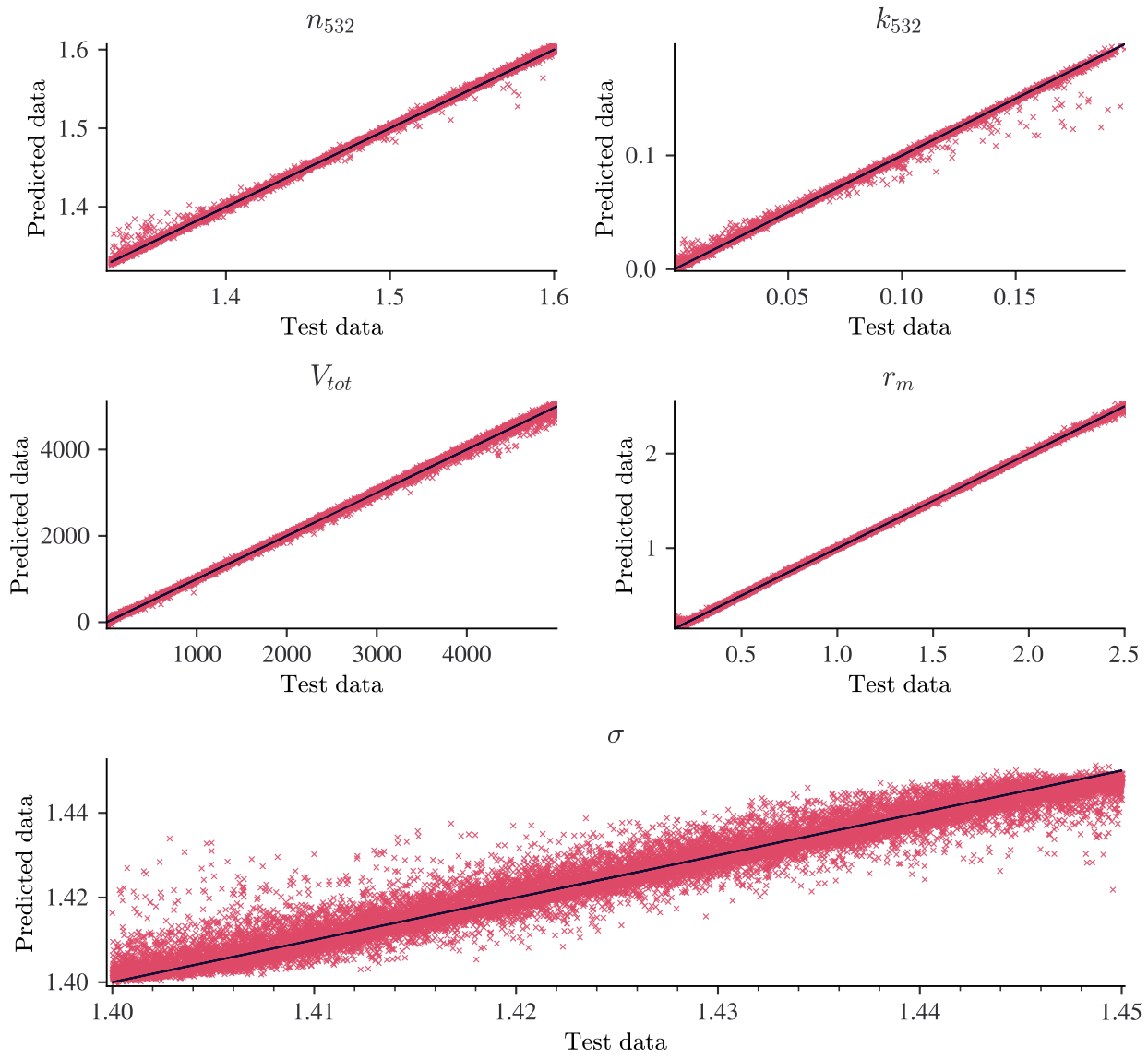

Figure 12: Microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. Samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity.

5.1.4 Qualitative comparison of predictions with test data

Lastly we are going to qualitatively consider the predictions made by our best model. Figure 13 shows five predictions of the forward process and their true counterparts from the test set. Although the predicted curves follow the general trends of the true curves from the test data, they contain high frequency variations that are absent in all of the depicted curves from the test set.

Figure 14 displays Five predictions of the log-normal size distribution delivered by the inverse process and their true counterparts from the test set. The prediction of the median optical radius r_m, the standard deviation σ and the total volume concentration V_{tot} seem to fit for all the depicted cases but for the the curve with the highest overall concentration, for which our prediction seems still slightly off with respect to σ and V_{tot}.

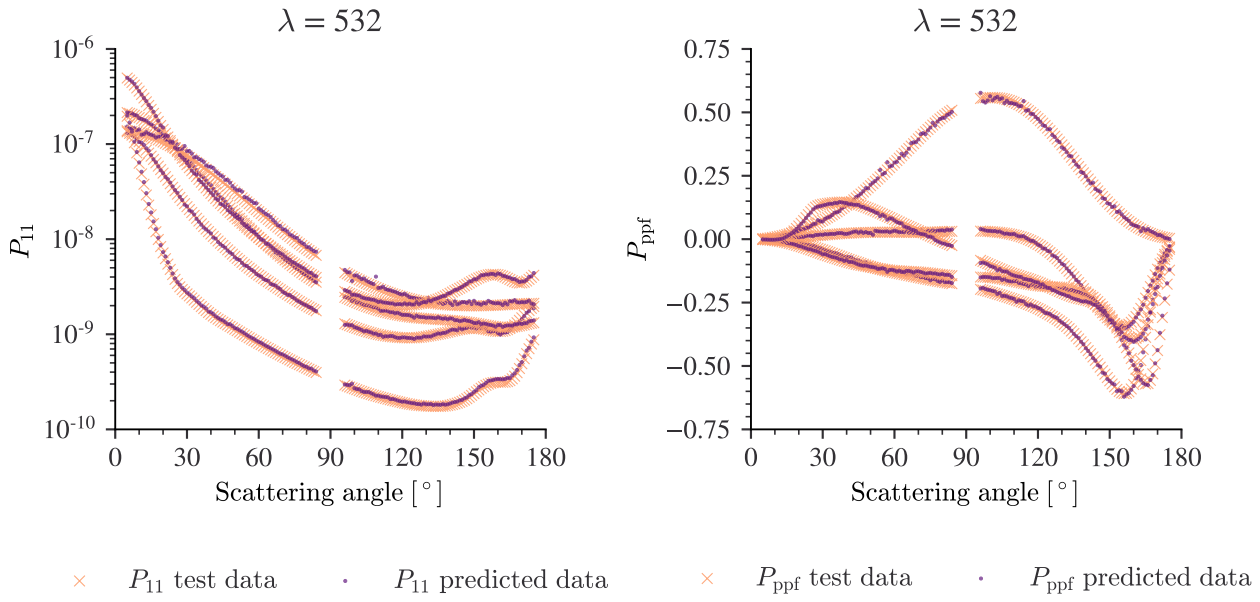

Figure 13: Qualitative comparison between curves that we predicted from sets of microphysical aerosol parameters from the test set and the respective curves simulated by GRASP-OPEN.

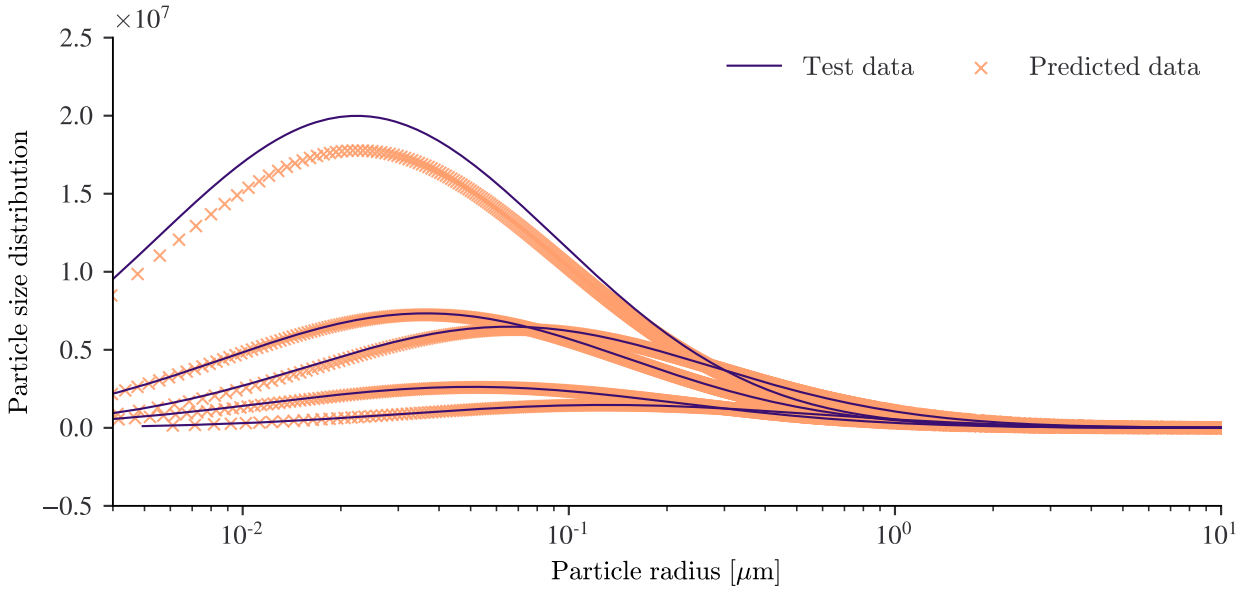

Figure 14: Qualitative comparison between the log-normal size distributions retrieved using our network when applied to five samples of light scattering data of the test set to the true log-normal size distributions that were used to simulate the respective samples of light scattering data.

5.2 Model trained on two wavelengths

The following section discusses models trained on the subset of the data corresponding to the wavelengths $\Lambda = \{450\,\text{nm}, 630\,\text{nm}\}$.

The best of the 30 models trained within the scope of this hyperparameter scan was chosen according to the highest coefficient of determination, R^2, scored on the validation data set for the inverse process at the last epoch of training. Figure 15 shows the R^2 value of the validation data plotted against the R^2 value of the training data for each model at the end of training. This representation suggests that the forward process is slightly overfitting, while the inverse processes would be expected to generalize perfectly. It should further be noted that this hyperparameter scan delivered numerous models that would potentially be more accurate surrogates for the forward process than the one chosen following our simple selection criterion.

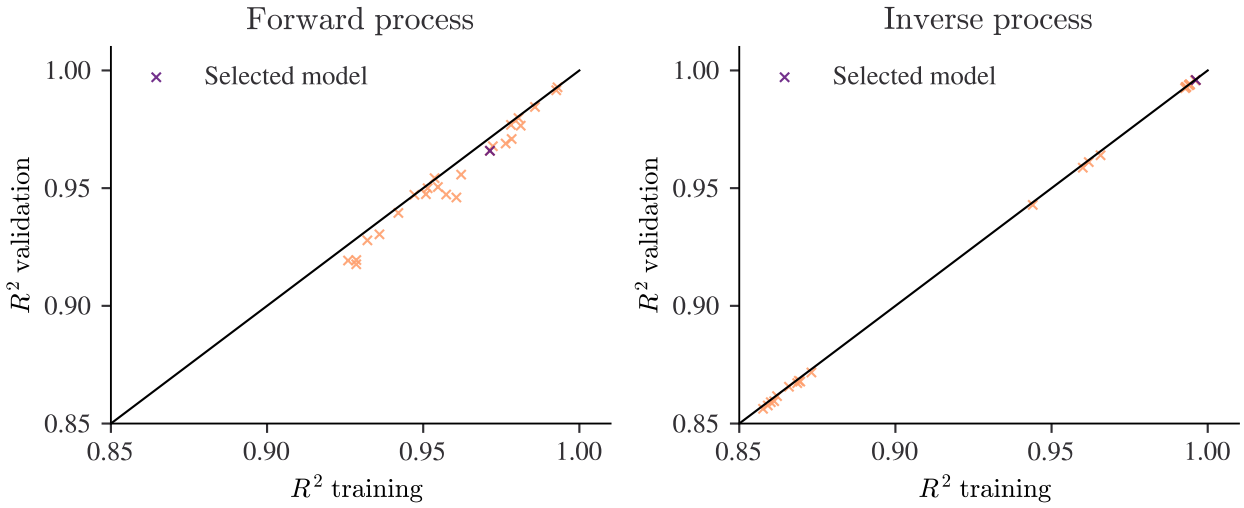

Figure 15: Results of the hyperparameter scan. The coefficient of determination, R^2, of the validation data is plotted against the R^2 value of the training data.

Table 7: Hyperparameters of the selected model.

Blocks	Depth	Width	Nominal dimension	w_p	w_r	w_x	w_y	w_z	σ_r
4	2	118	646	2	316	147	350	228	0.122

The hyperparameters of the selected model are given in Table 7. At the last epoch of training, the coefficients of determination of the forward process of this model scored $R^2_{train} = 0.971$ on the training data and $R^2_{val} = 0.966$ on the validation data. The coefficients of determination of the inverse process scored $R^2_{train} = 0.996$ on the training data and $R^2_{val} = 0.996$ on the validation. On one core of an Intel Xeon Gold 6152 processor, the average prediction time of the forward process of the selected model is 0.478 ms; the average prediction time of the inverse process 23.7 ms.

5.2.1 Training history

Figure 16 depicts the history of the coefficient of determination, R^2, of the forward and the inverse process. The maximum of R^2 over the 80 training epochs of the forward process is reached around epoch 37, the overall trend is ambiguous. Contrarily, the R^2 value of the inverse process is on the rise up until the last epoch. Figure 17 shows the evolution of the mean absolute error of both the forward and the inverse processes over 80 training epochs.

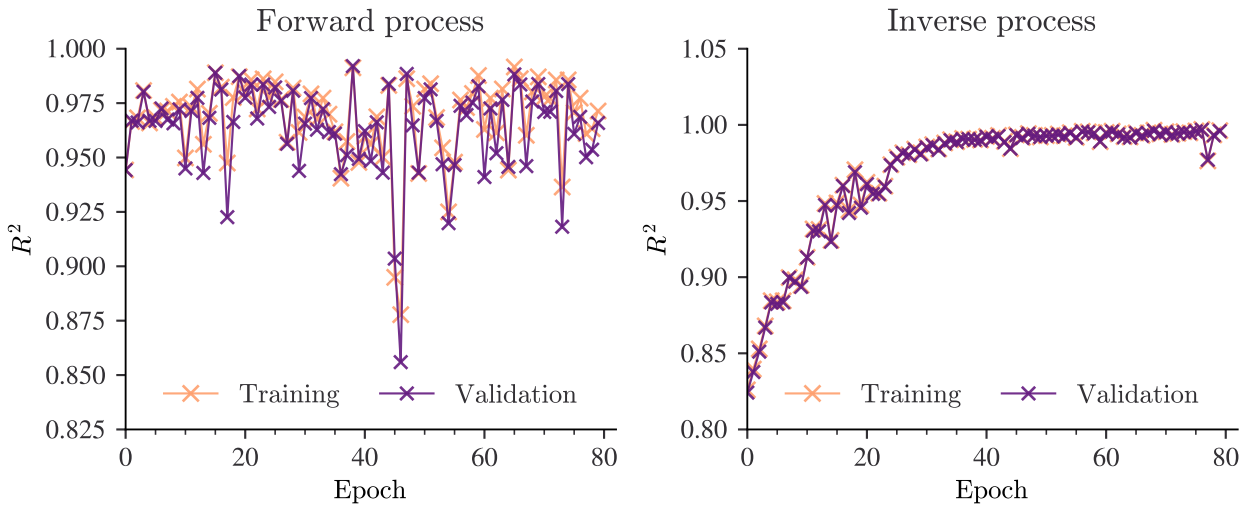

Figure 16: History of the coefficient of determination, R^2, of the forward and the inverse process.

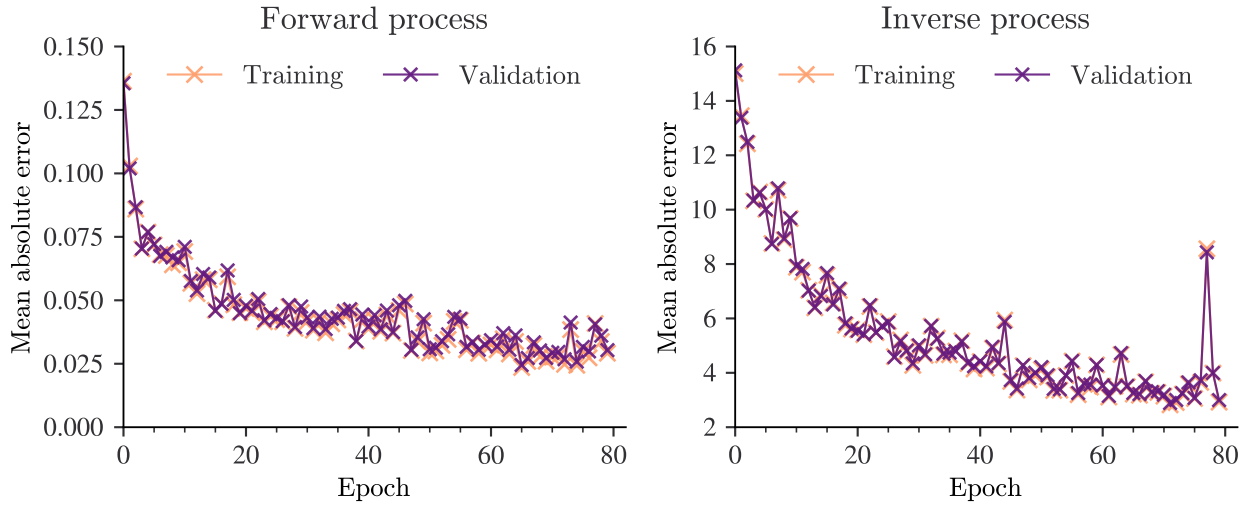

Figure 17: Evolution of the mean absolute error of both the forward and the inverse process over 80 training epochs.

5.2.2 Generalization of the forward process

The most compact assessment of the quality of the forward process is given in Table 8, which contains the maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel} \log(P_{11}(\theta))$, and of the absolute errors of the polarized phase function, $\Delta_{abs} P_{\text{ppf}}(\theta)$. A less compact but more nuanced representation of the accuracy is given in Figure 18, which shows the 50th, 75th, 90th, 95th and 99th percentiles of $\Delta_{rel} \log(P_{11}(\theta))$ and $\Delta_{abs} P_{\text{ppf}}(\theta)$ as a function of the scattering angle. Note that besides broader trends in the errors across the angles, there are individual angles that significantly differ in accuracy from their respective neighbourhoods. Overall, the predictions of P_{ppf} seem to be more accurate at the wavelength of $\lambda = 630$ compared to the wavelength of $\lambda = 450$, mainly due to the region of increased error between about 40° and 85° of P_{ppf} in the case of $\lambda = 450$.

Figure 19 depicts the coefficient of determination, R^2, once for the logarithm of the scattering phase function, $\log(P_{11})$, and once for the polarized phase function, P_{ppf}, plotted against the scattering angles. Here can be

observed that the angles with the lowest R^2 are close to 5° in the case of $\log(P_{11})$ and close to 175° in the case of P_{ppf}. Note again that besides these broader trends across the angles, there are individual angles, in particular for P_{11}, that significantly differ in accuracy from their respective neighbourhoods. The overall trends in accuracy are similar between the two wavelengths, with the exception of a small dip of R^2 of P_{ppf} of $\lambda = 630\,\text{nm}$ compared to $\lambda = 450\,\text{nm}$ around the scattering angle of 142°. The overall variation in R^2 is, however, merely of the order 10^{-3}.

Table 8: The maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel}\log(P_{11}(\theta))$, on one hand and of the absolute errors of the polarized phase function, $\Delta_{abs}P_{\text{ppf}}(\theta)$, on the other. Both are rounded to three decimal places.

$\max_{\theta\in\Theta}$	$\Delta_{rel}\log(P_{11}(\theta))$ [%]	$\Delta_{abs}P_{\text{ppf}}(\theta)$
50%	0.203	0.007
75%	0.350	0.013
90%	0.524	0.021
95%	0.665	0.028
99%	1.156	0.047

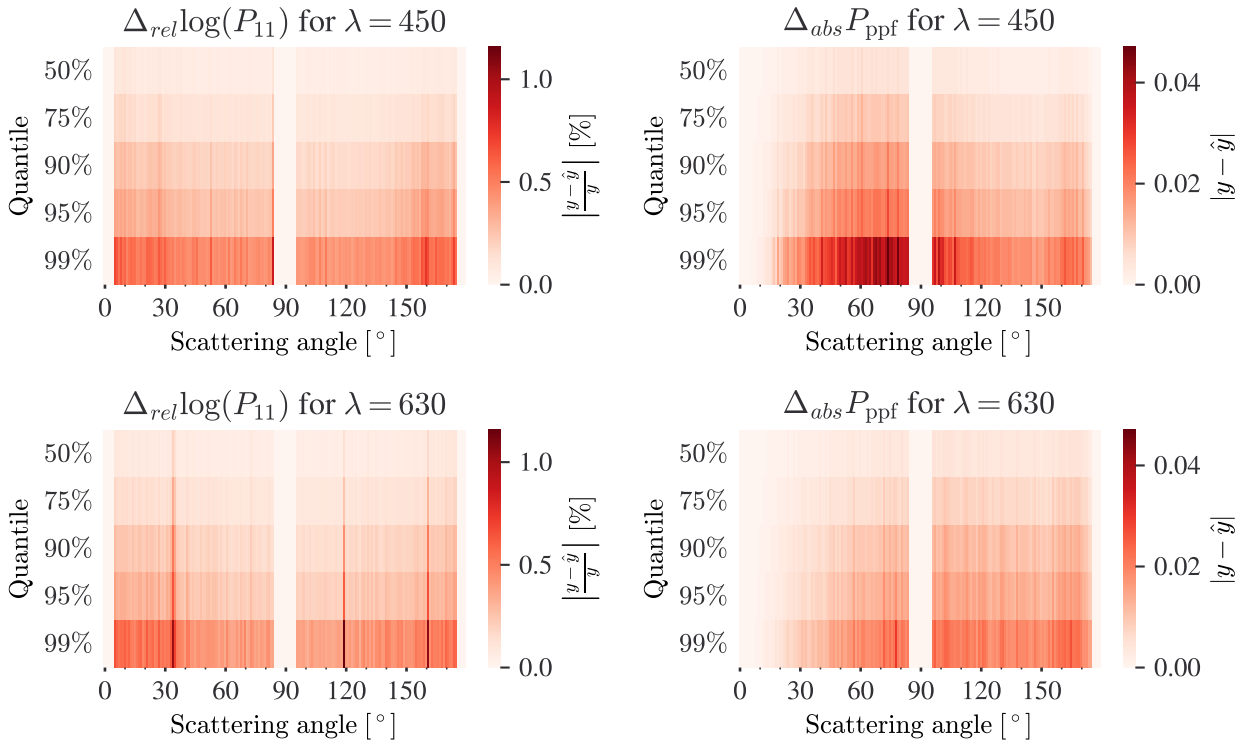

Figure 18: 50th, 75th, 90th, 95th and 99th percentiles of the relative error of the predicted logarithm of the scattering phase function, $\log(P_{11})$, and of the absolute error of the predicted polarized phase function, P_{ppf}, as a function of the scattering angle.

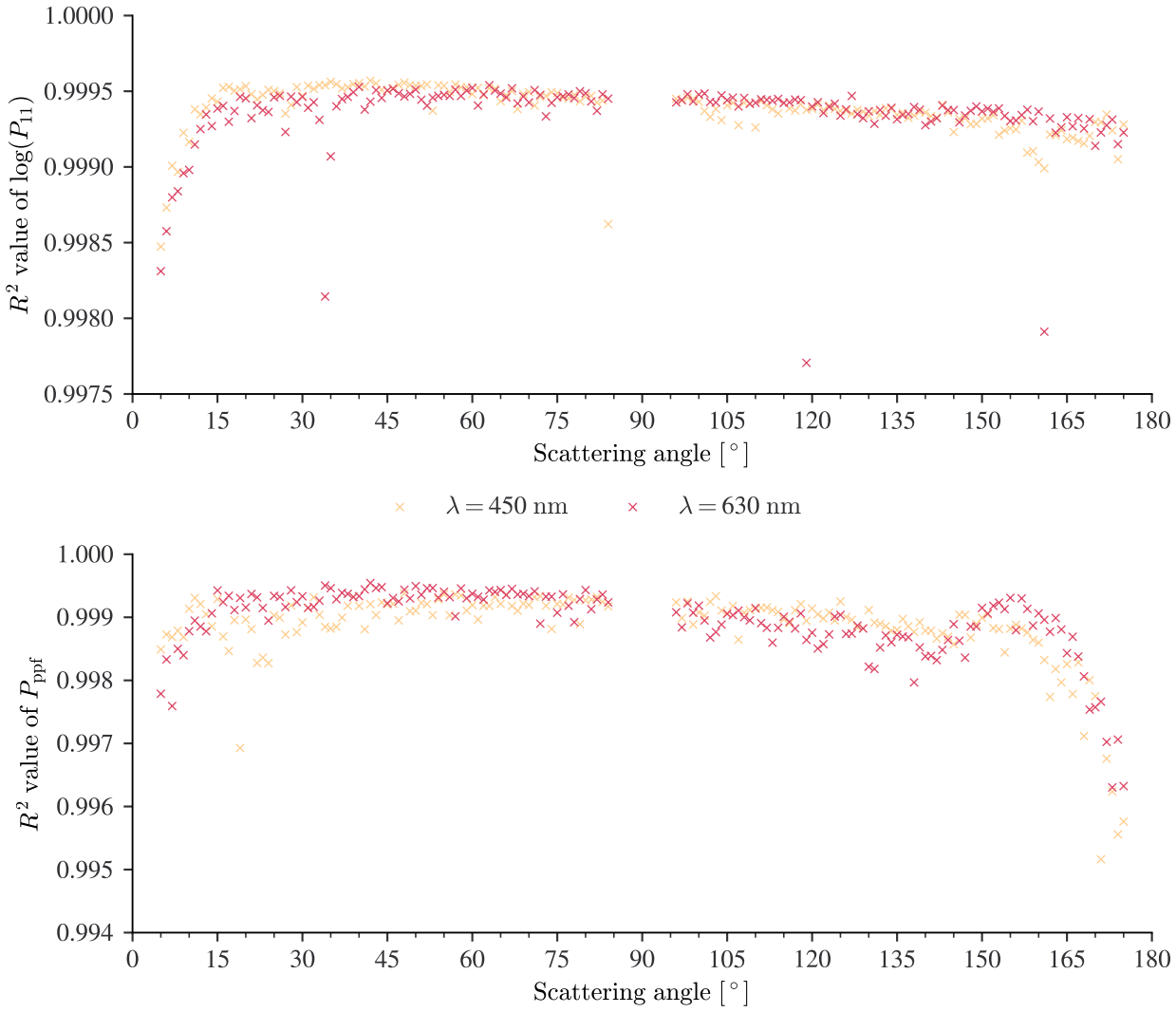

Figure 19: Coefficient of determination, R^2, once for $\log(P_{11})$ and once for P_{ppf} plotted against the scattering angles.

5.2.3 Generalization of the inverse process

To quantify the inaccuracy of the inverse process, Table 9 contains the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties. As before, an alternative representation is given in Figure 20, where the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties are normalized by the respective range from which the parameters were sampled. According to this representation, the errors of the standard deviation σ of the log-normal function are significantly larger than for the other features. Also the R^2_{adj} value of σ is smaller than the R^2_{adj} values of the other microphysical particle properties given in Table 10.

Table 9: 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties rounded to three decimal places.

Percentiles of Δ_{abs}	n_{450}	n_{630}	k_{450}	k_{630}	V	R	σ
50%	0.001	0.001	0.001	0.000	15.841	0.006	0.001
75%	0.002	0.001	0.001	0.001	28.453	0.011	0.002
90%	0.002	0.002	0.002	0.001	44.397	0.016	0.003
95%	0.003	0.003	0.002	0.002	56.587	0.020	0.004
99%	0.005	0.004	0.003	0.005	90.102	0.032	0.008

50%	0.333 %	0.296 %	0.300 %	0.204 %	0.317 %	0.272 %	2.600 %
75%	0.556 %	0.519 %	0.550 %	0.409 %	0.569 %	0.477 %	4.400 %
90%	0.852 %	0.778 %	0.851 %	0.664 %	0.888 %	0.694 %	6.801 %
95%	1.111 %	1.000 %	1.051 %	0.971 %	1.132 %	0.851 %	8.801 %
99%	1.741 %	1.630 %	1.651 %	2.505 %	1.803 %	1.362 %	15.601 %
	n_{450}	n_{630}	k_{450}	k_{630}	V_{tot}	r_m	σ

Figure 20: The absolute error of the retrieved microphysical aerosol particle properties, normalized by the respective range from which the parameters were sampled, i.e. $\Delta_{abs}\hat{y}_i/(\max_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i - \min_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i)$.

Table 10: R^2_{adj} of the retrieved microphysical aerosol particle properties rounded to two decimal places.

n_{450}	n_{630}	k_{450}	k_{630}	V	R	σ
1.00	1.00	1.00	1.00	1.00	1.00	0.97

Figure 12 displays the microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. In this representation, samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity. The dispersion is the highest for the standard deviation σ of the unimodal log-normal size distribution function and the lowest for the median optical radius r_m.

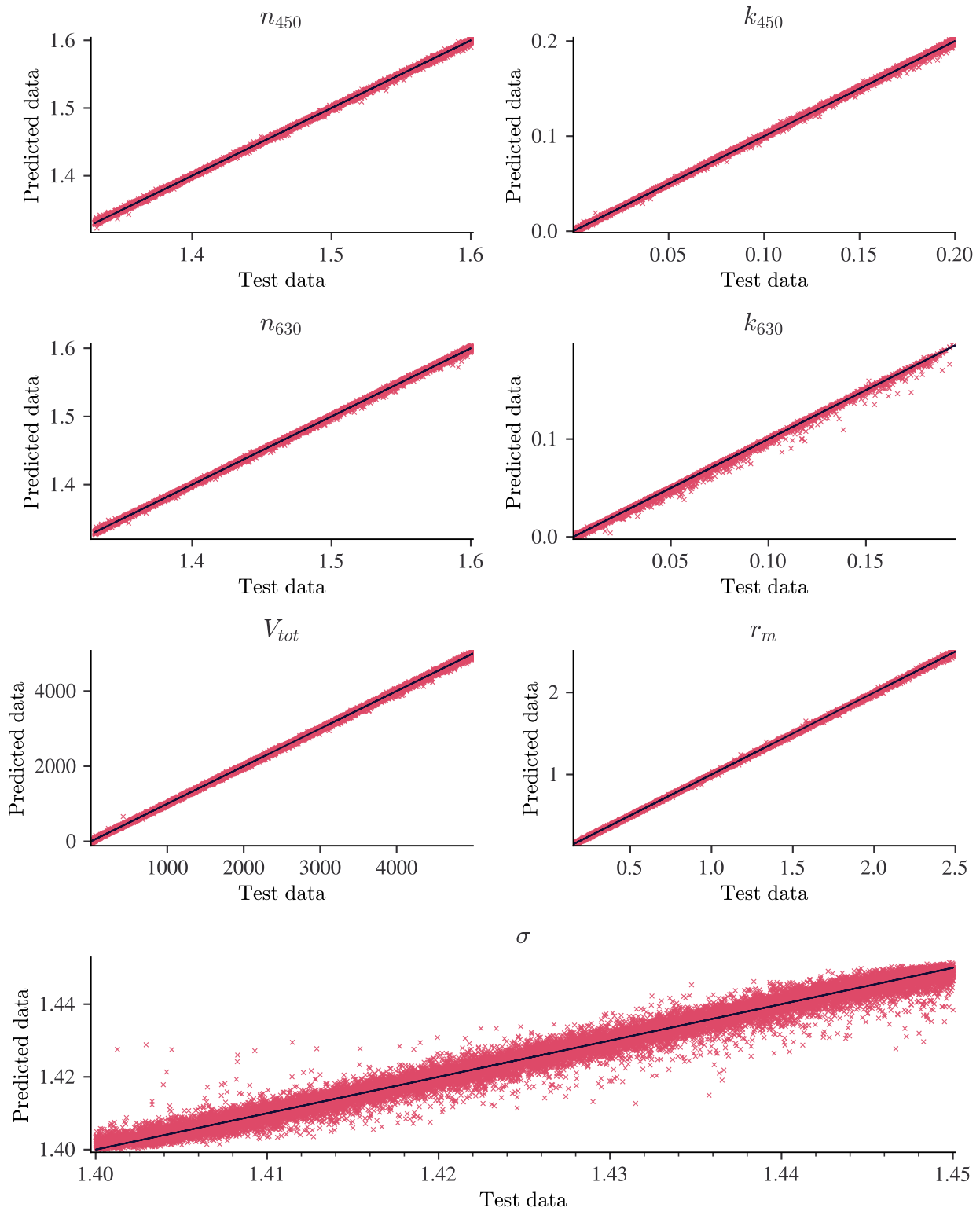

Figure 21: Microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. Samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity.

5.2.4 Qualitative comparison of predictions with test data

Lastly we are going to qualitatively consider the predictions made by our best model. Figure 22 shows five predictions of the forward process and their true counterparts from the test set. Although the predicted curves follow the general trends of the true curves from the test data, they contain high frequency variations that are absent in all of the depicted curves from the test set.

Figure 23 displays Five predictions of the log-normal size distribution delivered by the inverse process and their true counterparts from the test set. As in Figure 14, the prediction of the median optical radius r_m, the standard deviation σ and the total volume concentration V_{tot} seem to fit for all the depicted cases but for the the curve with the highest overall concentration, for which our prediction still seems slightly off with respect to r_m, σ and V_{tot}, albeit less than the model trained on data from one wavelength only in σ and V_{tot}.

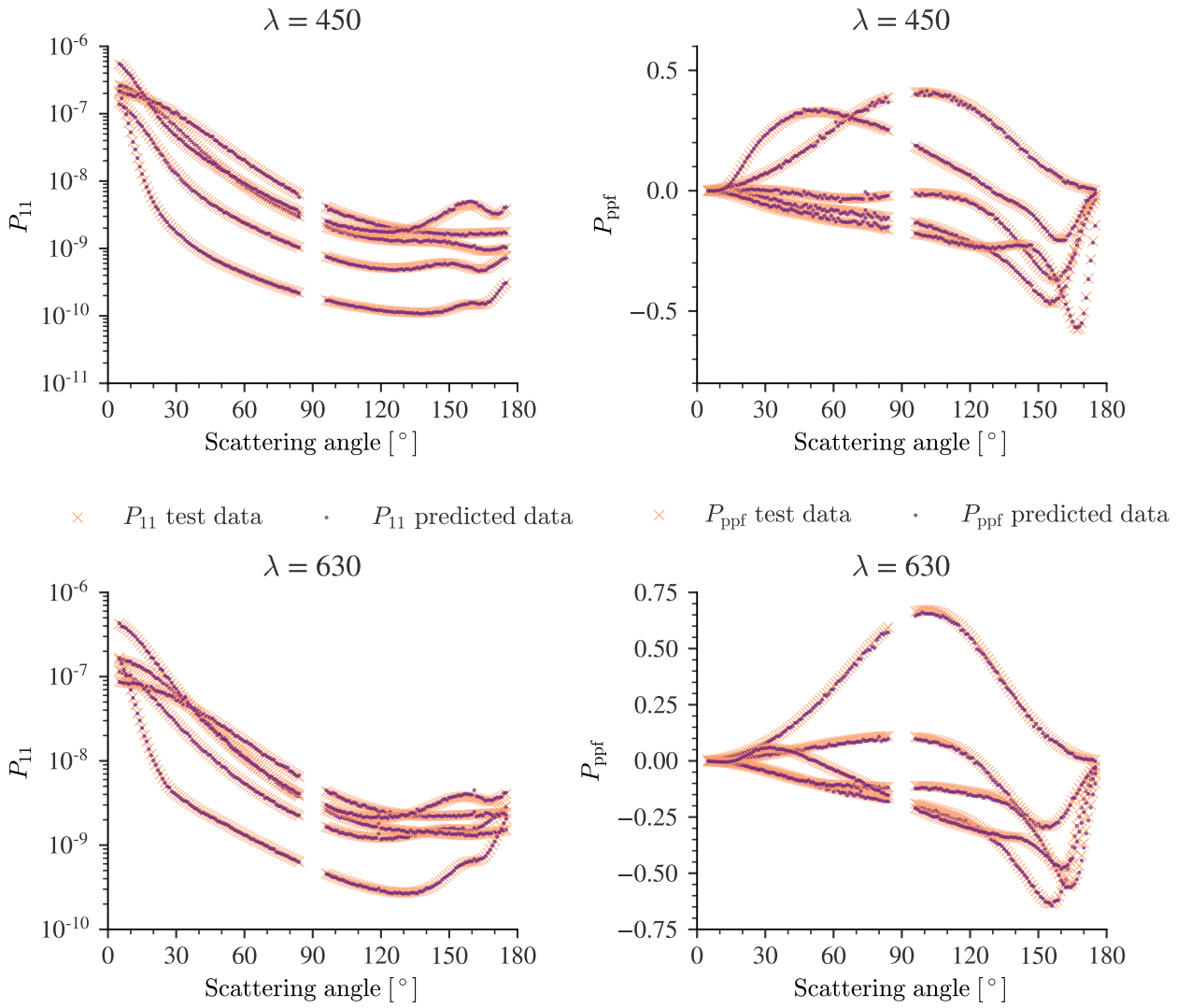

Figure 22: Qualitative comparison between curves that we predicted from sets of microphysical aerosol parameters from the test set and the respective curves simulated by GRASP-OPEN.

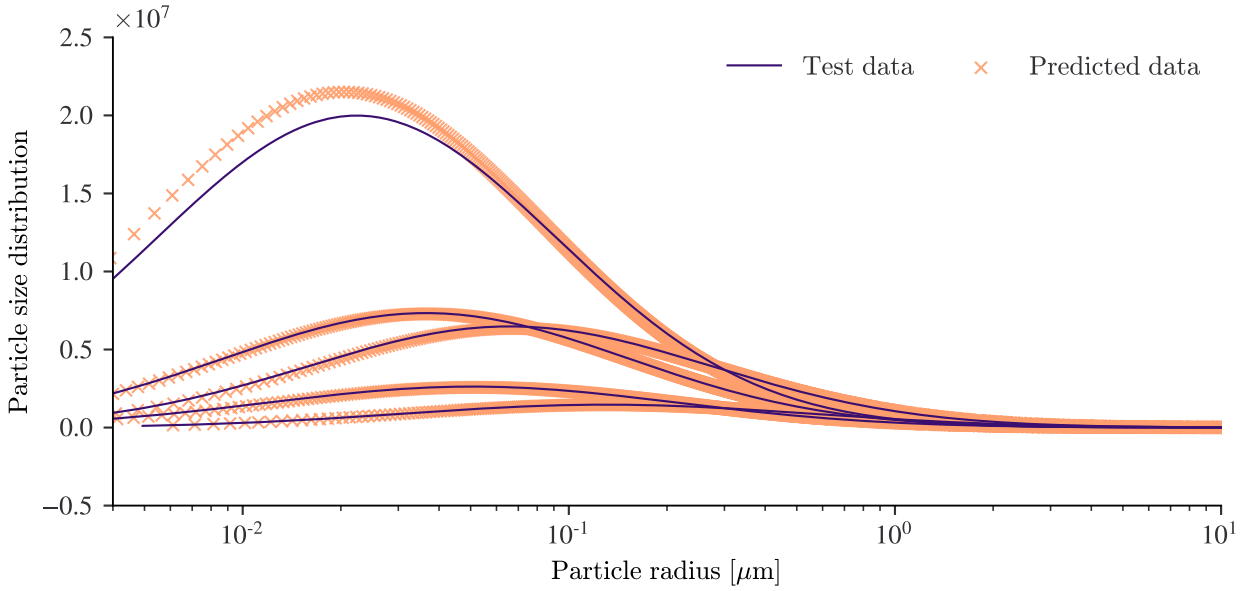

Figure 23: Qualitative comparison between the log-normal size distributions retrieved using our network when applied to five samples of light scattering data of the test set to the true log-normal size distributions that were used to simulate the respective samples of light scattering data.

5.3 Model trained on three wavelengths

The following section discusses models trained on data of all three wavelengths $\Lambda = \{450\,\text{nm}, 532\,\text{nm}, 630\,\text{nm}\}$.

The best of the 14 models trained within the scope of this hyperparameter scan was chosen according to the highest coefficient of determination, R^2, scored on the validation data set for the inverse process at the last epoch of training. Figure 24 shows the R^2 value of the validation data plotted against the R^2 value of the training data for each model at the end of training. This representation suggests that the forward process is slightly overfitting, while the inverse processes would be expected to generalize perfectly. It should further be noted that this hyperparameter scan delivered about 4 models that would potentially be more accurate surrogates for the forward process than the one chosen following our simple selection criterion.

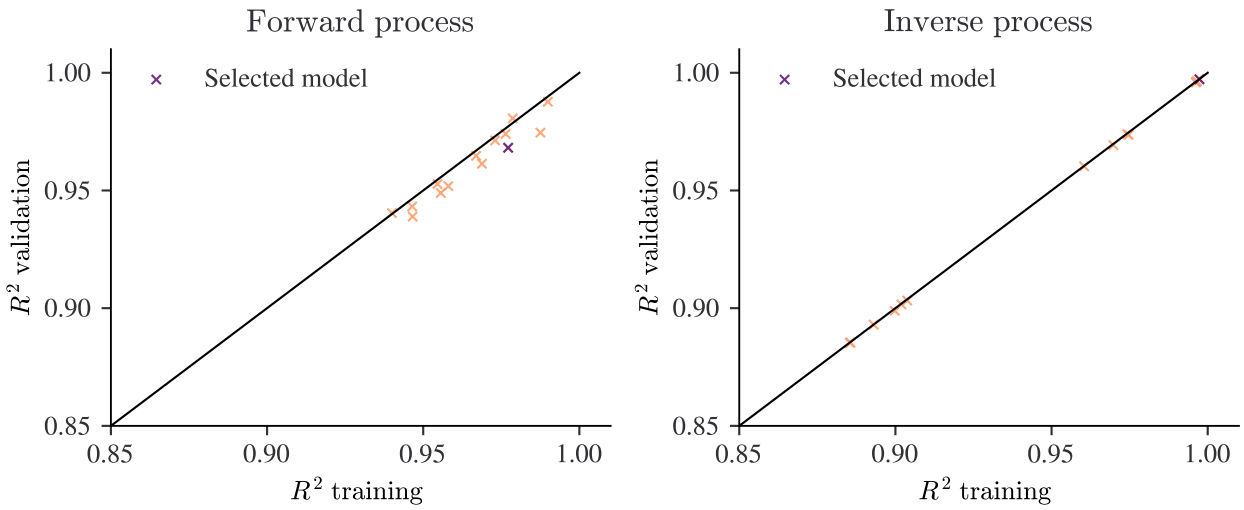

Figure 24: Results of the hyperparameter scan. The coefficient of determination, R^2, of the validation data is plotted against the R^2 value of the training data for each model at the end of training.

Table 11: Hyperparameters of the selected model.

Blocks	Depth	Width	Nominal dimension	w_p	w_r	w_x	w_y	w_z	σ_r
4	2	93	964	18	323	142	350	339	0.171

The hyperparameters of the selected model are given in Table 11. At the last epoch of training, the coefficients of determination of the forward process of this model scored $R^2_{train} = 0.977$ on the training data and $R^2_{val} = 0.968$ on the validation data. The coefficients of determination of the inverse process scored $R^2_{train} = 0.997$ on the training data and $R^2_{val} = 0.997$ on the validation. The average prediction time of the forward process of the selected model is 0.624 ms. The average prediction time of the inverse process 30.7 ms.

5.3.1 Training history

Figure 25 depicts the history of the coefficient of determination, R^2, of the forward and the inverse process. The maximum of R^2 over the 70 training epochs of the forward process is reached around epoch 55, the overall trend is ambiguous. Contrarily, the R^2 value of the inverse process is on the rise up until the last epoch. Figure 26 shows the evolution of the mean absolute error of both the forward and the inverse processes over 70 training epochs.

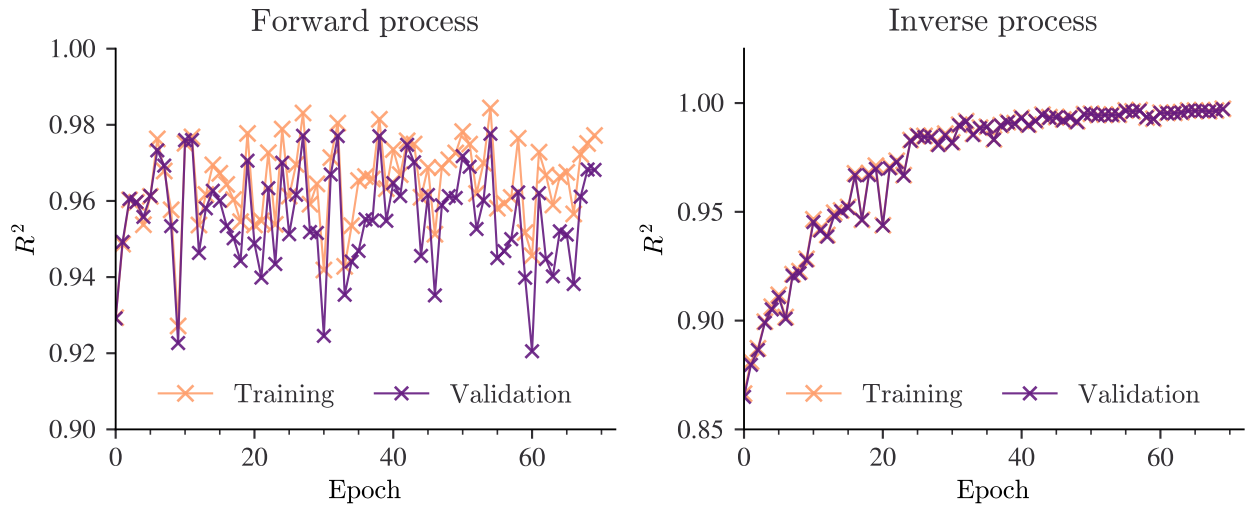

Figure 25: History of the coefficient of determination, R^2, of the forward and the inverse process.

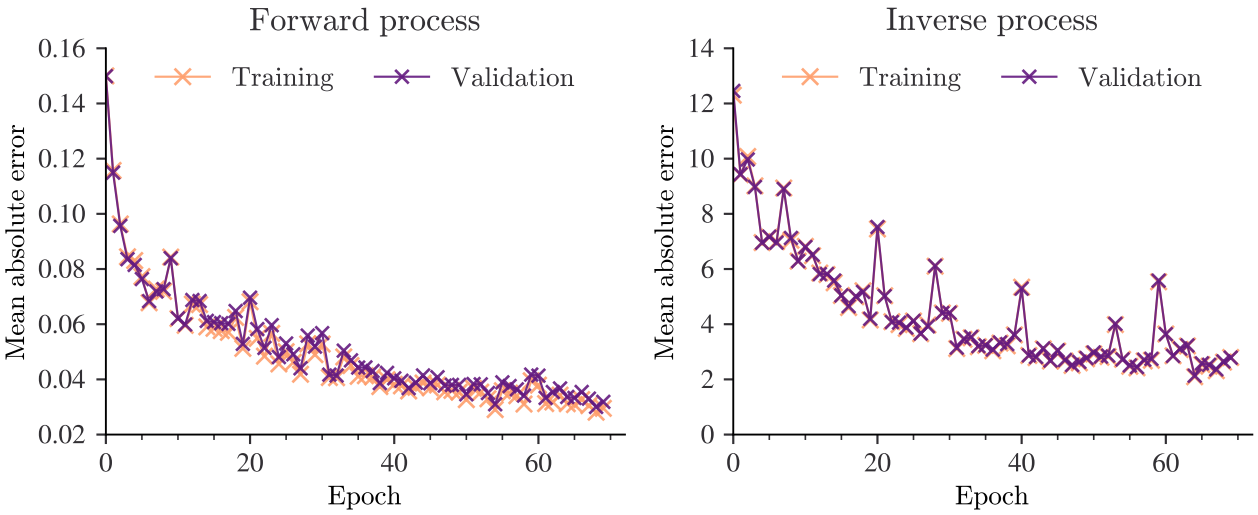

Figure 26: Evolution of the mean absolute error of both the forward and the inverse process over 80 training epochs.

5.3.2 Generalization of the forward process

The most compact assessment of the quality of the forward process is given in Table 12, which contains the maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel} \log(P_{11}(\theta))$, and of the absolute errors of the polarized phase function, $\Delta_{abs} P_{\text{ppf}}(\theta)$. A less compact but more nuanced representation of the accuracy is then given in Figure 27, which shows the 50th, 75th, 90th, 95th and 99th percentiles of $\Delta_{rel} \log(P_{11}(\theta))$ and $\Delta_{abs} P_{\text{ppf}}(\theta)$ as a function of the scattering angle. Note that besides broader trends in the errors across the angles, there are individual angles that significantly differ in accuracy from their respective neighbourhoods. Overall, the predictions of P_{ppf} seem to be more accurate at the wavelengths of $\lambda = 532$ and $\lambda = 630$ compared to the wavelength of $\lambda = 450$, again, mainly due to the region of increased error between about 40° and 85° of P_{ppf} in the case of $\lambda = 450$.

Figure 28 depicts the coefficient of determination, R^2, once for the logarithm of the scattering phase function, $\log(P_{11})$, and once for the polarized phase function, P_{ppf}, plotted against the scattering angles. Here can be observed that the angles with the lowest R^2 are close to 5° in the case of $\log(P_{11})$ and close to 175° in the case of P_{ppf}. Further, we can observe that the R^2 values of P_{ppf} for $\lambda = 450$ lower than the R^2 values of P_{ppf} for both $\lambda = 532$ and $\lambda = 630$ between the angles of 95° to 175°. Note again that besides these broader trends across the angles, there are individual angles that significantly differ in accuracy from their respective neighbourhoods. The overall variation in R^2 is, however, merely of the order of 10^{-3}.

Table 12: The maximum over all scattering angles of the 50th, 75th, 90th, 95th and 99th percentiles of the relative errors of the logarithm of the scattering phase function, $\Delta_{rel} \log(P_{11}(\theta))$, on one hand and of the absolute errors of the polarized phase function, $\Delta_{abs} P_{\text{ppf}}(\theta)$, on the other. Both are rounded to three decimal places.

$\max_{\theta \in \Theta}$ percentiles	$\Delta_{rel} \log(P_{11}(\theta))$ [%]	$\Delta_{abs} P_{\text{ppf}}(\theta)$
50%	0.274	0.012
75%	0.482	0.022
90%	0.778	0.031
95%	1.078	0.038
99%	2.007	0.055

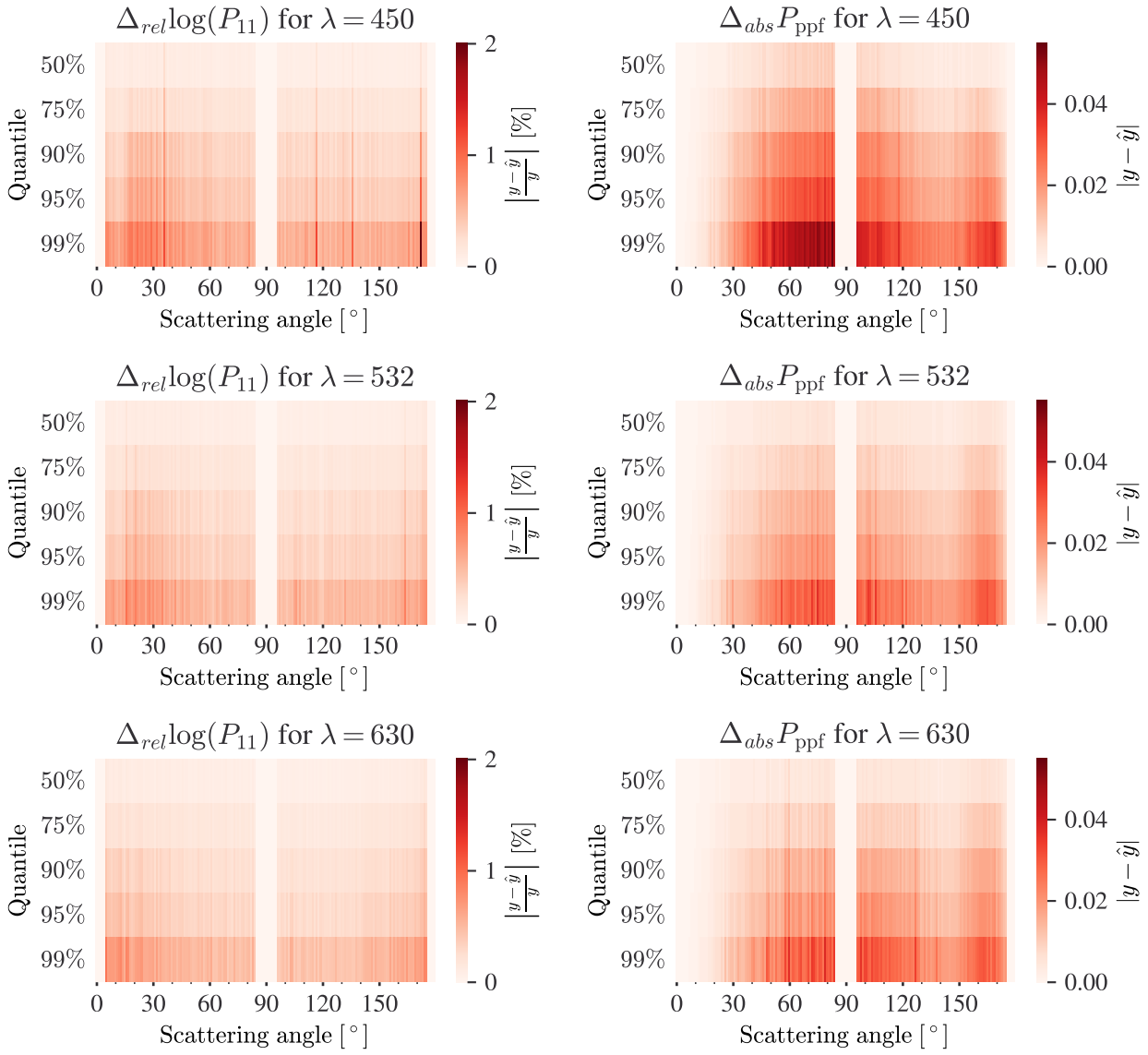

Figure 27: 50th, 75th, 90th, 95th and 99th percentiles of the relative error of the predicted logarithm of the scattering phase function, $\log(P_{11})$, and of the absolute error of the predicted polarized phase function, P_{ppf}, as a function of the scattering angle.

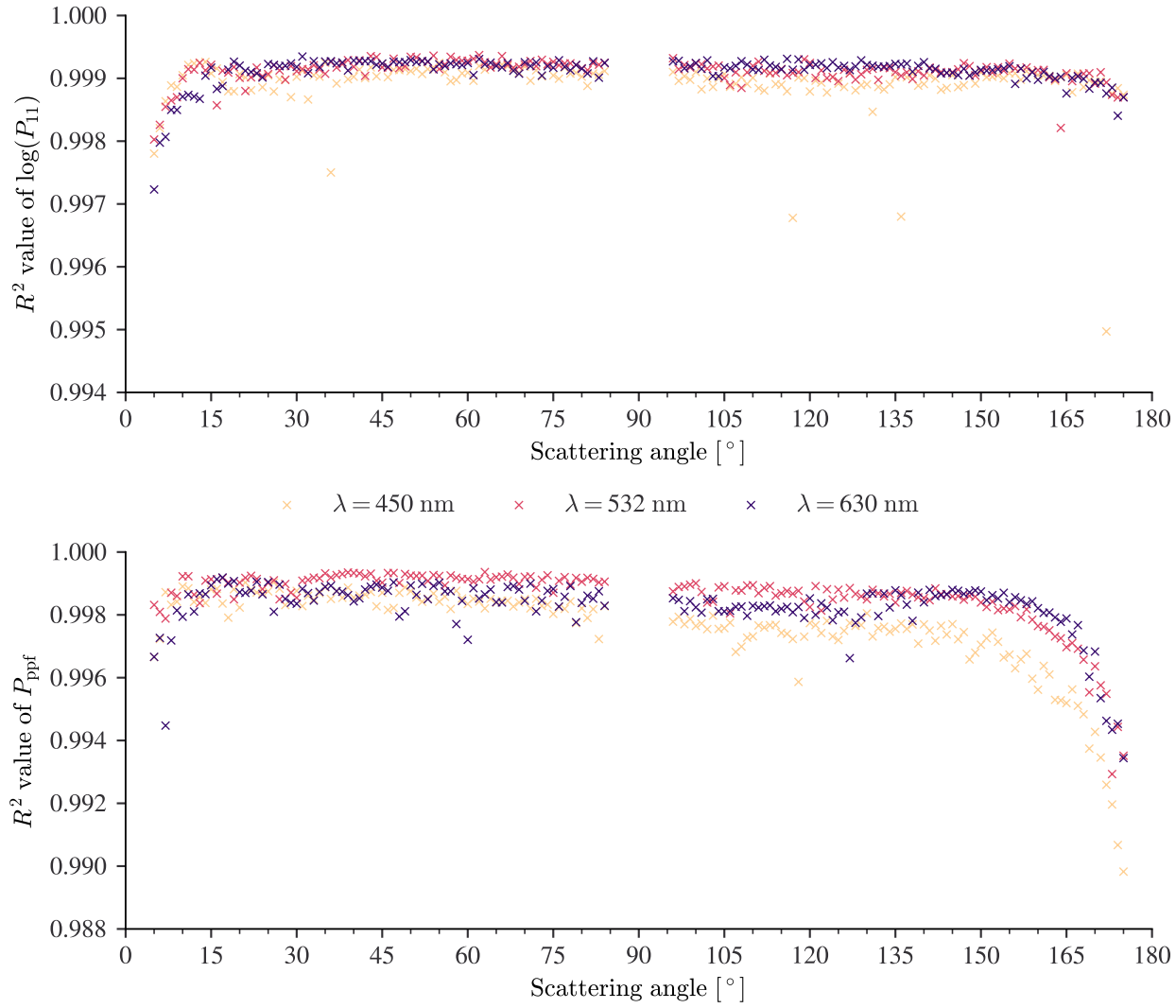

Figure 28: Coefficient of determination, R^2, once for $\log(P_{11})$ and once for P_{ppf} plotted against the scattering angles.

5.3.3 Generalization of the inverse process

To quantify the inaccuracy of the inverse process, Table 13 contains the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties. As before, an alternative representation is given in Figure 29, where the 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties are normalized by the respective range from which the parameters were sampled. According to this representation, the errors of the standard deviation σ of the log-normal function are significantly larger than for the other features. Also the R^2_{adj} value of σ is smaller than the R^2_{adj} values of the other microphysical particle properties given in Table 14.

Table 13: 50th, 75th, 90th, 95th and 99th percentiles of the absolute errors of the retrieved microphysical aerosol particle properties rounded to three decimal places.

Percentiles of Δ_{abs}	n_{450}	n_{532}	n_{630}	k_{450}	k_{532}	k_{630}	V_{tot}	r_m	σ
50%	0.001	0.001	0.001	0.001	0.001	0.000	20.526	0.005	0.001
75%	0.002	0.001	0.002	0.001	0.001	0.001	35.453	0.008	0.002
90%	0.002	0.002	0.002	0.002	0.002	0.001	53.037	0.012	0.003
95%	0.003	0.003	0.003	0.002	0.002	0.002	65.358	0.015	0.004
99%	0.004	0.004	0.004	0.004	0.003	0.004	95.219	0.022	0.007

	n_{450}	n_{532}	n_{630}	k_{450}	k_{532}	k_{630}	V_{tot}	r_m	σ
50%	0.296 %	0.296 %	0.333 %	0.250 %	0.405 %	0.204 %	0.411 %	0.200 %	2.600 %
75%	0.556 %	0.519 %	0.556 %	0.500 %	0.658 %	0.409 %	0.709 %	0.349 %	4.400 %
90%	0.815 %	0.815 %	0.852 %	0.750 %	0.911 %	0.613 %	1.061 %	0.519 %	6.401 %
95%	1.037 %	1.037 %	1.074 %	1.001 %	1.063 %	0.818 %	1.308 %	0.647 %	8.201 %
99%	1.630 %	1.556 %	1.667 %	1.751 %	1.619 %	1.789 %	1.905 %	0.941 %	13.401 %

Figure 29: The absolute error of the retrieved microphysical aerosol particle properties, normalized by the respective range from which the parameters were sampled, i.e. $\Delta_{abs}\hat{y}_i/(\max_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i - \min_{\tilde{y}\in\boldsymbol{Y}_D}\tilde{y}_i)$.

Table 14: R^2_{adj} of the retrieved microphysical aerosol particle properties rounded to two decimal places.

n_{450}	n_{532}	n_{630}	k_{450}	k_{532}	k_{630}	V_{tot}	r_m	σ
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98

Figure 30 displays the microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. In this representation, samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity. As in Figure 12 and 21, The dispersion is the highest for the standard deviation σ of the unimodal log-normal size distribution function and the lowest for the median optical radius r_m.

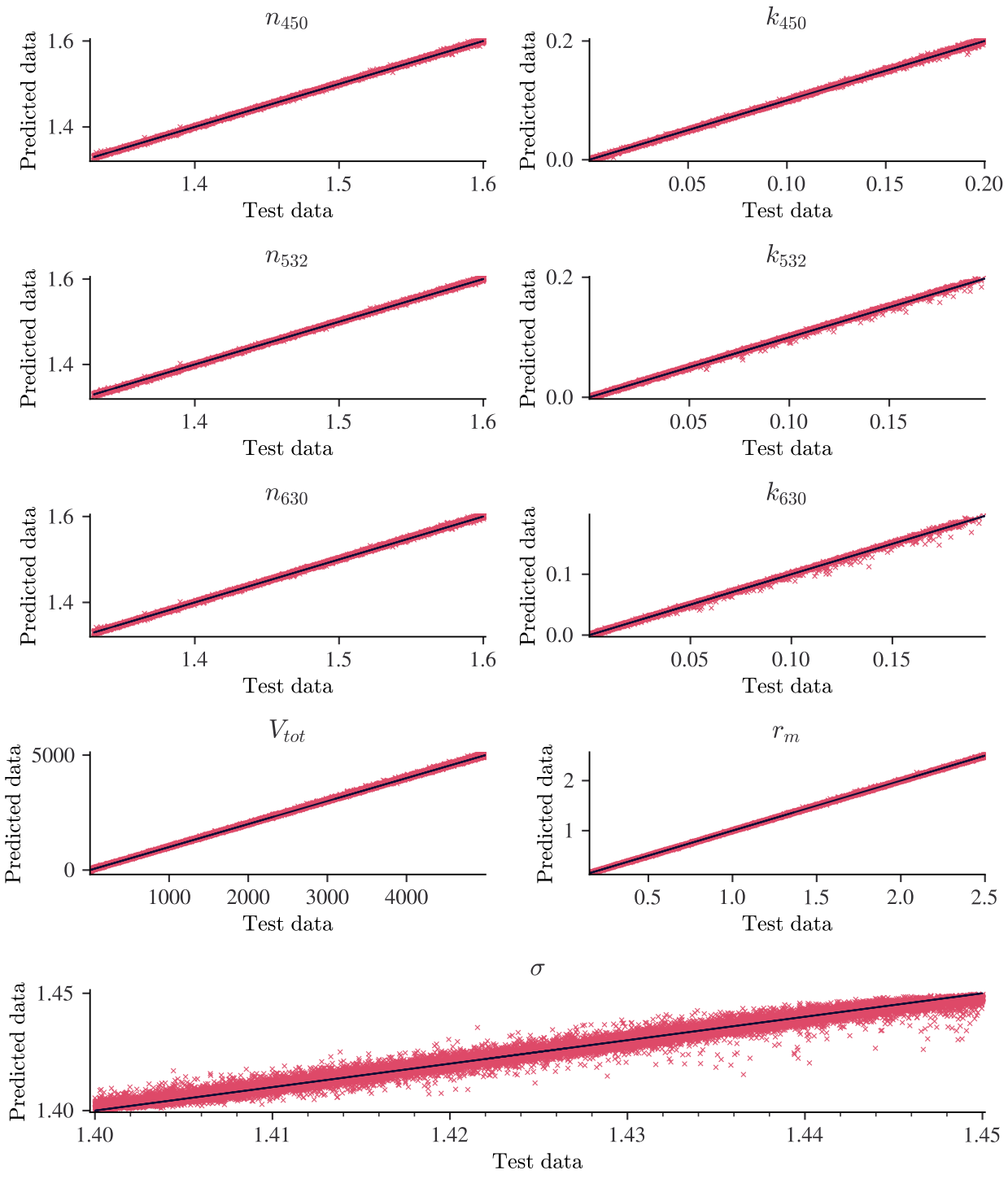

Figure 30: Microphysical aerosol properties retrieved from scattering data plotted against their true counterparts from the test data set. Samples come to lie on the black line on the diagonal if the predicted data is equal to the true value, i.e. their ratio unity.

5.3.4 Qualitative comparison of predictions with test data

Lastly we are going to qualitatively consider the predictions made by our best model. Figure 32 shows five predictions of the forward process and their true counterparts from the test set. Although the predicted curves follow the general trends of the true curves from the test data, they contain high frequency variations that are absent in all of the depicted curves from the test set. Figure 31 displays Five predictions of the log-normal size distribution delivered by the inverse process and their true counterparts from the test set. Contrary to Figure 14 and Figure 23, the prediction of the median optical radius r_m, the standard deviation σ and the total volume concentration V_{tot} seem to fit for all the depicted cases.

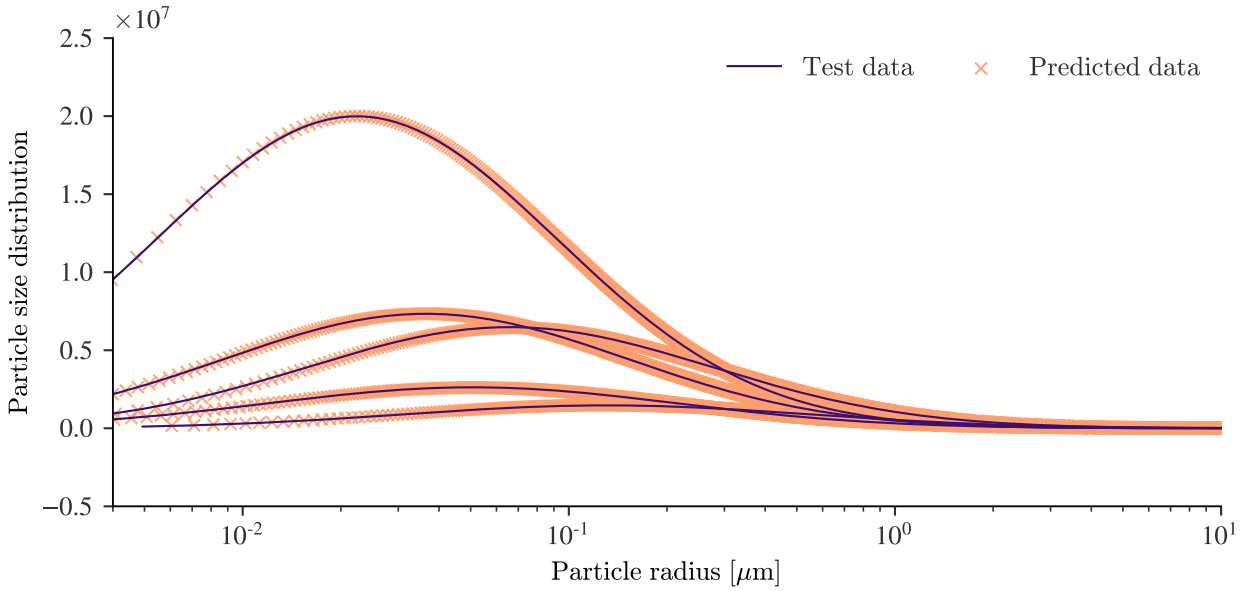

Figure 31: Qualitative comparison between the log-normal size distributions retrieved using our network when applied to five samples of light scattering data of the test set to the true log-normal size distributions that were used to simulate the respective samples of light scattering data.

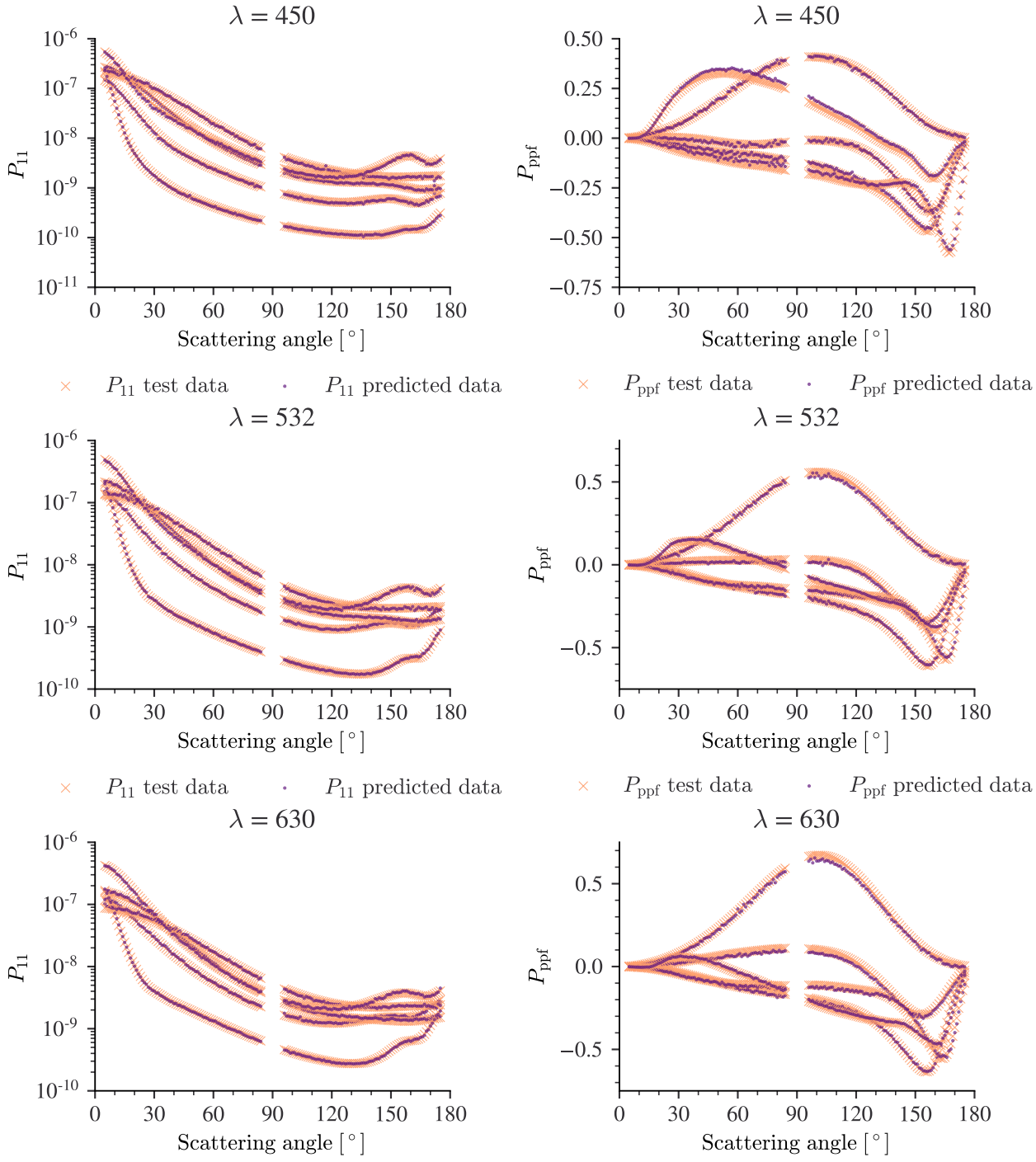

Figure 32: Qualitative comparison between curves that we predicted from sets of microphysical aerosol parameters from the test set and the respective curves simulated by GRASP-OPEN.

6 Discussion and outlook

The fluctuations in the R^2 value observable in Figure 7, 16 and 25 suggest that future work is needed to determine the criteria according to which one chooses the epoch at which the invertible neural network is considered best. It could also occur that the forward process commonly starts overfitting while the inverse process is still underfitting. Unfortunately, it was not possible for me to adequately investigate this aspect within the scope of this report due to time constraints.

In the prediction of P_{11} in particular, we sometimes encounter individual angles that significantly differ in accuracy to their respective surroundings, as can be observed in the figures depicting the relative quantile error of $\log(P_{11})$ and the absolute quantile error of P_{ppf} but also those depicting the R^2 value per scattering angle for the forward process, i.e. Figure 9, 18, 27 and Figure 10, 19, 28 respectively. If we compare these figures among each other, we note that a subset of these angles vary, suggesting the possibility that these type of errors could be reduced by the use of an ensemble method, e.g. a weighted average per angle of the predictions of several different models.

Another approach to increase robustness and accuracy of both the forward and the inverse process might be to integrate the fact that most both P_{11} and P_{ppf} seem to be very smooth, as can be seen when comparing our predictions to the test data in Figure 13, 22 and 32. Contrary to the test data, our predictions contain high variations at different sections. This is in itself unsurprising, for our invertible neural network simply regards them as separate data points with no constraints on the differences in values between neighbouring input values. However, this could be changed by deploying a pre and post processing involving spline interpolation, convolution based, or Fast Fourier transform based approach to filter out high frequency oscillations, be they artificial or measurement noise, which might help increasing the overall accuracy. I also was unable to investigate this due to time constraints, but the idea seems natural and promising.

Considering the results in Figure 13, 22 and 32, one might also contemplate the possibility of estimating the missing values of P_{11} and P_{ppf} between 85° to 95° degree as well as beyond the truncation angles at 5° and 175°, given the rest of the curve. Contrary to the parts of the curves between the angles of 85° to 95°, for which a linear extension of the curves might already do a good job at approximating the missing values, it seems that one would need a more intricate approach for completing the curves beyond the truncation angles of 5° and 175°, where higher order derivatives of the curves are likely not negligible. Being able to accurately complete P_{11} and P_{ppf} for these angles is desirable, for it might path the way to simpler measurement devices and more efficient measurement procedures. Bearing in mind the need for continuous and efficient monitoring of aerosols in various areas, these prospects are very tempting. If one were interested in completing curves of variable input length, then the logical form of this task would become equivalent to the one posed for time series forecasting in statistical inference. Solving the task might constitute an interesting application of recurrent neural networks which are often used in the latter area of research.

For the inverse process, the standard deviation, σ, of the size distribution function has the biggest prediction error across all models, as can be seen in Figure 11, 20 and 29. Comparing these plots among each other, we can further note that the accuracy of retrieval is always higher for the model using more scattering data, i.e. scattering data originating from more wavelengths. Also as expected, the difference in accuracy is smaller between the models using $\Lambda = \{450\,\text{nm}, 532\,\text{nm}, 630\,\text{nm}\}$ and $\Lambda = \{450\,\text{nm}, 630\,\text{nm}\}$ respectively than between the latter and $\Lambda = \{532\,\text{nm}\}$.

As already hinted at in Sec. 3, it can not be expected that the models trained on our current data set perform well on real data, unless for cases in which the assumption holds that $n_{450} \approx n_{532} \approx n_{630}$ and the measured aerosols are within the range of our simulated training set. Nonetheless, the quality of our predictions and retrievals suggest that given more realistic training data, this approach seems practical, fulfilling the promise of constituting a fast and accurate surrogate alternative for GRASP-OPEN.

Our work shows that invertible neural networks are a promising tool for both simulating the scattering phase function P_{11} and the polarized phase function P_{ppf} as well as retrieving microphysical aerosol properties from such light scattering data. After the initial training phase, their evaluation is cheap, fast and accurate, a combination that was currently not reachable using classical computer algorithms.

If the accuracy of the invertible model were insufficient for a given task, it might still be possible to use it to initialise more precise but also more expensive physically-based iterative algorithms, such as GRASP-OPEN, and thus diminish the time it takes to reach a given accuracy.

References

[1] L. Ardizzone, J. Kruse, S. Wirkert, D. Rahner, E. W. Pellegrini, R. S. Klessen, L. Maier-Hein, C. Rother, and U. Köthe. Analyzing inverse problems with invertible neural networks. *arXiv preprint arXiv:1808.04730*, Aug. 2018. URL http://arxiv.org/pdf/1808.04730v3.

[2] R. Bellotti. Accelerating accelerators: Fast surrogate models for beam prediction. Master's thesis, ETH Zürich, June 2020.

[3] V. V. Berdnik and V. A. Loiko. Retrieval of size and refractive index of spherical particles by multiangle light scattering: neural network method application. *Applied optics*, 48:6178–6187, 2009. ISSN 0003-6935. doi: 10.1364/ao.48.006178.

[4] V. V. Berdnik and V. A. Loiko. Neural networks for particle parameter retrieval by multi-angle light scattering. In *Light Scattering Reviews 10*, pages 291–340. Springer, 2016. doi: 10.1007/978-3-662-46762-6_7.

[5] C. Chen, C. Chen, O. Dubovik, D. Fuertes, P. Litvinov, T. Lapyonok, A. Lopatin, F. Ducos, Y. Derimian, M. Herman, D. Tanré, L. A. Remer, A. Lyapustin, A. M. Sayer, A. M. Sayer, R. C. Levy, N. C. Hsu, J. Descloitres, L. Li, B. Torres, Y. Karol, M. Herrera, M. Herreras, M. Aspetsberger, M. Wanzenboeck, L. Bindreiter, D. Marth, A. Hangler, and C. Federspiel. Validation of grasp algorithm product from polder/parasol data and assessment of multi-angular polarimetry potential for aerosol monitoring. *Earth System Science Data*, 12:3573–3620, 12 2020. ISSN 1866-3516. doi: 10.5194/essd-12-3573-2020. URL https://essd.copernicus.org/articles/12/3573/2020/essd-12-3573-2020.pdf.

[6] G. Dolgos and J. V. Martins. Polarized imaging nephelometer for in situ airborne measurements of aerosol light scattering. *Optics express*, 22(18):21972–21990, 2014. ISSN 1094-4087. doi: 10.1364/oe.22.021972.

[7] B. N. Holben, T. F. Eck, I. Slutsker, D. Tanré, J. P. Buis, A. Setzer, E. Vermote, J. A. Reagan, Y. J. Kaufman, T. Nakajima, F. Lavenu, I. Jankowiak, and A. Smirnov. Aeronet–a federated instrument network and data archive for aerosol characterization. *Remote sensing of environment*, 66(1):1–16, 1998.

[8] P. G. Hull and M. S. Quinby-Hunt. Neural network to extract size parameter from light scattering data, 1997.

[9] A. Ishimaru, S. Kitamura, R. J. Marks, L. Tsang, C. M. Lam, and D. C. Park. Particle-size distribution determination using optical sensing and neural networks. *Optics letters*, 15(21):1221–1223, 1990. ISSN 0146-9592. doi: 10.1364/ol.15.001221.

[10] W. John. Size distribution characteristics of aerosols. *Aerosol Measurement: Principles, Techniques, and Applications, 3rd Edn., edited by: Kulkarni, P., Baron, PA, and Willeke, K., John Wiley and Sons, New York*, pages 41–54, 2011. doi: 10.1002/9781118001684.ch4.

[11] Y. J. Kaufman, D. Tanré, and O. Boucher. A satellite view of aerosols in the climate system. *Nature*, 419(6903):215–223, Sept. 2002. doi: 10.1038/nature01091. URL https://ui.adsabs.harvard.edu/abs/2002Natur.419..215K.

[12] A. Kratsios. The universal approximation property. *Annals of Mathematics and Artificial Intelligence, 2020*, Oct. 2019. doi: 10.1007/s10472-020-09723-1.

[13] L. Li, K. Jamieson, A. Rostamizadeh, E. Gonina, M. Hardt, B. Recht, and A. Talwalkar. A system for massively parallel hyperparameter tuning. *Conference on Machine Learning and Systems 2020*, Oct. 2018.

[14] R. L. Modini, A. Moallemi, and M. B. Gysel. personal communication, Mar. 2021.

[15] G. Myhre, C. Myhre, B. Samset, and T. Storelvmo. Aerosols and their relation to global climate and climate sensitivity. *Nature Education Knowledge*, 4(5):7, 2013.

[16] C. A. O. Nascimento, R. Guardani, and M. Giulietti. Use of neural networks in the analysis of particle size distributions by laser diffraction. *Powder Technology*, 90(1):89–94, 1997. ISSN 0032-5910. doi: 10.1016/s0032-5910(96)03192-0.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[18] A. B. Pluchino, S. S. Goldberg, J. M. Dowling, and C. M. Randall. Refractive-index measurements of single micron-sized carbon particles. *Applied optics*, 19(19):3370–3372, 1980. ISSN 0003-6935. doi: 10.1364/ao.19.003370.

[19] S. Self, J.-X. Zhao, R. E. Holasek, R. C. Torres, and A. J. King. *The Atmospheric Impact of the 1991 Mount Pinatubo*, pages 1089–1115. Univ. Washington Press, 1993.

[20] P. M. Shamjad, S. N. Tripathi, N. M. Thamban, and H. Vreeland. Refractive index and absorption attribution of highly absorbing brown carbon aerosols from an urban indian city-kanpur. *Scientific reports*, 6(1):1–7, 2016. ISSN 2045-2322. doi: 10.1038/srep37735.

[21] Z. Ulanowski, Z. Wang, P. H. Kaye, and I. K. Ludlow. Application of neural networks to the inverse light scattering problem for spheres. *Applied optics*, 37(18):4027, 1998. ISSN 0003-6935. doi: 10.1364/ao.37.004027.