



Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations

A. Di Noia¹, O. P. Hasekamp¹, G. van Harten², J. H. H. Rietjens¹, J. M. Smit¹, F. Snik², J. S. Henzing³, J. de Boer², C. U. Keller², and H. Volten⁴

¹SRON Netherlands Institute for Space Research, Sorbonnelaan 2, 3584 CA Utrecht, the Netherlands

²Leiden Observatory, Leiden University, Niels Bohrweg 2, 2333 CA Leiden, the Netherlands

³Netherlands Organisation for Applied Research (TNO), Princetonlaan 6, 3584 CB Utrecht, the Netherlands

⁴National Institute for Public Health and the Environment (RIVM), Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, the Netherlands

Correspondence to: A. Di Noia (a.di.noia@sron.nl)

Received: 11 July 2014 – Published in Atmos. Meas. Tech. Discuss.: 8 September 2014

Revised: 4 December 2014 – Accepted: 5 December 2014 – Published: 14 January 2015

Abstract. In this paper, the use of a neural network algorithm for the retrieval of the aerosol properties from ground-based spectropolarimetric measurements is discussed. The neural network is able to retrieve the aerosol properties with an accuracy that is almost comparable to that of an iterative retrieval. By using the outcome of the neural network as first guess in the iterative retrieval scheme, the accuracy of the retrieved fine- and coarse-mode optical thickness is further improved, while for the other parameters the improvement is small or absent. The resulting scheme (neural network + iterative retrieval) is compared to the original one (look-up table + iterative retrieval) on a set of simulated ground-based measurements, and on a small set of real observations carried out by an accurate ground-based spectropolarimeter. The results show that the use of a neural-network-based first guess leads to an increase in the number of converging retrievals, and possibly to more accurate estimates of the aerosol effective radius and complex refractive index.

1 Introduction

Atmospheric aerosols influence Earth's radiative balance and climate in multiple ways. First, through their scattering and absorption properties, they modify the amount of solar radiation reaching the Earth's surface and the lowest atmospheric layers (direct effect, Yu et al., 2006, and references therein). Second, acting as cloud condensation nuclei, they influence the cloud formation processes, thereby affecting

cloud microphysical properties and precipitations (indirect effect; Lohmann and Feichter, 2005, and references therein). Third, highly absorbing aerosols have a warming effect on the atmosphere, which can cause the evaporation of cloud particles and hence lead to a reduction in the cloud cover (semi-direct effect; Hansen et al., 1997; Kaufman and Koren, 2006).

The net effect of the atmospheric aerosols on climate is highly uncertain, both in magnitude and sign (Boucher et al., 2013). This uncertainty is partly due to the uncertainties in the parametrization of aerosol processes in models (Liu et al., 2012; Rosenfeld et al., 2013), but is also due to the intrinsic limitations in our current knowledge of the global distribution of the most important aerosol microphysical and optical properties (Mishchenko et al., 2004, 2005).

Further motivation for studying atmospheric aerosols is provided by their impact on human health (Anderson et al., 2012) and aviation safety (Alexander, 2013).

The sensitivity of polarized radiance measurements to the aerosol size distribution, shape and refractive index (Hansen and Travis, 1974) provides, in principle, a means for retrieving these aerosol properties from remote measurements either from space (Mishchenko and Travis, 1997; Hasekamp and Landgraf, 2007) or from the ground (Boesche et al., 2006).

A limited number of techniques for retrieving the aerosol properties using intensity and polarization of the scattered sunlight are reported in the literature. Some of these are based on look-up tables (LUTs), in that they select the

aerosol model that best fits the observed total and polarized intensities from a set of pre-calculated values (e.g. Deuzé et al., 2000, 2001). Other methods consist of variational retrievals, where full radiative transfer calculations are used to minimize a nonlinear cost function with respect to a set of aerosol microphysical and optical properties. In particular, Dubovik et al. (2011) describe a variational retrieval scheme where satellite intensity and polarization measurements are combined in order to retrieve a set of aerosol (total aerosol load, aerosol size distribution in a number of size bins, fraction of spherical particles, spectrally dependent complex refractive index and aerosol layer height) and surface reflection parameters, relying on the assumption that all the aerosol size bins have the same refractive index, and imposing a number of smoothness constraints on the size distribution and on the spatial and temporal variability of the aerosol properties. An alternative approach is described in Hasekamp et al. (2011), where a different set of aerosol parameters is retrieved using a scheme based on Phillips–Tikhonov regularization, assuming a bimodal log-normal aerosol size distribution and allowing for a fine and a coarse mode characterized by different refractive indices, which are assumed to be wavelength-independent. Again, a nonlinear cost function – depending on the discrepancy between observed and simulated intensities and degrees of linear polarization (DLPs) at multiple wavelengths and viewing angles – is iteratively minimized, starting from a first guess provided by a LUT. Further details are given in Sect. 2.

The choice of a LUT-based first guess is also common to other variational aerosol retrieval schemes (e.g. Waquet et al., 2009; Knobelspiesse et al., 2011). The accuracy of a LUT retrieval depends on the amount of entries (aerosol models, other atmospheric parameters, viewing angles, etc.) stored in the table. However, enhancing the accuracy of a LUT retrieval by increasing the LUT dimension is an expensive task, in terms of memory requirements (need for bigger arrays to be stored) and computation time (more models to be tried in order to select the one that best fits the measurements).

A possible alternative to a LUT for generating a first guess in a variational aerosol retrieval scheme is represented by a neural network (NN). Neural networks are powerful tools for the mapping of nonlinear relationships, and have been successfully applied as stand-alone retrieval algorithms to several remote sensing problems, including aerosol-related applications (Han et al., 2006; Niang et al., 2006; Radosavljevic et al., 2010; Picchiani et al., 2011). Nevertheless, the output of a NN retrieval scheme is also suitable for use as the initial step of an iterative retrieval method, giving rise to a “neuro-variational” retrieval scheme. A number of neuro-variational retrieval schemes, also applied to aerosol retrievals, are already reported in the literature. Jamet et al. (2005) and Brajard et al. (2006, 2012) used a set of NNs to build a forward model that simulates satellite-measured reflectances given a set of atmospheric and oceanic parameters, and is then inverted with respect to such parameters using an iterative

method. Diouf et al. (2013) further refined this scheme by using an unsupervised neural network to generate a first guess for the iterative retrieval.

In our experiments we have investigated a different type of neuro-variational retrieval algorithm. Specifically, we have trained a multilayer perceptron (MLP) neural network (Werbos, 1974) to retrieve a set of aerosol parameters from simulated intensities and DLPs at a number of wavelengths and angles, and we have used the set of aerosol parameters retrieved by the NN as first guess for the subsequent iterative scheme using a full-physics forward model and an inversion method based on Phillips–Tikhonov regularization. We will show that the use of a NN-based first guess instead of the LUT-based one can substantially improve the performance of the retrieval algorithm, with respect to the number of retrievals that reach an acceptable minimum of the cost function and with respect to the accuracy of the retrieved parameters. The resulting retrieval scheme is also less demanding in terms of memory allocation, and is faster than the scheme that uses a LUT-based first guess.

The paper is organized as follows. Section 2 gives an overview of the original retrieval algorithm, including details on the LUT. Section 3 describes the actual setup of the experiments performed in this paper. Section 4 describes the design of the NN that is then used as a replacement for the LUT. In Sect. 5 the algorithm version using a NN-based first guess is compared to that using the LUT-based first guess on a data set of simulated observations, and then on a number of measurements performed by a ground-based spectropolarimeter based on the spectral modulation technique (van Harten et al., 2014). Section 6 is dedicated to the conclusions and to the discussion of future work.

2 Variational aerosol retrieval algorithm

2.1 General setup

The retrieval scheme used in this work is a modified version of the algorithm discussed by Hasekamp et al. (2011) for PARASOL satellite data. The main modifications are the following: (i) the forward model is adapted to the simulation of ground-based measurements instead of satellite measurements, (ii) a different method is used in the choice of the regularization parameter at each iteration of the inversion (a heuristic procedure, described in the following subsection, is used instead of the L-curve method), and (iii) the state vector is defined in a different manner. In particular, with respect to the latter point, the retrieved aerosol quantities are the effective radius, the complex refractive index, and the aerosol column amount for the fine and the coarse mode.

2.2 Inversion method

According to the Phillips–Tikhonov regularization method, the retrieval cost function is defined as

$$J(\mathbf{x}) = [\mathbf{y} - \mathbf{F}(\mathbf{x})]^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x})] + \gamma (\mathbf{x} - \mathbf{x}_a)^T \mathbf{H} (\mathbf{x} - \mathbf{x}_a), \quad (1)$$

where \mathbf{x} is the state vector (i.e. the set of quantities to be retrieved), \mathbf{y} is the vector of measurements, \mathbf{F} is the vector-valued forward model function that maps the state vector into the measurement vector, \mathbf{x}_a is a priori for the state vector, \mathbf{S}_ϵ is the error covariance matrix, \mathbf{H} is the regularization matrix and γ is the regularization parameter. Since \mathbf{F} is nonlinear with respect to \mathbf{x} , the minimization of J is carried out iteratively, using the Gauss–Newton method. Starting from a first guess \mathbf{x}_0 for the state vector, a sequence of solutions $\{\mathbf{x}_i\}$ is generated by linearizing \mathbf{F} around \mathbf{x}_i at each step i , and then finding the next solution \mathbf{x}_{i+1} as the one that minimizes J for the linearized version of \mathbf{F} . The process is stopped when the goodness of fit no longer changes significantly or when a maximum number of iterations is reached. The linear approximation of \mathbf{F} around \mathbf{x}_i can be written as

$$\mathbf{F}(\mathbf{x}) \approx \mathbf{F}(\mathbf{x}_i) + \mathbf{K}_i (\mathbf{x} - \mathbf{x}_i), \quad (2)$$

where $\mathbf{K}_i = \partial \mathbf{F} / \partial \mathbf{x}$ evaluated for $\mathbf{x} = \mathbf{x}_i$. Depending on the adopted forward model, \mathbf{K}_i can be computed either analytically or by finite differencing. By replacing \mathbf{F} with Eq. (2) in Eq. (1), differentiating J with respect to \mathbf{x} and setting the result to zero, the new solution is found as (Rodgers, 2000)

$$\mathbf{x}_{i+1} = \mathbf{x}_a + (\mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} \mathbf{K}_i + \gamma_i \mathbf{H})^{-1} \mathbf{K}_i^T \mathbf{S}_\epsilon^{-1} [\mathbf{y} - \mathbf{F}(\mathbf{x}_i) + \mathbf{K}_i (\mathbf{x}_i - \mathbf{x}_a)]. \quad (3)$$

As indicated by the subscript i , the regularization parameter γ is also updated during the iterative process. While in Hasekamp et al. (2011) this is done using the L-curve method (Hansen and O’Leary, 1993), in this version of the algorithm the choice of γ_i is made heuristically by trying different values for γ_i and selecting the one that produces the solution with the best fit between the measurements and the simulations. The goodness of fit is measured using the parameter

$$\chi^2(\gamma_i) = \frac{1}{m} \sum_{j=1}^m \frac{[y_j - F_j(\mathbf{x}_{i+1}(\gamma_i))]^2}{\sigma_j^2}, \quad (4)$$

where m is the dimension of \mathbf{y} , y_j and F_j are the j th components of \mathbf{y} and \mathbf{F} respectively, and σ_j^2 is the j th element on the principal diagonal of \mathbf{S}_ϵ . In Eq. (4), the dependence of χ^2 and \mathbf{x}_{i+1} on the chosen γ_i has been made explicit. According to this formalism, the chosen γ_i is thus the one that minimizes χ^2 .

The regularization matrix \mathbf{H} adopted in our retrieval scheme is diagonal, and the elements on its principal diagonal can be written as

$$h_{ii} = \frac{1}{(x_i^{\max} - x_i^{\min})^2} \quad i = 1, \dots, n, \quad (5)$$

where n is the dimension of the state vector and x_i^{\max} and x_i^{\min} are respectively the maximum and minimum allowed values for the elements of the state vector. In the Phillips–Tikhonov inversion scheme the matrix \mathbf{H} determines the relative weight between the different parameters in the side constraint, whereas γ determines the overall contribution of the side constraint, for which a proper value is found using the procedure described above.

In the original version of the algorithm, the first guess \mathbf{x}_0 is generated by a LUT. The LUT stores the coefficients of the azimuthal Fourier expansion of the Stokes vectors simulated for 576 predefined aerosol models (i.e. combinations of the aerosol parameters) at a number of wavelengths and viewing angles, stratified by solar zenith angle, surface pressure and surface albedo. The reflectances and the DLPs are computed by recombining the aforementioned coefficients for each relative azimuth angle. For all the predefined aerosol models in the LUT, a retrieval of the fine- and coarse-mode aerosol optical thickness (AOT) is performed using the LUT as forward model, and the aerosol model that provides the best fit to the measurements is selected as first guess.

The values of the aerosol parameters and of the other radiative transfer model parameters used to generate the LUT are listed in Table 1.

2.3 Forward model

The forward model provides a simulation of the instrument measurements corresponding to a certain atmospheric state (combination of aerosol and other atmospheric parameters) and a certain observation geometry. The forward model consists of two main parts. The first part relates the microphysical aerosol properties (optical thickness, single scattering albedo, phase matrix) to their optical properties in different layers of a model atmosphere. For this part of the forward model, pre-calculated optical properties as a function of the size parameters and of the real and imaginary part of the refractive index (Dubovik et al., 2006) are employed. The optical properties are pre-calculated for spheres and spheroids with an axis ratio distribution as proposed by Dubovik et al. (2006), and the aerosols are described by a mixture of spheres and spheroids. The optical properties for a given refractive index are obtained by spline interpolation from the tabulated values, and subsequently their integration over the aerosol size distribution is performed. The second part of the forward model solves the vector radiative transfer equation for given optical properties in different altitude layers in the model atmosphere. For this, the model of Hasekamp and Landgraf (2002, 2005), based on

Table 1. LUT entries for each radiative transfer model parameter. The model parameters are solar zenith angle (SZA), viewing zenith angle (VZA), surface pressure (p_s), surface albedo (a_s), aerosol effective radius (r_{eff}), effective variance (v_{eff}), refractive index real part (m_r) and imaginary part (m_i), fraction of spherical particles (n_s) and aerosol optical thickness (AOT). The superscripts f and c on the symbols of the aerosol parameters indicate the fine and the coarse mode respectively.

Parameter	Values
Wavelengths (nm)	440, 675, 870
SZA	15° to 75° in steps of 5°
VZA	0° to 65° in steps of 5°
p_s (hPa)	700, 1013
a_s (440, 675 nm)	0.02, 0.04, 0.06, 0.10, 0.15
a_s (870 nm)	0.10, 0.15, 0.20, 0.30, 0.45
r_{eff}^f (μm)	0.05, 0.1, 0.2, 0.3
v_{eff}^f (μm^2)	0.19
m_r^f	1.37, 1.45, 1.52, 1.6
m_i^f	0.001, 0.02, 0.1
n_s^f	1.0
τ_{aer}^f	0.025, 0.25, 0.5, 0.8, 1.0, 1.5, 3.0
r_{eff}^c (μm)	1.0, 2.5
v_{eff}^c (μm^2)	0.19
m_r^c	1.38, 1.44, 1.5
m_i^c	0.001, 0.05
n_s^c	0.1
τ_{aer}^c	0.025, 0.15, 0.5, 0.8, 1.0, 3.0

the Gauss–Seidel iterative method, is used. For the ground-based, upward-looking geometry, the analytical calculation of the derivatives of the measurement vector with respect to the state vector has not been implemented yet. For this reason, the derivatives are currently calculated by means of finite differencing.

3 Experimental setup

For this study, a spectropolarimeter designed for multi-angle ground-based observations based on the spectral modulation technique (Snik et al., 2009) is used. The key idea behind this technique is that the intensity of the incoming light (e.g. sunlight scattered in the atmosphere) becomes spectrally modulated after passing through a set of optical components (viz., an achromatic quarter-wave retarder, an athermal multiple-order retarder and a polarizing beam splitter). The corresponding modulation amplitude is proportional to the DLP, and the modulation phase is related to the angle of linear polarization (ALP). The DLP and the ALP of the incoming radiation can be derived from the spectrally modulated intensity through a demodulation algorithm as prescribed by Snik et al. (2009). This measurement concept eliminates the need for moving parts in the polarization measurement

device. Multi-angle measurements are obtained by sequentially pointing the instrument described above toward multiple zenith directions. A complete scan of the principal plane is obtained within 5–10 min.

The family of instruments constructed according to the concept described above has been given the name Spectropolarimeter EXperiment (SPEX). The ground-based spectropolarimeter used in this work will be hereinafter referred to as groundSPEX. A more detailed description of the instrument is given in van Harten et al. (2014).

4 Design of the neural network retrieval scheme

In order to train the NN, a set of scenarios simulating groundSPEX observations has been created. Multispectral and multi-angular reflectances and DLP measured by a groundSPEX-like instrument have been simulated for different combinations of aerosol parameters, solar zenith angles and values of the surface pressure and surface albedo. In the generation of this synthetic data set, cloud-free conditions have been assumed, and it has been hypothesized that the instrument performs observations in the principal plane. The principal plane is indeed the most favourable one for the retrieval of the aerosol properties, as it allows for sampling of the aerosol scattering matrix in the widest range of scattering angles. Since groundSPEX is currently unable to look less than 6° away from the Sun because of stray light, for simplicity of modelling it has been assumed that the groundSPEX-like instrument only scans the part of the principal plane with relative azimuth angle $\tilde{\varphi} = 180^\circ$ (i.e. the instrument only looks away from the Sun). With these settings, more than 10^6 simulations have been produced. A fraction of the simulated data has been used to train a NN for the retrieval of the aerosol properties; another fraction has been used to test the NN and to compare the retrieval scheme using the NN-based first guess to that using the LUT-based first guess.

It has been assumed that the measurements are performed at six viewing zenith angles (VZAs) (0, 6, 20, 32, 46 and 60°) and three wavelengths (440, 675 and 870 nm) that correspond to the wavelengths at which groundSPEX has been radiometrically calibrated. This choice allowed for a straightforward application of the method to real measurements. Finally, it has been assumed that the instrument carries out its observations on a surface covered by green vegetation. A surface albedo of 0.05 has been assumed at 440 and 675 nm, whereas the 870 nm albedo has been varied between 0.3 and 0.5 in the simulations.

Approximately 8×10^5 of the simulations described above have been used to train a NN to approximate the inverse relationship between the set of aerosol parameters x and the set of spectropolarimetric measurements y . Furthermore, about 1.65×10^5 independent simulations have been used to monitor the generalization ability of the NN during the training phase (validation set) so as to make sure that no overtrain-

ing was taking place, and an equal amount of simulations have been used to validate the NN retrievals after the training phase was terminated (test set).

Prior to their presentation to the NN, the multispectral and multi-angular reflectances and DLPs have been preprocessed using a linear principal component analysis (PCA), with the twofold goal of compressing and denoising the measurements. A method for processing multi-angular spectropolarimetric data using PCA prior to the presentation to a NN has been previously reported by Tsekeri et al. (2011), who separately applied the PCA to noise-free reflectances and DLPs, and selected the number of retained principal components (PCs) based on the percentage of explained variance. In this work, instead, an approach similar to that devised by Aires et al. (2002) for the processing of satellite infrared radiances was used. Specifically, the PCA was carried out on a set of simulated noise-free measurements in order to derive the transformation matrix to be applied to the measurement data. Then, a set of noisy simulated measurements was compressed and decompressed using the previously derived matrix, the reconstruction error was monitored as a function of the number of retained PCs versus the corresponding noise-free radiances, and the number of retained PCs corresponding to the minimum reconstruction root-mean-square error (RMSE) was finally chosen.

In mathematical terms, let \mathbf{y} be a generic vector of noise-free measurements (e.g. reflectance or DLP at a number of wavelengths and angles), let $\tilde{\mathbf{y}}$ be a noisy version of \mathbf{y} , and let \mathbf{P} be the matrix obtained applying the PCA to \mathbf{y} . The columns of such a matrix contain an orthonormal basis of eigenvectors of $\mathbf{S}_{\mathbf{y}}$, the covariance matrix of \mathbf{y} . If the truncation of \mathbf{P} obtained by retaining the eigenvectors associated to the n largest eigenvalues of $\mathbf{S}_{\mathbf{y}}$ is denoted by \mathbf{P}_n , it is possible to write the reconstructed vector of noisy measurements as $\tilde{\mathbf{y}}_n^r = \mathbf{P}_n^T \mathbf{P}_n \tilde{\mathbf{y}}$. Then, the error $E_n = \|\tilde{\mathbf{y}}_n^r - \tilde{\mathbf{y}}\|$ over the whole data set is monitored as a function of n , and the n that minimizes E_n is chosen.

This pre-processing scheme, separately applied to the reflectances and the DLPs, finally led to 7 PCs being retained for the reflectance and 10 PCs for the DLP.

In Fig. 1, the reconstruction RMSE is plotted as a function of the number of retained PCs for the radiances and the DLPs. As can be seen from the figure, the RMSE for the reflectance initially has a steep decrease for an increasing number of PCs until a number of 6–7 PCs (the latter being the optimal number according to the results) is reached. Then the RMSE starts to slowly increase again. A similar behaviour is observed for the DLP, with the difference that the RMSE decreases until 10 PCs are used. The increase in the reconstruction RMSE for a number of PCs higher than the optimal one indicates that the least significant PCs (11 for the reflectance, 8 for the DLP) contribute more noise than signal in the reconstruction of the observed quantities, and should therefore be discarded. Since the signal reconstructed with all the PCs (in our case, 18) is the same as the original noisy signal, it

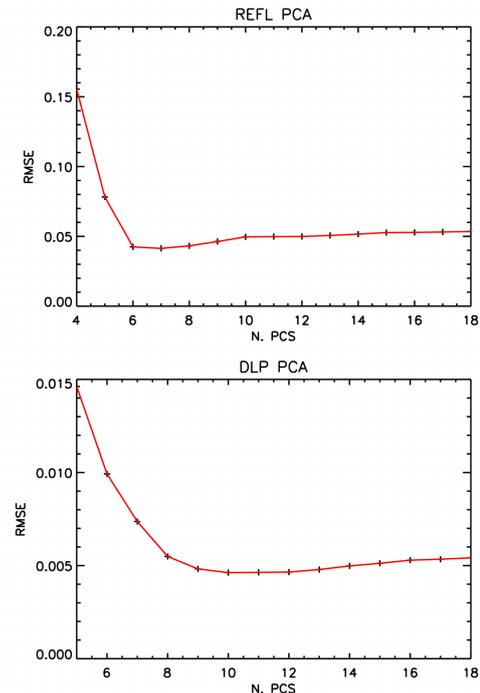


Figure 1. Reconstruction error versus number of retained principal components for the reflectances (above) and the degrees of linear polarization (below).

can be argued that the signal reconstructed with 7 (for the reflectance) or 10 (for the DLP) PCs is on average closer to the noise-free one than is the original noisy signal, and thus that the PCA has a beneficial denoising effect on the data. Similar results, and also similar plots, have been shown by other authors who have applied a similar type of processing to infrared hyperspectral data (see, for example, Antonelli et al., 2004).

An important point is that the applicability of our method to real data relies on the assumptions that (i) the sample of simulated data used for the computation of the covariance matrix for the PCA is representative enough of the natural variability of the measured quantities and (ii) a realistic noise model is used when the denoising performance of the algorithm is evaluated. As will be seen in Sect. 5.2, relying on these assumptions did not seem to significantly impair the performance of our retrieval method on real data.

The settings adopted in the generation of the training data set for the NN, as well as the NN input and output variables, are summarized in Table 2, and the range assumed for each of the parameters to be retrieved is given in Table 3, together with the mean and the standard deviation of each parameter. The solar zenith angle (SZA) and the surface pressure have been used as additional input variables for the NN, in order to account for the dependence of the simulated reflectances and DLPs on these parameters. The NN has been trained using an online error backpropagation

Table 2. Simulation settings and list of the quantities used as input and output variables in the design of the aerosol retrieval neural network.

Settings	Wavelengths for reflectance and DLP: 440, 675, 870 nm Viewing zenith angles: 0, 6, 20, 32, 46, 60° Relative azimuth angle: 180° Assumed aerosol size distribution: bimodal log-normal Assumed effective variance: $0.2 \mu\text{m}^2$ for both modes Assumed fraction of spherical particles: 100 % for fine mode, 10 % for coarse mode Assumed aerosol vertical distribution: Gaussian with peak at 1 km and FWHM* = 2 km Assumed surface type: green grass Assumed surface albedo at 440 and 675 nm: 0.05 Log-reflectance standard error: 0.026 at 440 nm, 0.046 at 675 nm, 0.076 at 870 nm DLP standard error: 0.004 at 440 nm, 0.006 at 675 and 870 nm
Inputs	Reflectance (7 principal components) DLP (10 principal components) Solar zenith angle Surface pressure
Outputs	Fine-mode effective radius (r_{eff}^f) Fine-mode refr. index real part (m_r^f) Fine-mode refr. index imaginary part (m_i^f) Fine-mode AOT at 550 nm (τ_{aer}^f) Coarse-mode effective radius (r_{eff}^c) Coarse-mode refr. index real part (m_r^c) Coarse-mode refr. index imaginary part (m_i^c) Coarse-mode AOT at 550 nm (τ_{aer}^c) Surface albedo at 870 nm (a_s)

* Full width at half maximum

algorithm (Rumelhart et al., 1986) accelerated by means of a learning rate annealing technique (Böös and Amari, 1999) that consists in reducing the learning rate of the NN while the training proceeds. The choice of the online backpropagation learning instead of more sophisticated algorithms like scaling conjugate gradient (Møller, 1993) or Levenberg–Marquardt (Hagan and Menhaj, 1994) – which are reported to be, in principle, faster and more stable than the online backpropagation – is motivated by the fact that these latter algorithms require that the entire training set be presented to the NN before each update of the network weights. This can cause slow training when the training set is very large (Wilson and Martinez, 2000), as is the case in this work. The training process has been regularized by adding Gaussian noise to the simulated measurements during the training phase. This technique prevents the NN from yielding unreasonable responses when stimulated by noisy inputs (Holmström and Koistinen, 1992; Bishop, 1995b), and is therefore of critical importance for the applicability of a NN to real-world data.

As pointed out in Table 2, the NN designed following the method described above has an input layer of 19 neurons (7 reflectance PCs and 10 DLP PCs, plus SZA and surface pressure) and an output layer of 9 neurons (8 aerosol parameters plus the surface albedo at 870 nm). The number of hidden layers and neurons was selected by comparing a number of different architectures and selecting the architecture that

gave the lowest RMSE on the validation set. The chosen architecture consists of 3 hidden layers with 40 neurons per layer. The hidden neurons have a hyperbolic tangent activation function (the reader may refer to Appendix A for the meaning of the term “activation function”), whereas the activation function of the output neurons is linear.

At a first glance, the result found for the number of hidden layers may seem surprising, given that NNs with more than 2 hidden layers are rarely reported in the literature, at least in remote sensing applications (in other fields of application, “deeper” NNs are less uncommon; see, for example, LeCun et al., 1989, for an interesting application of a MLP with three hidden layers). The most often reported reason why simpler NN architectures are usually chosen is perhaps that more complex architectures (i.e. NNs with more hidden layers and neurons) are more prone to overfitting, i.e. a situation where the training errors are small but poor performance is obtained for cases that are not included in the training set. This is, in principle, true, but it is important to keep in mind that the relationship between complexity and generalization also depends on the complexity of the function to be approximated and on the number of training data (the more training data, the lower the risk of overfitting for a given architecture). The fact that the NN described in this paper has been trained with a very large number of data, together with the fact that the backpropagation algorithm seems to have a bias towards

Table 3. Minimum, maximum, mean and standard deviation of the aerosol parameters used to generate the synthetic data set for this study. Please refer to Table 2 for the meaning of the symbols in the first column.

Parameter	Min	Max	Mean	SD
$r_{\text{eff}}^{\text{f}}$ (μm)	0.020	0.620	0.320	0.175
m_{r}^{f}	1.330	1.590	1.460	0.076
m_{i}^{f}	0.001	0.410	0.080	0.128
$\tau_{\text{aer}}^{\text{f}}$	0.000	3.000	0.928	0.922
$r_{\text{eff}}^{\text{c}}$ (μm)	0.800	4.000	2.400	0.980
m_{r}^{c}	1.330	1.590	1.460	0.076
m_{i}^{c}	0.001	0.410	0.080	0.128
$\tau_{\text{aer}}^{\text{c}}$	0.000	2.000	0.598	0.587
a_{s}	0.350	0.500	0.425	0.048

smooth functions even when the NN has many parameters (Lawrence et al., 1997; Caruana et al., 2001), makes the results found for the number of hidden layers more plausible than it may be intuitively expected.

Table 4 summarizes a number of error statistics used to evaluate the performance of the NN on the training, validation and test set. The RMSE, the mean absolute error (MAE), the mean bias and the Pearson correlation coefficient ρ between retrieved and actual aerosol parameters are reported for each aerosol parameter. The effective radius and the complex refractive index for each mode are provided only if the retrieved AOT of the considered mode is larger than 0.05. It can be seen that no significant differences are present in the performance of the NN on the three sets. This indicates that the NN can be reliably applied to data that are not included in the training set.

5 Use of the neural network retrievals as first guess in an iterative scheme

A further step in this study has been the use of the aerosol parameters retrieved by the NN as first guess for the variational retrieval scheme described in Sect. 2, as a replacement for the LUT. In principle, there are at least two reasons why it might be worthwhile applying a variational retrieval after the NN instead of using the NN as a stand-alone method. First, since the NN is trained using a cost function that encompasses a wide range of physical situations, its estimate might not be the optimal one for individual cases, and the application of a variational retrieval might refine the NN output by defining a case-specific cost function. Second, although some methods exist for characterizing the uncertainty associated with the output of a NN (see, for example, Dybowski and Roberts, 2001, for a general review, and Ristovski et al., 2012, for an application to aerosol retrievals), it is still not straightforward to associate the most widespread diagnostics (averaging kernels, full covariance matrix of the retrieval, degrees

of freedom for signal) to NN retrievals, whereas such diagnostics can be provided using a variational retrieval scheme. Furthermore, the variational approach provides a means for checking the credibility of each individual retrieval using the χ^2 as a diagnostic.

The drawback of the proposed approach is that the advantage of a superior speed offered by a NN retrieval is lost because of the need for multiple calls to a radiative transfer model. This drawback might be overcome, in principle, by training a second NN to approximate the radiative transfer model (Schiller, 2007), but the accuracy requirements for such a network would be stringent and difficult to meet for a forward model with many independent input parameters.

In order to assess the impact of using the NN as first guess in the variational retrieval algorithm described in Sect. 2, the scheme with a NN-based first guess (hereinafter named NN + PT, where PT stands for Phillips–Tikhonov) has been compared to the scheme with a LUT-based first guess (hereinafter named LUT + PT) on two test data sets, one consisting of simulated data and one consisting of real observations. The two data sets, as well as the results of the comparisons, are described in the following subsections.

5.1 Application to simulated data

The synthetic data set chosen for the comparisons consists of 5000 combinations of the aerosol parameters, randomly drawn from the data set used to test the NN after the training process. The only reason why only a subset of the test data set has been used to compare the two retrieval algorithms, instead of the whole data set, is that running both the retrieval schemes on 1.65×10^5 data would have required much more time, while randomly selecting a few thousands cases is expected to be enough in order to draw reliable conclusions about the performance of the two retrieval schemes.

The combinations of aerosol parameters to be retrieved have been used as input for new radiative transfer simulations, aimed at creating a “pseudo-reality” in which the retrieval algorithm has to operate. In order to obtain error statistics that are as indicative as possible of those that can be obtained in an operational scenario, the values of some of the non-retrieved aerosol parameters (effective variance, fraction of spherical particles of the coarse mode, aerosol layer height) have been randomly perturbed with respect to those assumed in the training of the NN and in the retrieval algorithm by replacing the assumed values with values drawn from a uniform distribution (between 0.1 and 0.3 for the effective variance of both modes, between 0.03 and 0.06 for the surface albedo at 440 and 675 nm, between 0 and 1 for the fraction of spherical particles of the coarse mode).

The results of the comparison between the two retrieval schemes are summarized in Table 5. The variational retrievals have been assumed to have converged if the final value of the goodness of fit (χ^2 as defined in Eq. 4) was less than 2 within a maximum of 20 iterations. Again, the effec-

Table 4. Root-mean-square error (RMSE), mean absolute error (MAE), mean bias and Pearson correlation coefficient (ρ) between the retrieved and the actual aerosol parameters, computed on the training, test and validation set. Please refer to Table 2 for the meaning of the symbols in the first column.

Variable	Training set				Validation set				Test set			
	RMSE	MAE	Bias	ρ	RMSE	MAE	Bias	ρ	RMSE	MAE	Bias	ρ
$r_{\text{eff}}^{\text{f}}$	0.061	0.038	-0.012	0.940	0.061	0.038	-0.012	0.939	0.061	0.038	-0.012	0.939
m_{r}^{f}	0.046	0.035	0.011	0.806	0.046	0.034	0.011	0.806	0.046	0.035	0.011	0.804
m_{i}^{f}	0.048	0.020	0.004	0.930	0.048	0.020	0.004	0.930	0.048	0.020	0.004	0.931
$\tau_{\text{aer}}^{\text{f}}$	0.166	0.105	-0.004	0.984	0.167	0.105	-0.004	0.983	0.168	0.106	-0.005	0.983
$r_{\text{eff}}^{\text{c}}$	0.752	0.569	-0.130	0.654	0.751	0.570	-0.126	0.654	0.754	0.571	-0.130	0.651
m_{r}^{c}	0.062	0.050	-0.001	0.569	0.062	0.050	-0.001	0.566	0.062	0.050	-0.001	0.568
m_{i}^{c}	0.085	0.038	-0.020	0.772	0.085	0.038	-0.020	0.768	0.086	0.038	-0.021	0.770
$\tau_{\text{aer}}^{\text{c}}$	0.166	0.106	0.011	0.960	0.165	0.106	0.010	0.960	0.166	0.106	0.011	0.959
a_{s}	0.036	0.029	-0.005	0.665	0.036	0.029	-0.005	0.664	0.036	0.029	-0.005	0.663

tive radius and the complex refractive index for each mode are included in the statistics only if the retrieved AOT of the considered mode is larger than 0.05. It is clear from Table 5 that the fraction of converging retrievals for the NN + PT algorithm was considerably higher than that achieved by the LUT + PT scheme (slightly higher than 55 % for the latter, slightly lower than 95 % for the former). The second and third panel of Table 5 summarize a number of error statistics computed for the two retrieval schemes on the benchmark data set. In particular, the second panel of Table 5 reports RMSE, MAE, bias and Pearson correlation coefficient for the estimates produced by the LUT alone and at the end of the LUT + PT scheme, taking only the converging LUT + PT retrievals into account. The same is done in the third panel for the NN and the NN + PT retrievals. A slight difference between the LUT and the NN first guess is that the LUT does not fit the surface albedo at 870 nm, for which a prior of 0.4 is always assumed. However, this difference does not compromise the comparison between the two methods, because the iterative retrieval is in most cases able to recover from a wrong initial value for this parameter, as shown in Table 5.

A number of elements are noteworthy in the observed results.

1. By comparing the RMSE or the MAE of the LUT + PT to that of the NN + PT scheme (upper and lower right panels in Table 5), it is possible to see that the NN + PT scheme seems to outperform the LUT + PT scheme in the estimation of all the aerosol parameters. This is especially true for the fine-mode effective radius and refractive index (real and imaginary part), and for the imaginary part of the coarse-mode refractive index. For these variables, the decrease in the RMSE is also accompanied by a substantial rise in the correlation coefficient. The improvement in the fine-mode AOT, instead, is not very big. This is probably because the relationship between the measurements and this parameter is

closer to a linear one, so that the iterative method is less dependent on the first guess for this particular variable.

2. The differences seem to a large extent attributable to the fact that the quality of the first guess produced by the NN is much better than that of the first guess produced by the LUT. This applies to all the retrieved quantities. It can be seen that the NN retrievals alone already seem to perform better than the full LUT + PT retrievals.
3. For both algorithms, the coarse-mode effective radius and real part of refractive index seem the most difficult parameters to retrieve. Even though an improvement is noticed by replacing the LUT with a NN, the errors are quite large and the correlation coefficients are below 0.50 for the effective radius and even below 0.45 for the real part of the refractive index. This behaviour is partly explained by the uncertainty in the fraction of spherical coarse-mode particles, which is assumed to be constant in the retrieval algorithm but is allowed to vary in the pseudo-reality. Furthermore, the absence of viewing angles near the forward scattering direction reduces the sensitivity to coarse-mode aerosols, because the region of enhanced intensity around such a direction – which is a distinctive feature of large particles – is not sampled.
4. By comparing the error statistics of the NN retrievals to those of the NN + PT retrievals (third panel in Table 5), no improvement is observed in many of the retrieved parameters after the application of the iterative retrieval to the NN outputs. The only parameters that improve are the fine-mode AOT and the surface albedo at 870 nm. In Fig. 2 the histograms of the absolute errors in the retrieved parameters for the NN alone and for the NN + PT schemes are shown. It is clear that the iterative retrieval generally improves the estimates of the surface albedo and of the fine- and coarse-mode AOTs with respect to the NN alone, whereas the improvement

Table 5. Root-mean-square error (RMSE), mean absolute error (MAE), mean bias and Pearson correlation coefficient (ρ) between retrieved and actual aerosol parameters. Please refer to Table 2 for the meaning of the symbols in the first column.

	Method				N converging retrievals (fraction)			
	LUT + Phillips–Tikhonov				2756 (55.12 %)			
	NN + Phillips–Tikhonov				4321 (93.60 %)			
Variable	LUT only				LUT + Phillips–Tikhonov			
	RMSE	MAE	Bias	ρ	RMSE	MAE	Bias	ρ
$r_{\text{eff}}^{\text{f}}$	0.147	0.109	−0.088	0.676	0.109	0.066	−0.032	0.752
m_{r}^{f}	0.134	0.113	0.094	0.158	0.085	0.062	0.028	0.530
m_{i}^{f}	0.097	0.045	−0.023	0.391	0.075	0.034	−0.012	0.687
$\tau_{\text{aer}}^{\text{f}}$	0.405	0.272	−0.112	0.897	0.290	0.161	−0.034	0.946
$r_{\text{eff}}^{\text{c}}$	1.289	1.032	−0.599	0.142	1.430	1.024	−0.422	0.192
m_{r}^{c}	0.090	0.073	−0.012	0.100	0.095	0.074	0.002	0.320
m_{i}^{c}	0.118	0.065	−0.026	0.168	0.109	0.051	−0.029	0.444
$\tau_{\text{aer}}^{\text{c}}$	0.559	0.333	0.019	0.581	0.356	0.202	0.017	0.784
a_{s}	0.054	0.042	−0.025	0.000	0.044	0.022	−0.001	0.727
Variable	NN only				NN + Phillips–Tikhonov			
	RMSE	MAE	Bias	ρ	RMSE	MAE	Bias	ρ
$r_{\text{eff}}^{\text{f}}$	0.077	0.048	0.000	0.895	0.078	0.044	0.008	0.896
m_{r}^{f}	0.052	0.038	0.010	0.742	0.059	0.041	0.010	0.707
m_{i}^{f}	0.060	0.023	−0.003	0.883	0.063	0.025	−0.007	0.869
$\tau_{\text{aer}}^{\text{f}}$	0.226	0.144	0.027	0.971	0.210	0.117	0.026	0.975
$r_{\text{eff}}^{\text{c}}$	0.882	0.688	−0.218	0.505	0.948	0.723	−0.206	0.474
m_{r}^{c}	0.066	0.054	0.001	0.506	0.079	0.061	0.005	0.448
m_{i}^{c}	0.090	0.043	−0.019	0.735	0.089	0.042	−0.014	0.733
$\tau_{\text{aer}}^{\text{c}}$	0.231	0.146	−0.011	0.922	0.229	0.129	−0.033	0.926
a_{s}	0.045	0.036	0.004	0.477	0.033	0.016	−0.002	0.816

is more limited in the effective radius and in the complex refractive index of both modes.

These results may seem to advocate the use of the NN as a stand-alone scheme without applying the iterative method afterwards. However, such a solution would result in a less complete error analysis, and in the impossibility of checking the plausibility of individual retrievals (see discussion at the beginning of Sect. 5).

A word of caution is necessary in the interpretation of Table 5, because there are some parameters (fine- and coarse-mode effective radius, coarse-mode imaginary part of refractive index) whose mean values in the training data set are close to the maximum values stored in the LUT or much larger. As a result, the comparison between the NN and the LUT alone may be unfair towards the LUT, since the error of the LUT retrieval will necessarily be large when the true value of one of these parameters is far larger than the LUT maximum. To some extent, this unfairness may also affect the comparison between the full retrievals, because in cases like those described above it may be expected that the first guess will be far from the true value and the convergence to such value will be difficult to achieve. In order to

assess to what extent this issue affects our findings, we compared the LUT + PT and the NN + PT schemes on a subset of cases where all the true values of the aerosol parameters were within the LUT limits. The number of such cases was 774. On this subset, the fraction of successful retrievals was around 59 % for the LUT + PT scheme and around 88 % for the NN + PT scheme. Therefore, also looking at this subset, the main point of this comparison – i.e. that using the NN as first guess for the iterative retrieval improves the number of successful retrievals – remains largely unaffected. On the reduced data set, the accuracy of the LUT + PT retrievals was found comparable to that of the NN + PT retrievals for the fine-mode effective radius, whereas the NN + PT scheme still performed better than the LUT + PT scheme on all the other parameters.

A comparison between the LUT + PT and the NN + PT scheme in terms of speed is reported in Fig. 3, where histograms of the number of iterations needed to converge and of the total retrieval time are shown. It can be seen that the Phillips–Tikhonov scheme takes on average two iterations less to converge when it is initialized using the NN as a first guess than it does when the LUT is used as a first guess.

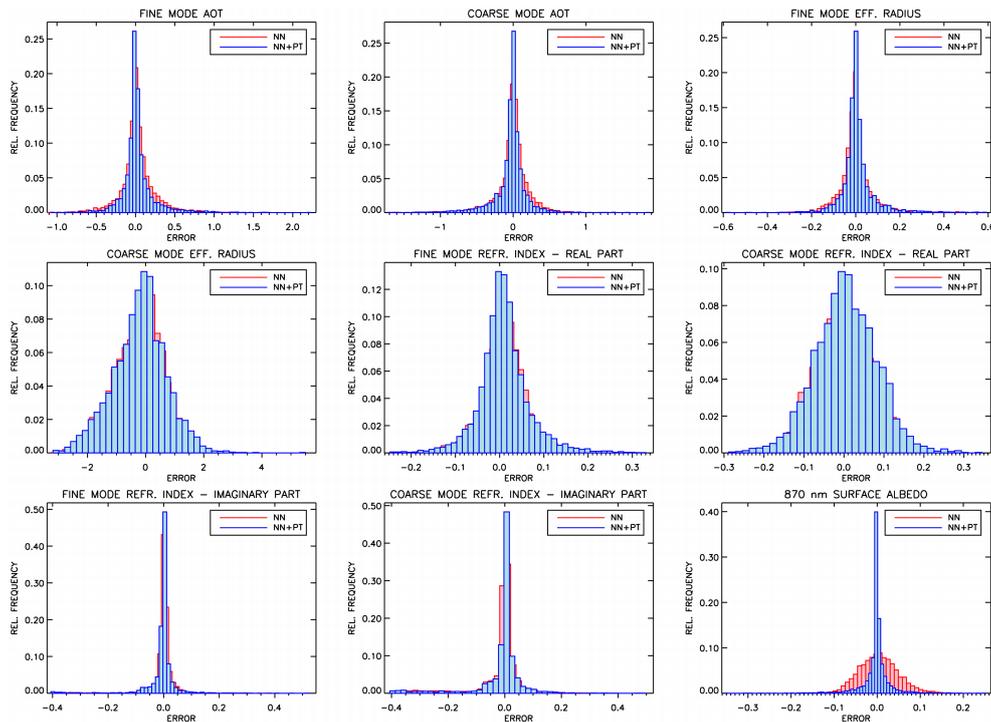


Figure 2. Histograms of the absolute errors in the retrieved parameters over the benchmark synthetic data set for the NN-only retrievals (red) and for the Phillips–Tikhonov retrievals with the NN-based first guess (blue).

The average difference in the total retrieval time is about 7 s. This difference is partly explained by the smaller number of iterations and partly by the time needed to generate the first guess. The generation of the first guess (histograms not shown) takes about 2 s when the LUT is used, and about 0.2 s when the NN is used. These 0.2 s are dominated by the radiative transfer simulation performed to evaluate the fit between the NN retrieval and the measurements, as the NN retrieval alone only takes about 4 ms. The total computation time, which is on average around 15 s for the NN + PT scheme, can be significantly reduced if the numerical computation of the Jacobians of the forward model is replaced by an analytical one. In the satellite retrieval algorithm by Hasekamp et al. (2011), where this has been done, the total retrieval time is of about 2 s, and replacing the LUT with a NN may result in a further speed-up.

5.2 Application to real measurements

The measurements used in this study have been performed using the groundSPEX instrument between 7 and 9 July 2013 and on 5 September 2013 at the Cabauw Experimental Site for Atmospheric Research (CESAR Observatory, <http://www.cesar-observatory.nl>), a few tens of kilometres southwest of Utrecht, in the Netherlands (Monna and Bosveld, 2013). The measurements were made under clear-sky conditions. During the July measurements, an elevated smoke layer was present over the Netherlands as a consequence

of air advection from North America (KNMI, 2013). The LUT + PT and the NN + PT retrieval schemes have been applied to this set of measurements, and the retrieved aerosol parameters have been compared to those estimated by the Aerosol Robotic Network (AERONET) station located at the CESAR site.

Given the unavailability of AERONET level 2 data for the considered dates at the time of writing, level 1.5 data have been used for the comparison. The total AOTs (fine plus coarse mode) and the Ångström exponents between 440 and 675 nm estimated by groundSPEX using the LUT + PT and the NN + PT schemes have been compared to those provided by the AERONET direct sun product (Holben et al., 1998). The fine- and coarse-mode effective radii and complex refractive indices have been compared to those provided by the AERONET product based on the almucantar algorithm described by Dubovik and King (2000). As for the complex refractive index, the almucantar product provides a single value that applies to both modes and is spectrally dependent, while our algorithm provides a spectrally independent value for each mode. Therefore, prior to the comparison, the AERONET refractive indices have been averaged with respect to the wavelength, and the fine- and coarse-mode refractive indices estimated by groundSPEX have been combined in a weighted average using the AOTs of the two modes as weights.

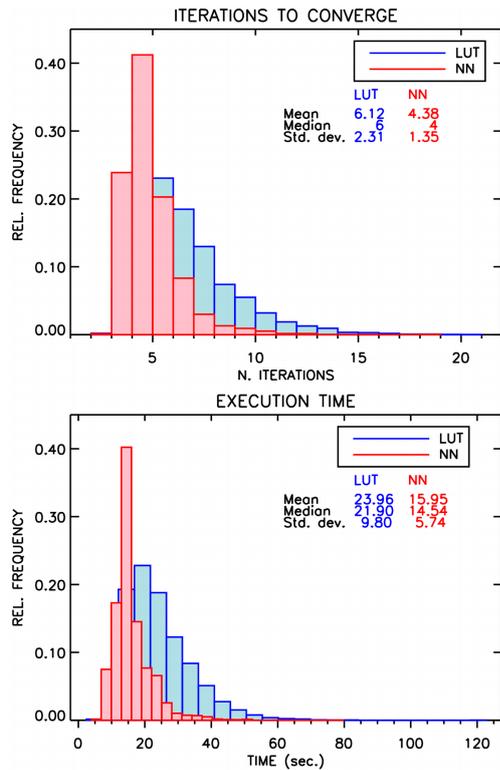


Figure 3. Histograms of the number of iterations before convergence (above) and of the total retrieval time (below) for the Phillips–Tikhonov retrievals using the LUT first guess (blue) and the NN first guess (red).

While AERONET total AOTs and Ångström exponents obtained from direct sun measurements are intrinsically robust reference values, AERONET almucantar retrievals of the aerosol microphysical parameters are often affected by non-negligible uncertainties (Dubovik et al., 2000). As a consequence, it is important to bear in mind that the comparison between groundSPEX and AERONET microphysical parameters should not be regarded as a validation of groundSPEX against a completely established truth. The meaning of this intercomparison is rather that of a consistency check between retrievals of a similar set of parameters made by means of two profoundly different methodologies. Because of this, cases of consistency between the retrievals reinforce our confidence in both the retrieval approaches, whereas cases of inconsistency do not necessarily imply that groundSPEX is performing worse than AERONET.

In total, 71 measurements were carried out by groundSPEX during the considered period. While this is by no means an extensive benchmark data set, it is useful in order to have at least a preliminary idea of the applicability of the NN – trained on synthetic data – to a real-world scenario. Out of the attempted retrievals, 48 of the LUT + PT retrievals satisfied the χ^2 criterion mentioned in Section 5.1, whereas the number of successful NN + PT retrievals was 59. In Fig. 4

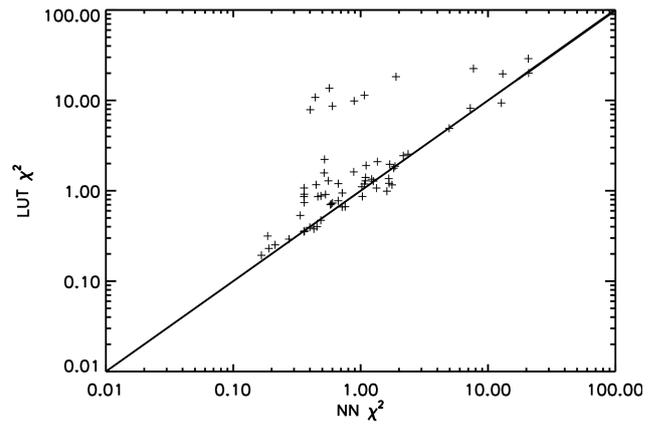


Figure 4. Plot of the final χ^2 obtained using the LUT + PT scheme (ordinate) versus that obtained using the NN + PT scheme (abscissa) for the considered groundSPEX measurements.

the final χ^2 obtained by the LUT + PT scheme is plotted against that obtained by the NN + PT scheme applied to the considered set of groundSPEX measurements. Points that are close to the 1 : 1 line indicate cases where the two algorithms performed similarly. Points that are well above the 1 : 1 line indicate cases where the NN + PT scheme performed better than the LUT + PT (i.e. LUT + PT achieved a higher χ^2 than NN + PT), whereas points located well below the line represent cases where LUT + PT performed better than NN + PT. From this plot, it is straightforward to note that the number of points above the identity line is considerably larger than the number of points located below the line. This means that, besides yielding a higher number of successful retrievals, the NN + PT algorithm also has the tendency to produce final solutions that fit the measured reflectances and DLPs better than those obtained using the LUT + PT scheme.

Figures 5, 6 and 7 respectively show time series of total AOTs, fine- and coarse-mode effective radii and AOT-weighted refractive indices retrieved by the two algorithms and those estimated by AERONET during the 4 days considered in this study. Superimposed on the retrieved parameters are error bars derived by error propagation using the formalism of Rodgers (1990). According to this formalism, the retrieval error can be decomposed in a contribution due to random noise and a contribution due to the a priori error (called null space error). The null space error depends on the assumed a priori error. As explained in Sect. 2, our retrieval scheme uses the first guess produced by the LUT or by the NN as a priori. However, a method to derive the uncertainty associated with the LUT or the NN retrievals alone has not been implemented yet; therefore the a priori error is set to the reciprocal of the half width of the physical variation range assumed for each parameter. As a consequence, it is likely that the width of the error bars shown in this paper is overestimated, especially in the case of the NN retrievals. The error

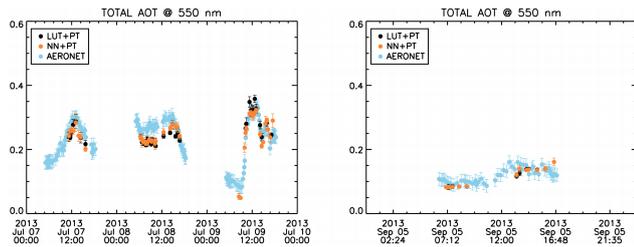


Figure 5. Time series of the total AOT retrieved by the LUT + PT scheme (black circles), the NN + PT scheme (orange circles), and the AERONET direct sun level 1.5 product (sky-blue circles): between 7 and 9 July 2013 (left) and on 5 September 2013 (right).

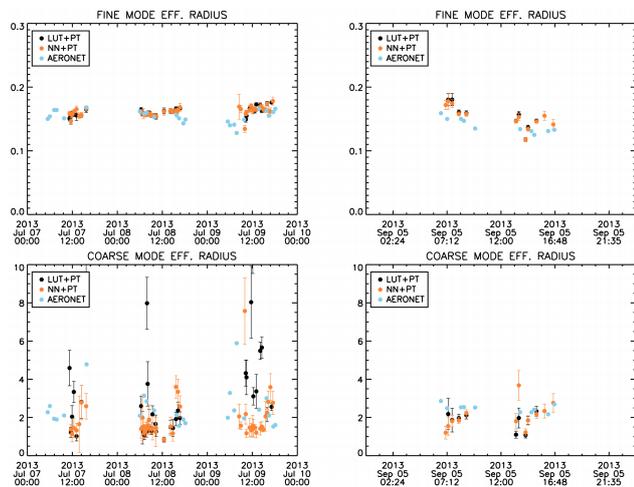


Figure 6. Time series of the fine- (above) and coarse-mode (below) effective radius retrieved by the LUT + PT scheme (black circles), the NN + PT scheme (orange circles), and the AERONET almucantar level 1.5 product (sky-blue circles): between 7 and 9 July 2013 (left) and on 5 September 2013 (right).

bars for the AERONET retrievals are the same as those used by van Harten et al. (2014).

For both algorithms, the agreement seems generally good for the AOTs (absolute deviation 95th percentile around 0.05). During the 3 days in July, the same applies to the fine-mode effective radii (absolute deviation 95th percentile around 0.02 μm) and to the real part of the AOT-weighted refractive indices (groundSPEX–AERONET differences well within the estimated AERONET error bars). On 5 September, the agreement for these parameters seems less good. It must be noted, however, that the lower AOTs that were generally observed on this latter date make the retrieval of the aerosol microphysical parameters more difficult for both groundSPEX and AERONET. The NN + PT retrievals also seem more capable than the LUT + PT retrievals of reproducing the temporal variation of the imaginary part of the refractive index, although the agreement between the values varies from day to day. This is particularly evident during 9 July (Fig. 7, lower left panel), when the

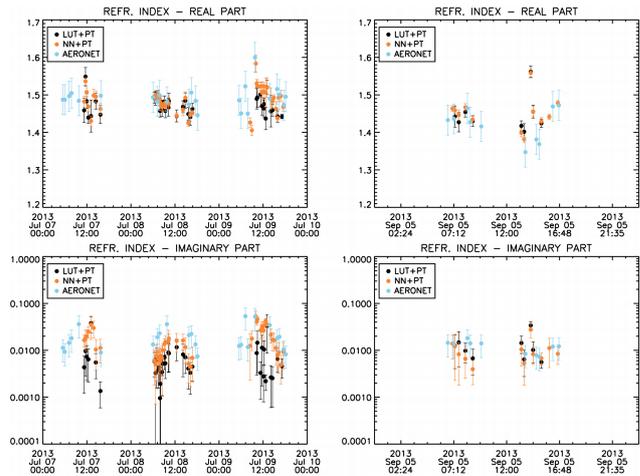


Figure 7. Time series of the real (above) and imaginary part (below) of the (AOT-weighted) refractive index retrieved by the LUT + PT scheme (black circles), the NN + PT scheme (orange circles), and the AERONET almucantar level 1.5 product (sky-blue circles): between 7 and 9 July 2013 (left) and on 5 September 2013 (right).

temporal variation of the imaginary part of the refractive index retrieved by AERONET is followed very well by the groundSPEX retrievals based on the NN + PT scheme (again, from visual inspection of Fig. 7 it is clear that the differences between groundSPEX and AERONET are well within the AERONET error bars), whereas the estimates produced using the LUT + PT scheme are more distant from the AERONET values.

A closer look at the total AOTs shown in Fig. 5 seems to indicate that a slight underestimation tendency exists in both retrieval schemes. In order to explain this fact, it might be interesting to compare the fine- and coarse-mode AOTs retrieved by groundSPEX to those provided by the AERONET product based on the spectral deconvolution algorithm (SDA) by O’Neill et al. (2003). This is done in Fig. 8. In the interpretation of this comparison, it must be kept in mind that the definition of the aerosol fine and coarse modes in our algorithm is different from that used in the SDA product, owing to differences in the assumptions made about the aerosol size distribution. Despite the fact that the groundSPEX and the SDA optical thicknesses are not rigorously comparable for the reason mentioned above, it is interesting to note that an excellent agreement exists between the groundSPEX and the SDA fine-mode AOTs, with the NN + PT that follows the diurnal cycle of the fine-mode AOT particularly well (absolute deviation 95th percentile around 0.06, but 85th percentile smaller than 0.03). Instead, the coarse-mode AOTs estimated by the two groundSPEX retrieval methods are almost always lower than those provided by the AERONET SDA product. This may suggest that the underestimation tendency in the total AOT could be, in some way, attributable to the coarse mode. However, the different definitions of the aerosol modes in

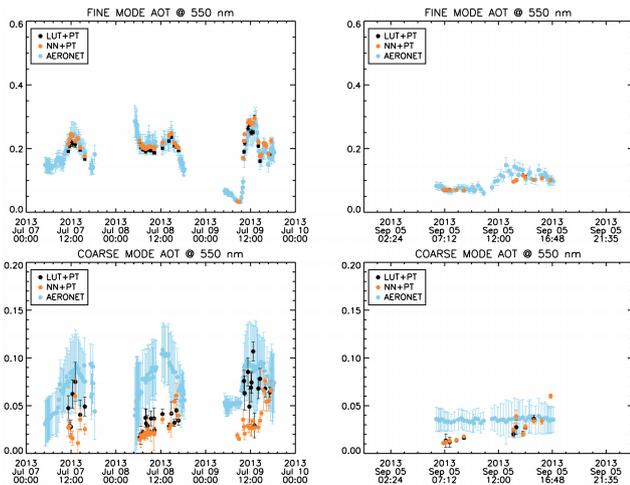


Figure 8. Time series of the fine- (above) and coarse-mode (below) AOT retrieved by the LUT + PT scheme (black circles), the NN + PT scheme (orange circles), and the AERONET SDA level 1.5 product (sky-blue circles): between 7 and 9 July 2013 (left) and on 5 September 2013 (right).

the two algorithms, as well as the lack of a more appropriate benchmark for the groundSPEX fine- and coarse-mode AOTs, do not allow us to provide stronger evidence for this statement.

Another parameter that is also indicative of the aerosol particle size and is less dependent on the mode definition is the Ångström exponent. In Fig. 9 the Ångström exponents between 440 and 670 nm retrieved by the two groundSPEX algorithms are compared to those measured by AERONET. The agreement seems generally good during July, whereas on 5 September groundSPEX systematically estimates smaller Ångström exponents than AERONET. This would indicate that groundSPEX sees larger particles than AERONET during 5 September. However, this conclusion does not seem very consistent with the results previously shown for the coarse-mode effective radius and the coarse-mode AOTs. In view of this, no conclusive interpretation of the differences between groundSPEX and AERONET with respect to the particle size can be given at this moment in cases of low AOT.

It is interesting to compare the time series of the aerosol parameters retrieved by the NN + PT scheme to those retrieved by the NN alone. Examples of such comparisons are shown in Fig. 10 for the fine-mode AOT and effective radius, as well as for the AOT-weighted refractive index, during July 2013. It appears that the application of the iterative retrieval after the NN estimate brings the retrieved values closer to the AERONET values for the fine-mode AOT and effective radius and for the real part of the refractive index. For the imaginary part of the refractive index, the updates brought by the Phillips–Tikhonov method to the NN retrievals do not seem to improve the accuracy of the final estimates significantly. On the one hand, this might be con-

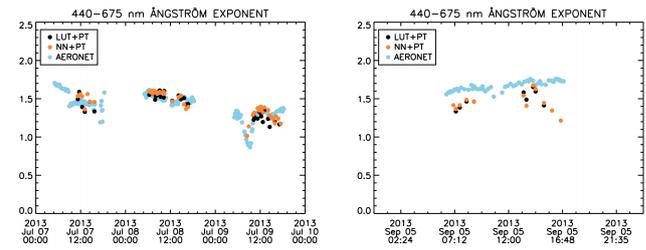


Figure 9. Time series of the 440–670 nm Ångström exponent retrieved by the LUT + PT scheme (black circles), the NN + PT scheme (orange circles), and the AERONET direct sun level 1.5 product (sky-blue circles): between 7 and 9 July 2013 (left) and on 5 September 2013 (right).

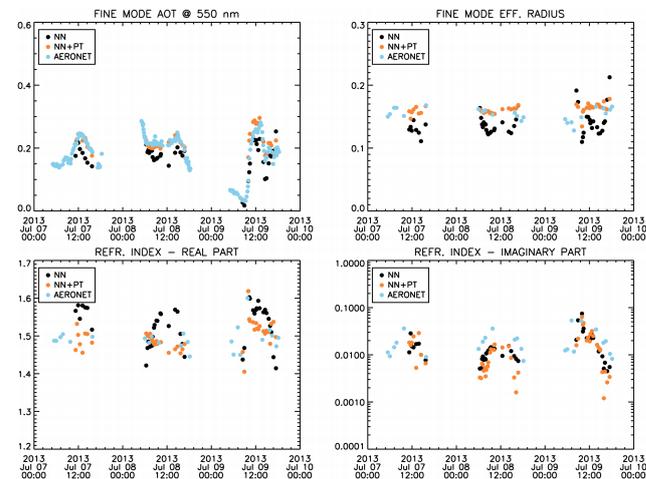


Figure 10. Fine-mode AOT (upper left) and effective radius (upper right) and AOT-weighted real (lower left) and imaginary part (lower right) of the refractive index estimated using the NN alone (black circles), the NN + PT scheme (orange circles), and reference values from AERONET (sky-blue circles).

tradicting the results shown obtained with simulated data, but on the other hand it must be kept in mind that the range of situations sampled by the real measurements during July 2013 is much narrower than that encompassed by the simulations; therefore one should not necessarily expect consistent error statistics between the two cases. In any case, the results observed so far on real data seem to suggest that there is still merit in applying a full physics retrieval after the NN inversion, even at the cost of a longer processing time.

6 Conclusions

In this paper, the use of neural networks for the retrieval of aerosol properties from ground-based multi-angle measurements of intensity and polarization has been investigated. A neural network for the retrieval of aerosol properties from ground-based multi-angular spectropolarimetric data has been developed, and has been used as a replace-

ment for a look-up table as an initial guess for an iterative aerosol retrieval algorithm. In order to assess the impact of the neural-network-based first guess, the iterative algorithm initialized using the neural first guess and its version initialized using the LUT-based first guess have been compared on a data set consisting of simulated measurements, and on a set of real observations carried out by the groundSPEX instrument.

The synthetic data set consisted of 5000 cases generated by simulating the reflectance and the degree of linear polarization that would be measured by a groundSPEX-like instrument for a wide range of combinations of the aerosol parameters. On this data set, the replacement of the LUT-based first guess with a NN-based one has led to a remarkable increase in the fraction of successful retrievals (i.e. retrievals that achieved a goodness of fit indicator smaller than 2 within 20 iterations). While approximately 55 % of the retrievals successfully converged in the case of the LUT first-guess algorithm, almost 95 % of the retrievals converged when the NN was used instead of the LUT as initial guess. Furthermore, the converging retrievals based on the NN first guess exhibited smaller errors than those based on the LUT first guess in the coarse-mode effective radius and refractive index. The application of the iterative algorithm after the NN retrieval improved the estimates of the fine- and coarse-mode AOT with respect to the NN-only retrieval. Instead, no significant improvements were noticed in the other aerosol parameters, for which the quality of the NN output alone was found to be comparable to that of the final retrieval.

Also, when the two algorithms were applied to real measurements carried out by the groundSPEX instrument at the CESAR atmospheric observatory, a higher number of successful retrievals was observed after replacing the LUT first guess with the NN first guess. The observed improvement in the number of converging retrievals was not as large as that observed with simulated data. This might be due to the fact that the groundSPEX measurements were very concentrated in time, with small intrinsic variations in the aerosol properties, whereas the synthetic data set samples a much broader range of situations, including cases where the first guess produced by the LUT might be much worse than that produced by the NN.

The comparison between the groundSPEX aerosol retrievals and the AERONET retrievals showed a good agreement in the total AOT (absolute differences most often below 0.05) and in the AOT-weighted refractive index. In particular, replacing the LUT with the NN led to a remarkable improvement in the retrieval of the imaginary part of the refractive index. A good agreement (absolute errors often below $0.02 \mu\text{m}$) was found in the fine-mode effective radius during the 3 days in July 2013, whereas the agreement was worse on 5 September, probably because the smaller fine-mode AOT observed on that day made the retrieval of the aerosol microphysical properties more challenging for both groundSPEX and AERONET. The coarse-mode effective radius seems, in all the cases, the most difficult parameter to retrieve, and this seems to confirm the results obtained with simulated data. A negative bias in the total AOT was observed in the groundSPEX retrievals, independent of the first guess. Qualitative comparisons against the SDA fine- and coarse-mode AOTs seem to suggest that this bias could be attributed to the coarse-mode AOT. However, a rigorous quantification of the bias using the SDA product is not possible, because the definitions of fine and coarse mode made in this product are different from those used by the groundSPEX algorithm. Also, the analysis of the retrieved Ångström exponents, which are indicative of the particle size but are independent of the mode definitions, did not lead to conclusive results, because it seemed to suggest that groundSPEX overestimates the particle sizes with respect to AERONET, which is contrary to our previous findings.

A follow-up of this work will be the extension of the method to the case of downlooking observations, such as satellite or aircraft measurements. The main difficulty of such an approach lies in training a NN for an instrument whose set of viewing angles changes from pixel to pixel, as is the case for satellite instruments. In fact, the variability of the observed geometry makes the definition of a feature set to train a NN problematic. This is not a problem for ground-based observations, where the set of viewing geometries adopted by the instrument can always be programmed so as to satisfy the needs of the retrieval algorithm. Solutions aimed at overcoming this difficulty with satellite measurements are under study.

Appendix A: Neural networks: general concepts and terminology

This section summarizes some aspects of the NN theory and language that may be useful for the reader of this paper. For a more complete treatment, the reader can refer to Bishop (1995a) or Haykin (1999).

Given the large variety of neural network models, a general definition of a neural network is difficult to give. In very broad terms, an artificial neural network can be seen as a system or model that is able to change its internal parameters through an interaction with an external environment in order to perform a given task or to accomplish a prescribed behaviour, in a way that resembles the adaptation of the synaptic connections in a biologic neural system. The process of adapting the internal parameters of a NN is what is called “learning”, while the role of the environment is played by the so-called “training set”. The word “network” emphasizes that a NN is usually made of simple computational units interconnected between each other. Such computational units are called “neurons”. A neuron usually applies very simple computations to its input (e.g. multiplying the input by a vector of weights, computing the distance between the input and another vector, computing a simple nonlinear function, etc., depending on the particular type of NN).

Learning in a NN can take place in several ways. In “supervised” learning, the NN is presented with a set of input-output pairs, and the NN parameters are changed based on a measure of the discrepancy between the NN response to an input and the desired output corresponding to that input. In “unsupervised” learning, the NN is only presented with a set of input vectors, and the goal is to differentiate the NN response based on the region of the input space the input vector belongs to, thereby obtaining a clustering of the input space that can be used, e.g. for classification of the input vectors. In “reinforcement” learning, the goal is to train the NN to perform a sequence of actions in response to a given input. Differently from what happens in the supervised learning, the parameters of the NN are not changed based on the correct output for a given input, but based on a “critic”, which is a score used to evaluate the quality of the NN response to the considered input.

Since the NN described in this paper is a supervised one, it is worthwhile providing the reader with further information about supervised learning. As said above, in supervised learning the discrepancy between the NN response to an input vector and the desired response for that vector is used to adjust the parameters of the NN. In practice, a cost function Q , based on the aforementioned discrepancy, is minimized with respect to the vector \mathbf{w} of the NN parameters. Since Q is, in all the practical cases, a nonlinear function, its minimization is carried out iteratively, and leads to an update of the NN parameters in the form

$$\mathbf{w}_{i+1} = \mathbf{w}_i + \alpha_i \mathbf{d}_i, \quad (\text{A1})$$

with $\alpha_i \in \mathbb{R}$ and \mathbf{d}_i a descent direction for Q in a neighbourhood of \mathbf{w}_i . In this case, there always exists an $\bar{\alpha}_i$, such that if $\alpha_i < \bar{\alpha}_i$, then $Q(\mathbf{w}_{i+1}) < Q(\mathbf{w}_i)$. The value of this $\bar{\alpha}_i$ depends on the behaviour of Q around \mathbf{w}_i , and is not known a priori unless second-order information about Q is known. For this reason, learning algorithms that do not use second-order information may occasionally diverge if a step size $\alpha_i > \bar{\alpha}_i$ is chosen.

An important distinction, recalled in Sect. 4, is made between “online” and “offline” (or “batch”) learning algorithms. In online algorithms, the cost function is evaluated (and the NN parameters updated accordingly) after the presentation of each training input. The cost function has the form

$$Q(\mathbf{w}_i, \mathbf{x}_i) = \frac{1}{2} \left\| \mathbf{y}_i(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{t}_i \right\|^2 + \gamma R(\cdot), \quad (\text{A2})$$

where \mathbf{y}_i is the NN response to \mathbf{x}_i with parameter vector \mathbf{w}_i , $\gamma \in \mathbb{R}$ and R is a regularization function that might depend, in principle, on the previous values assumed by the vector of parameters, or on higher order derivatives of the NN output with respect to the input or to the parameter vector. In offline algorithms, the update of the NN parameters takes place only once the entire training set has been presented to the network. The cost function has the form

$$Q(\mathbf{w}) = \frac{1}{2N} \sum_{i=1}^N \left\| \mathbf{y}_i(\mathbf{x}_i, \mathbf{w}_i) - \mathbf{t}_i \right\|^2 + \gamma R(\cdot), \quad (\text{A3})$$

where N is the number of training data. It can be said that online learning algorithms realize a stochastic approximation of the gradient of Q , which individuates the descent direction. Offline algorithms usually have a more stable behaviour and a smoother convergence with respect to online algorithms, thanks to the better estimates of \mathbf{d} they provide. This leads to a faster convergence provided that the training set is not so large. For large and redundant training sets, instead, online algorithms produce reliable estimates of \mathbf{d} , and may converge even faster than offline algorithms thanks to a much higher number of parameter updates per training cycle.

The type of supervised NN used in this paper is known as multilayer perceptron (MLP; Werbos, 1974). In this NN architecture, the neurons are organized in layers, with the outputs of the neurons in the i th layer being the inputs for those in the $(i+1)$ th layer. The first layer of a MLP is called the “input layer”, and does nothing except receive the input vector. The last layer is the “output layer”, and is the layer that provides the ultimate NN output. Between the input and the output layer, one or more “hidden layers” are situated. Each neuron is characterized by a vector of weights \mathbf{w} and a bias b , and its output y is related to its input \mathbf{x} through the relationship

$$y = \varphi(\mathbf{w}^T \mathbf{x} + b), \quad (\text{A4})$$

where φ is usually a C^∞ function called the “activation function” of the neuron. Hidden neurons always have nonlinear

activation functions, the most popular being the hyperbolic tangent and the logistic function. Output neurons, instead, can either have a linear or a nonlinear activation function.

Several algorithms exist to train MLPs, either in online or in offline mode. The most famous online learning algorithm is the standard error backpropagation (Rumelhart et al., 1986), which is obtained by setting $\gamma = 0$ in Eq. (A2) and by setting $\mathbf{d}_i = -\nabla Q(\mathbf{w}_i, \mathbf{x}_i)$ and α_i to a fixed value η (called the “learning rate parameter”) in Eq. (A1). By letting η decrease during the training of a MLP, the convergence of the standard backpropagation can be accelerated (Bös and Amari, 1999), and this solution has been used in this work

(see Sect. 4). Among the most common offline algorithms are the batch backpropagation, obtained in the same way as the standard backpropagation but starting from Eq. (A3), and the scaled conjugate gradient (Møller, 1993) and Levenberg–Marquardt (Hagan and Menhaj, 1994) methods, which select the descent direction \mathbf{d}_i and the step size α_i by approximating the Hessian matrix of Q at each iteration, thereby reducing the risk of generating a locally divergent parameter update. Offline algorithms tend to ensure a faster and more stable convergence than online algorithms provided that the training set is not extremely large. In the latter case, online algorithms may be preferable (Wilson and Martinez, 2000).

Acknowledgements. Antonio Di Noia was funded by the Principal Investigator Preparatory Programme (PIPP) of the Netherlands Space Office (NSO) under project PIPP 11-02.

The groundSPEX instrument was developed and built in the framework of the strategic research programme of the Dutch National Institute for Public Health and the Environment (RIVM).

The CESAR Observatory at Cabauw is operated by the Royal Netherlands Meteorological Institute (KNMI).

To ensure the high quality standards of the AERONET observations, funding is received from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 262254.

The anonymous reviewers who took part in the access review and in the discussion phase are gratefully acknowledged for their constructive and useful comments.

Edited by: A. Kokhanovsky

References

- Aires, F., Rossow, W. B., Scott, N. A., and Chédin, A.: Remote sensing from the infrared atmospheric sounding interferometer instrument 1. Compression, denoising, and first-guess retrieval algorithms, *J. Geophys. Res.*, 107, 4619, doi:10.1029/2001JD000955, 2002.
- Alexander, D.: Volcanic ash in the atmosphere and risks for civil aviation: A study in European crisis management, *Int. J. Disaster Risk Sci.*, 4, 9–19, doi:10.1007/s13753-013-0003-0, 2013.
- Anderson, J. O., Thundiyil, J. G., and Stolbach, A.: Clearing the air: A review of the effects of particulate matter air pollution on human health, *J. Med. Toxicol.*, 8, 166–175, doi:10.1007/s13181-011-0203-1, 2012.
- Antonelli, P., Revercomb, H. E., Sromovsky, L. A., Smith, W. L., Knuteson, R. O., Tobin, D. C., Garcia, R. K., Howell, H. B., Huang, H.-L., and Best, F. A.: A principal component noise filter for high spectral resolution infrared measurements, *J. Geophys. Res.*, 109, D23102, doi:10.1029/2004JD004862, 2004.
- Bishop, C. M.: *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995a.
- Bishop, C. M.: Training with noise is equivalent to Tikhonov regularization, *Neural Comput.*, 7, 108–116, doi:10.1162/neco.1995.7.1.108, 1995b.
- Boesche, E., Stammes, P., Ruhtz, T., Preusker, R., and Fischer, J.: Effect of aerosol microphysical properties on polarization of skylight: sensitivity study and measurements, *Appl. Optics*, 45, 8790–8805, doi:10.1364/AO.45.008790, 2006.
- Bös, S. and Amari, S.: Annealed online learning in multilayer neural networks, in: *On-Line Learning in Neural Networks*, edited by: Saad, D., Chap. 10, 209–229, Cambridge University Press, New York, NY, USA, doi:10.1017/CBO9780511569920.011, 1999.
- Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., Kerminen, V.-M., Kondo, Y., Liao, H., Lohmann, U., Rasch, P., Satheesh, S. K., Sherwood, S., Stevens, B., and Zhang, X. Y.: Clouds and aerosols, in: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by: Stocker, T. F., Qin, D., Plattner, K.-F., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Chap. 5, 571–657, Cambridge University Press, Cambridge, UK and New York, NY, USA, 2013.
- Brajard, J., Jamet, C., Moulin, C., and Thiria, S.: Use of a neuro-variational inversion for retrieving oceanic and atmospheric constituents from satellite ocean colour sensor: Application to absorbing aerosols, *Neural Networks*, 19, 178–185, doi:10.1016/j.neunet.2006.01.015, 2006.
- Brajard, J., Santer, R., Crépon, M., and Thiria, S.: Atmospheric correction of MERIS data for case-2 waters using a neuro-variational inversion, *Remote Sens. Environ.*, 126, 51–61, doi:10.1016/j.rse.2012.07.004, 2012.
- Caruana, R., Lawrence, S., and Giles, C. L.: Overfitting in neural nets: Backpropagation, conjugate gradient, and early stopping, in: *Proceedings of Neural Information Processing Systems (NIPS) Conference 2000*, edited by: Leen, T. K., Dietterich, T. G., and Tresp, V., 402–408, MIT Press, Cambridge, MA, USA, 2001.
- Deuzé, J. L., Goloub, P., Herman, M., Marchand, A., Perry, G., Susana, S., and Tanré, D.: Estimate of the aerosol properties over the ocean with POLDER, *J. Geophys. Res.*, 105, 15329–15346, doi:10.1029/2000JD900148, 2000.
- Deuzé, J. L., Bréon, F., Devaux, C., Goloub, P., Herman, M., Lafrance, B., Maignan, F., Marchand, A., Nadal, F., Perry, G., and Tanré, D.: Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements, *J. Geophys. Res.*, 106, 4913–4926, doi:10.1029/2000JD900364, 2001.
- Diouf, D., Niang, A., Brajard, J., Crépon, M., and Thiria, S.: Retrieving aerosol characteristics and sea-surface chlorophyll from satellite ocean color multi-spectral sensors using a neural-variational method, *Remote Sens. Environ.*, 130, 74–86, doi:10.1016/j.rse.2012.11.002, 2013.
- Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *J. Geophys. Res.*, 105, 20673–20796, doi:10.1029/2000JD900282, 2000.
- Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., and Slutsker, I.: Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, *J. Geophys. Res.*, 105, 9791–9806, doi:10.1029/2000JD900040, 2000.
- Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Muñoz, O., Veihelmann, B., van der Zande, W. J., Leon, J.-F., Sorokin, M., and Slutsker, I.: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, *J. Geophys. Res.*, 111, D11208, doi:10.1029/2005JD006619, 2006.
- Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations, *Atmos. Meas. Tech.*, 4, 975–1018, doi:10.5194/amt-4-975-2011, 2011.
- Dybowski, R. and Roberts, S. J.: Confidence intervals and prediction intervals for feedforward neural networks, in: *Clinical Applications of Artificial Neural Networks*, edited by: Dybowski, R. and Gant, V., Chap. 13, 298–326, Cambridge University Press, Cambridge, UK, doi:10.1017/CBO9780511543494.013, 2001.

- Hagan, M. T. and Menhaj, M. B.: Training feedforward networks with the Marquardt algorithm, *IEEE T. Neural Networ.*, 5, 989–993, doi:10.1109/72.329697, 1994.
- Han, B., Vucetic, S., Braverman, A., and Obradovic, Z.: A statistical complement to deterministic algorithms for the retrieval of aerosol optical thickness from radiance data, *Eng. Appl. Artif. Intel.*, 19, 787–795, doi:10.1016/j.engappai.2006.05.009, 2006.
- Hansen, J., Sato, M., and Ruedy, R.: Radiative forcing and climate response, *J. Geophys. Res.*, 102, 6831–6864, doi:10.1029/96JD03436, 1997.
- Hansen, J. E. and Travis, L. D.: Light scattering in planetary atmospheres, *Space Sci. Rev.*, 16, 527–610, doi:10.1007/BF00168069, 1974.
- Hansen, P. C. and O’Leary, D. P.: The use of the L-curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, 14, 1487–1503, doi:10.1137/0914086, 1993.
- Hasekamp, O. P. and Landgraf, J.: A linearized vector radiative transfer model for atmospheric trace gas retrieval, *J. Quant. Spectrosc. Ra.*, 75, 221–238, doi:10.1016/S0022-4073(01)00247-3, 2002.
- Hasekamp, O. P. and Landgraf, J.: Linearization of vector radiative transfer with respect to aerosol properties and its use in satellite remote sensing, *J. Geophys. Res.*, 110, D04203, doi:10.1029/2004JD005260, 2005.
- Hasekamp, O. P. and Landgraf, J.: Retrieval of aerosol properties over land surfaces: capabilities of multiple-viewing-angle intensity and polarization measurements, *Appl. Optics*, 46, 3332–3344, doi:10.1364/AO.46.003332, 2007.
- Hasekamp, O. P., Litvinov, P., and Butz, A.: Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements, *J. Geophys. Res.*, 116, D14204, doi:10.1029/2010JD015469, 2011.
- Haykin, S.: *Neural Networks: A Comprehensive Foundation*, Prentice Hall, Upper Saddle River, NJ, USA, 1999.
- Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y. J., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, 66, 1–16, doi:10.1016/S0034-4257(98)00031-5, 1998.
- Holmström, L. and Koistinen, P.: Using additive noise in backpropagation training, *IEEE T. Neural Networ.*, 3, 24–38, doi:10.1109/72.105415, 1992.
- Jamet, C., Thiria, S., Moulin, C., and Crépon, M.: Use of a neurovariational inversion for retrieving oceanic and atmospheric constituents from ocean color imagery: A feasibility study, *J. Atmos. Ocean. Tech.*, 22, 460–475, doi:10.1175/JTECH1688.1, 2005.
- Kaufman, Y. J. and Koren, I.: Smoke and pollution aerosol effect on cloud cover, *Science*, 313, 655–658, doi:10.1126/science.1126232, 2006.
- KNMI: OMI News: More North American wildfire smoke observed over Europe, available at: http://www.knmi.nl/omi/news/archive/fullFiles/2013-07-11_smoke_canadian_fires_en.php (last access: 25 November 2014), 2013.
- Knobelspiesse, K., Cairns, B., Ottaviani, M., Ferrare, R., Hair, J., Hostetler, C., Obland, M., Rogers, R., Redemann, J., Shinzuka, Y., Clarke, A., Freitag, S., Howell, S., Kapustin, V., and McNaughton, C.: Combined retrievals of boreal forest fire aerosol properties with a polarimeter and lidar, *Atmos. Chem. Phys.*, 11, 7045–7067, doi:10.5194/acp-11-7045-2011, 2011.
- Lawrence, S., Giles, C. L., and Tsoi, A. C.: Lessons in neural network training: Overfitting may be harder than expected, in: *Proceedings of the Fourteenth National Conference on Artificial Intelligence, AAAI-97*, 540–545, AAAI Press, Providence, RI, USA, 1997.
- LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D.: Backpropagation applied to handwritten ZIP code recognition, *Neural Comput.*, 1, 541–551, doi:10.1162/neco.1989.1.4.541, 1989.
- Liu, X., Easter, R. C., Ghan, S. J., Zaveri, R., Rasch, P., Shi, X., Lamarque, J.-F., Gettelman, A., Morrison, H., Vitt, F., Conley, A., Park, S., Neale, R., Hannay, C., Ekman, A. M. L., Hess, P., Mahowald, N., Collins, W., Iacono, M. J., Bretherton, C. S., Flanner, M. G., and Mitchell, D.: Toward a minimal representation of aerosols in climate models: description and evaluation in the Community Atmosphere Model CAM5, *Geosci. Model Dev.*, 5, 709–739, doi:10.5194/gmd-5-709-2012, 2012.
- Lohmann, U. and Feichter, J.: Global indirect aerosol effects: a review, *Atmos. Chem. Phys.*, 5, 715–737, doi:10.5194/acp-5-715-2005, 2005.
- Mishchenko, M. I. and Travis, L. D.: Satellite retrievals of aerosol properties over the ocean using polarization as well as intensity of reflected sunlight, *J. Geophys. Res.*, 102, 16989–17013, doi:10.1029/96JD02425, 1997.
- Mishchenko, M. I., Cairns, B., Hansen, J. E., Travis, L. D., Burg, R., Kaufman, Y. J., Martins, J. V., and Shettle, E. P.: Monitoring of aerosol forcing of climate from space: Analysis of measurement requirements, *J. Quant. Spectrosc. Ra.*, 88, 149–161, doi:10.1016/j.jqsrt.2004.03.030, 2004.
- Mishchenko, M. I., Cairns, B., Chowdhary, J., Geogdzhayev, I. V., Liu, L., and Travis, L. D.: Remote sensing of tropospheric aerosols from aircraft and satellites, *J. Phys. Conf. Series*, 6, 73–89, doi:10.1088/1742-6596/6/1/005, 2005.
- Møller, M. F.: A scaled conjugate gradient algorithm for fast supervised learning, *Neural Networks*, 6, 525–533, doi:10.1016/S0893-6080(05)80056-5, 1993.
- Monna, W. and Bosveld, F.: In Higher Spheres. 40 years of observations at the Cabauw Site, KNMI Publication 232, De Bilt, the Netherlands, available at: <http://www.knmi.nl/knmi-library/knmiplibmetnummer/knmiplib232.pdf> (last access: 25 November 2014), 2013.
- Niang, A., Badran, F., Moulin, C., Crépon, M., and Thiria, S.: Retrieval of aerosol type and optical thickness over the Mediterranean from SeaWiFS images using an automatic neural classification method, *Remote Sens. Environ.*, 100, 82–94, doi:10.1016/j.rse.2005.10.005, 2006.
- O’Neill, N. T., Eck, T. F., Smirnov, A., Holben, B. N., and Thulasiraman, S.: Spectral discrimination of coarse and fine mode optical depth, *J. Geophys. Res.*, 108, 4559, doi:10.1029/2002JD002975, 2003.
- Picchiani, M., Chini, M., Corradini, S., Merucci, L., Sellitto, P., Del Frate, F., and Stramondo, S.: Volcanic ash detection and retrievals using MODIS data by means of neural networks, *Atmos. Meas. Tech.*, 4, 2619–2631, doi:10.5194/amt-4-2619-2011, 2011.
- Radosavljevic, V., Vucetic, S., and Obradovic, Z.: A data-mining technique for aerosol retrieval across multiple accu-

- racy measures, *IEEE Geosci. Remote Sens. Lett.*, 7, 411–415, doi:10.1109/LGRS.2009.2037720, 2010.
- Ristovski, K., Vucetic, S., and Obradovic, Z.: Uncertainty analysis of neural-network-based aerosol retrieval, *IEEE T. Geosci. Remote*, 50, 409–414, doi:10.1109/TGRS.2011.2166120, 2012.
- Rodgers, C. D.: Characterization and error analysis of profiles derived from remote sounding measurements, *J. Geophys. Res.*, 95, 5587–5595, doi:10.1029/JD095iD05p05587, 1990.
- Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, London, UK, 2000.
- Rosenfeld, D., Wood, R., Donner, L. J., and Sherwood, S. C.: Aerosol Cloud-Mediated Radiative Forcing: Highly Uncertain and Opposite Effects from Shallow and Deep Clouds, in: *Climate Science for Serving Society*, edited by: Asrar, G. R. and Hurrell, J. W., Chap. 5, 105–149, Springer, Dordrecht, the Netherlands, 10.1007/978-94-007-6692-1_5, 2013.
- Rumelhart, D. E., Hinton, G. E., and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, 323, 533–536, doi:10.1038/323533a0, 1986.
- Schiller, H.: Model inversion by parameter fit using NN emulating the forward model – Evaluation of indirect measurements, *Neural Networks*, 20, 479–483, doi:10.1016/j.neunet.2007.04.022, 2007.
- Snik, F., Karalidi, T., and Keller, C. U.: Spectral modulation for full linear polarimetry, *Appl. Optics*, 48, 1337–1346, doi:10.1364/AO.48.001337, 2009.
- Tsekeri, A., Gross, B., Moshary, F., and Ahmed, S.: Potential retrieval of aerosol properties combining the multispectral, multi-angle Research Scanning Polarimeter (RSP) measurements of the intensity and linear polarization of light, in: *Proc. 91st American Meteorological Society Annual Meeting*, Seattle, WA, USA, 2011.
- van Harten, G., de Boer, J., Rietjens, J. H. H., Di Noia, A., Snik, F., Volten, H., Smit, J. M., Hasekamp, O. P., Henzing, J. S., and Keller, C. U.: Atmospheric aerosol characterization with a ground-based SPEX spectropolarimetric instrument, *Atmos. Meas. Tech.*, 7, 4341–4351, doi:10.5194/amt-7-4341-2014, 2014.
- Waquet, F., Cairns, B., Knobelspiesse, K., Chowdhary, J., Travis, L. D., Schmid, B., and Mishchenko, M. I.: Polarimetric remote sensing of aerosols over land, *J. Geophys. Res.*, 114, D01206, doi:10.1029/2008JD010619, 2009.
- Werbos, P. J.: *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, Ph.D. thesis, Harvard University, Cambridge, MA, USA, 1974.
- Wilson, D. R. and Martinez, T. R.: The inefficiency of batch training for large training sets, in: *Proc. IEEE-INNS-ENNS International Joint Conference on Neural Networks (IJCNN '00)*, Vol. 2, 113–117, doi:10.1109/IJCNN.2000.857883, 2000.
- Yu, H., Kaufman, Y. J., Chin, M., Feingold, G., Remer, L. A., Anderson, T. L., Balkanski, Y., Bellouin, N., Boucher, O., Christopher, S., DeCola, P., Kahn, R., Koch, D., Loeb, N., Reddy, M. S., Schulz, M., Takemura, T., and Zhou, M.: A review of measurement-based assessments of the aerosol direct radiative effect and forcing, *Atmos. Chem. Phys.*, 6, 613–666, doi:10.5194/acp-6-613-2006, 2006.