Atmospheric
Measurement
Techniques

Open Access

Satellite retrieval of aerosol microphysical and optical parameters using neural networks: a new methodology applied to the Sahara desert dust peak

M. Taylor[1], S. Kazadzis[1], A. Tsekeri[2], A. Gkikas[3], and V. Amiridis[2]

[1]Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas. Pavlou, Penteli, 15236, Athens, Greece
[2]Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens (NOA), Metaxa & Vas. Pavlou, Penteli, 15236, Athens, Greece
[3]Laboratory of Meteorology, Physics Department, University of Ioannina, Greece

Correspondence to: M. Taylor (patternizer@gmail.com)

Abstract. In order to exploit the full-earth viewing potential of satellite instruments to globally characterise aerosols, new algorithms are required to deduce key microphysical parameters like the particle size distribution and optical parameters associated with scattering and absorption from space remote sensing data. Here, a methodology based on neural networks is developed to retrieve such parameters from satellite inputs and to validate them with ground-based remote sensing data. For key combinations of input variables available from the MODerate resolution Imaging Spectro-radiometer (MODIS) and the Ozone Measuring Instrument (OMI) Level 3 data sets, a grid of 100 feed-forward neural network architectures is produced, each having a different number of neurons and training proportion. The networks are trained with principal components accounting for 98 % of the variance of the inputs together with principal components formed from 38 AErosol RObotic NETwork (AERONET) Level 2.0 (Version 2) retrieved parameters as outputs. Daily averaged, co-located and synchronous data drawn from a cluster of AERONET sites centred on the peak of dust extinction in Northern Africa is used for network training and validation, and the optimal network architecture for each input parameter combination is identified with reference to the lowest mean squared error. The trained networks are then fed with unseen data at the coastal dust site Dakar to test their simulation performance. A neural network (NN), trained with co-located and synchronous satellite inputs comprising three aerosol optical depth measurements at 470, 550 and 660 nm, plus the columnar water vapour (from MODIS) and the modelled absorption aerosol optical depth at 500 nm (from OMI), was able to simultaneously retrieve the daily averaged size distribution, the coarse mode volume, the imaginary part of the complex refractive index, and the spectral single scattering albedo – with moderate precision: correlation coefficients in the range $0.368 \leq R \leq 0.514$. The network failed to recover the spectral behaviour of the real part of the complex refractive index. This new methodological approach appears to offer some potential for moderately accurate daily retrieval of the total volume concentration of the coarse mode of aerosol at the Saharan dust peak in the area of Northern Africa.

1 Introduction

Aerosol particles reflect and absorb solar radiation in the atmosphere shading the earth's surface. They also reduce visibility and can have a direct effect on human health (Samet et al., 2000). Moreover, they are used to determine the earth's hydrological cycle (Remer et al., 2005). However, because of inadequate quantitative knowledge of the global spatial and temporal variation of aerosol optical properties (Hansen et al., 2005), there is uncertainty in the magnitude of their contribution to the earth's climate and planetary radiative-forcing (IPCC, 2007, 2013). With the expansion of the global

AErosol RObotic NETwork (AERONET) of high-quality remote sensing measurement instruments (Holben et al., 1998) and the development of advanced and robust inversion algorithms for the retrieval of aerosol parameters (Dubovik and King, 2000), our understanding of aerosol microphysics and optical properties has improved greatly. However, the size of the uncertainty associated with the aerosol contribution is known to be unacceptably large, and must be reduced by at least a factor of 3 (Schwartz, 2004). An attempt to address this uncertainty has been outlined in a recent report of Mishchenko et al. (2007), which provides the aerosol parameter retrieval accuracy requirements for climate studies. Retrieval of aerosol microphysical properties from inversion of direct sun and sky radiance measurements is provided by AERONET (Dubovik and King, 2000; Dubovik et al., 2002). Unfortunately, these retrievals have low and inhomogeneous spatial resolution (AERONET's ground-based remote sensing instruments are densely situated in and around cities and sparsely located elsewhere). Furthermore, AERONET stations are largely absent from vast uninhabited areas like deserts, oceans and the ice-caps which are the largest sources of planetary aerosol. Marine aerosol retrievals, in particular, are only available at island sites or in coastal regions. In contrast, space-bound satellite instruments like the Moderate Resolution Imaging Spectroradiometer (MODIS) instrument on board the satellites Terra and Aqua, sample the vertical atmospheric column of the whole earth, but their retrieval algorithms are not currently able to provide reliable proxies containing information on the mean particle size of fine and coarse aerosol, the complex refractive index and particle shape – all necessary for a full understanding of aerosol microphysics (Remer et al., 2005) and for globally characterizing different types of aerosols and sources (Tanré et al., 1996). Importantly, a statistically optimized inversion algorithm applied to multi-angle photo-polarimetric measurements has recently demonstrated that aerosol properties are obtainable from the POLarization and Directionality of the Earth's Reflectances (POLDER) instrument on the platforms Advanced Earth Observing Satellite-1 (ADEOS-1) (Deuzé et al., 2000, 2001) and Polarization & Anisotropy of Reflectances for Atmospheric Sciences coupled with Observations from a Lidar (PARASOL) (Dubovik et al., 2011; Hasekamp et al., 2011; Waquet et al., 2014) satellites, but these methods have not yet been independently validated with long data records. In this work, gridded (1 × 1 degree) data from operational MODerate resolution Imaging Spectro-radiometer (MODIS) and the Ozone Measuring Instrument (OMI) instruments was used in order to exploit a long 9-year period of data overlap with AERONET measurements.

1.1 Motivation

This paper focuses on the question of how to retrieve daily estimates of all aerosol parameters from satellite measurements. We assess the potential for achieving this by constructing neural network (NN) models and applying them to data from the region of Northern Africa – where the dust's global aerosol optical depth (AOD) peaks (Chin et al., 2002). This work is motivated then by the potential offered by capitalizing on the full-earth coverage of AOD, H_2O and absorption aerosol optical depth AAOD provided by satellite remote sensing instruments together with AERONET-quality retrievals of the aerosol volume size distribution (AVSD), complex refractive index (CRI), single scattering albedo (SSA) and the particle asymmetry factor (ASYM). The key to building the required bridge between ground and satellite retrievals is to train NNs on AERONET ground-truth data so as to learn the relationship between combinations of satellite AOD, H_2O and AAOD inputs and AERONET microphysical and optical parameters as outputs. The potential of the NNs to extrapolate is then tested by feeding them with unseen satellite inputs and comparing the outputs against co-located and synchronous ground-based AERONET data. In our study, we use the latest AERONET Level 2.0 Version 2 inversion products that are cloud-screened and quality assured (AERONET, 2012).

1.2 Contemporary studies

In the last 5 years or so, multivariate fitting techniques including function-approximating NNs have been brought to bear on problems in the field of aerosol science. Of paramount importance is the finding that a characteristic aerosol fine mode volume and effective radius can be derived from measurements of the AOD, the Ångström Exponent (å) and its curvature using a multi-functional approach (Gobbi et al., 2007). A further study constructed a multiple-input single output NN that took radiances, solar viewing angles, and terrain elevation from MODIS as input, and predicted the values of co-located AERONET AOD values as output (Radosavljevic et al., 2010). The study used data from 221 AERONET sites and demonstrated that AERONET AOD could be successfully estimated from satellite inputs. Taking this further, Ristovski et al. (2012) trained an NN-based estimator of retrieval accuracy which was globally validated on a large sample of co-located MODIS and AERONET AOD retrievals. Complementing this work, Albayrak et al. (2013) used an NN-based approach to perform a global bias adjustment of the MODIS-retrieved AOD relative to co-located AERONET data. NN models were also applied in a very recent study designed to detect and retrieve volcanic-ash-cloud properties from multi-spectral infrared MODIS measurements over Mount Etna during recent volcanic eruptions (Picchiani et al., 2011). In the context of the retrieval of vertical aerosol profiles, Sellitto et al. (2012) used an NN to invert SCanning Imaging Absorption SpectroMeter for Atmospheric CHartographY (SCIAMACHY) data and demonstrated that inclusion of visual radiation could reduce biases and increase the accuracy of ozone profiles at tropospheric levels. These

studies are a sign that the aerosol community is starting to embrace such methods.

AERONET's latest Level 2.0 Version 2 inversion algorithm retrieves all of the aforementioned aerosol microphysical and optical parameters from ground-based sensors by performing a numerical inversion of the observations, which must be performed for each case. On the contrary, the NNs are potentially able to simultaneously retrieve the AVSD, CRI, SSA, and ASYM for the entire data sample in a single step. NN retrieval schemes therefore (potentially) have the capacity to produce real-time retrievals for large data sets. To be more specific, the NN calculates a nonlinear regression function yielding an estimate for the atmospheric state given by the measurement vector (applying to all cases covered by the training space), whereas other methods (like look-up table and optimal estimator methods) match aerosol properties to the corresponding measurement vector. The calculation of this function may require considerable time since, depending on the size of the training data set, NN training can be long, but, once complete, the retrieval using the trained optimal NN is instantaneous. The theoretical basis underpinning the NN function approximation scheme is presented in Sect. 3.1.

1.3 Objectives

Motivated by the need to develop a methodology to produce global satellite retrievals of aerosol microphysical and optical parameters, and inspired by the success of recent NN models, this paper reports on the initial phase of AEROMAP (http: apcg.meteo.noa.gr/aeromap) a 2-year EU-funded project that began in March 2012. This, our first major study, has the following main objectives:

1. to assess the potential of performing aerosol typing by using Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model outputs to select suitable desert dust sites at the peak of dust extinction in Northern Africa,

2. to see if it is possible to standardize and optimize NN architectures capable of learning the relationship between the inputs and outputs for this region (i.e. for this aerosol type),

3. to validate the trained NNs with unseen data at a distant geolocation in the same region (i.e. aerosol type) and to assess their performance using statistical regression and timescale analysis.

1.4 Structure of the paper

The data used and an outline of the NN model are presented in Sect. 2. Section 3 then presents the theory involved in training and validating such NNs. In Sect. 4, the results of NN training and testing for different input configurations are presented and key findings, major impacts, as well as pros and cons of the method are noted and analyzed in Sect. 5.

Finally, we conclude in Sect. 6 by assessing the overall potential offered by the NN methodology for retrieving aerosol microphysical and optical parameters from space.

2 Methodology

Aerosol particles from different sources have different sizes, absorption properties, and shape. They are typically classified into a small number of types (≈ 5–10) including for example: desert or soil dust, smoke or organic and black carbon from biomass burning, urban sulphates, marine sea salt, volcanic ash as well as their mixtures. Researchers in the field have found that different aerosol types correlate strongly with pairs of different aerosol parameters, but no consensus has yet been reached on a single method to disambiguate and universally distinguish them. Therefore, in this work, in order to avoid as much as possible such potential sources of data inhomogeneity or inconsistency, we adopted an independent qualitative approach to aerosol typing which is described in Sect. 2.1.3.

2.1 Data selection

This work draws on 4 different data sources: satellite inputs from MODIS and the OMI, ground-based remote sensing data from AERONET, and global chemical model output data from the Georgia Institute of Technology's GOCART model (Chin et al., 2000, 2002; Ginoux et al., 2001). MODIS and OMI provide satellite inputs and co-located and synchronous values of these inputs as well as output parameters at the ground are provided by AERONET. The GOCART data is used for aerosol typing.

2.1.1 Satellite inputs

MODIS on board the Terra Earth Observation Satellite (EOS) (EOS-AM) and Aqua (EOS-PM) satellites has been capturing data in 36 spectral bands from 400 to 1440 nm since 1999 with a spatial resolution ranging from 250 m– 1 km. Collectively, the instruments image the entire earth's surface every 1–2 days. Daily averaged data was downloaded in hierarchical data format from the MODIS Level 3 Collection 5.1 Product (MODIS, 2012). From these files, AOD(470), AOD(550), AOD(660) time series provided at 1×1 degree spatial resolution were extracted. In addition, co-located and synchronous, Level 2, near-infrared, mean total columnar water vapour (H_2O) from the Aqua satellite (data set MYD05_L2) was also downloaded. Finally, the daily estimate of near-ultraviolet (UV) AAOD(500) was downloaded from the OMI Level 3 Near-UV Aerosol Data Product (OMAERUV) Product (OMI, 2012) for co-located and synchronous data (with MODIS) to test the impact of absorption on NN retrieval quality. As a result, daily averages of these parameters were obtained for the entire global
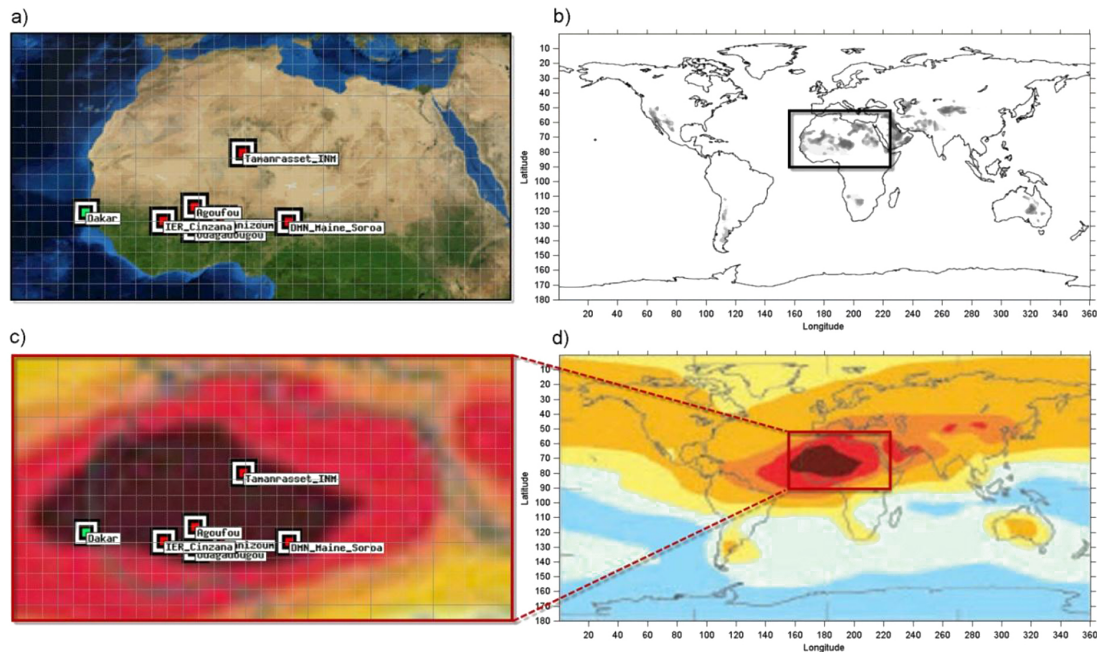
Figure 1. Schematic showing **(a)** the Northern African (NAF) AERONET sites used for NN training (red) and the coastal AERONET site at Dakar (green) used for simulation with data set A, **(b)** the NAF study region in the context of the global distribution of TOMS dust sources (Prospero et al., 2002), **(c)** an overlay of the AERONET sites on the peak of dust AOD extinction for the study region extracted from the mean global GOCART model output in shown in **(d)** (Chin et al., 2000, 2002).

domain, spanning the full temporal record of available data: 4 July 2002 to 4 July 2012.

2.1.2 AERONET products

The AERONET Level 2.0 Version 2 inversion products contain retrievals for 116 different aerosol parameters including the AVSD: $dV(r)/d\ln r$ (in $\mu m^3 \mu m^{-2}$) retrieved in 22 logarithmically equidistant radial bins spanning the range of particle radii: $0.05 \mu m \leq r \leq 15 \mu m$, the real and imaginary parts of the refractive index: CRI-R(λ), CRI-I(λ), and the optical parameters: AOD(λ), SSA(λ), and ASYM(λ) centered at 4 wavelengths: $\lambda = 440$, 675, 870 and 1020 nm. Daily averaged retrievals were downloaded for the entire global AERONET record (comprising 809 sites) and spanning the period: 1 March 1996 to 7 April 2012. For each site, its elevation (height above sea level in metres), its Eastern longitude and Northern latitude were extracted. In addition, although AERONET's Level 2.0 Version 2 inversion products also provide the mean geometric radii of the fine and coarse modes $r(f)$ and $r(c)$, their standard deviations $\sigma(f)$ and $\sigma(c)$, and their volume concentrations $V(f)$ and $V(c)$; the fine fraction η which is not provided was also calculated and appended to the AERONET data record. All of these parameters are calculated from the AVSD by specifying a mode separation point rs, and, in what follows, we will refer to them collectively as secondary microphysical parameters. Their calculation (required for comparing

satellite-driven NN simulated outputs with AERONET) is described briefly in Appendix A. Furthermore, it has been found that there is a (small) difference between AODs obtained by MODIS and AERONET which is important and non-negligible (Remer et al., 2005; Albayrak et al., 2013) Hence, the Ångström Exponent å (675 nm/440 nm) was calculated and used to extrapolate AERONET AODs to match those available from space at MODIS wavelengths with the following rearrangement:

$$\text{AOD}(\lambda_2) = \text{AOD}(\lambda_1)\left(\frac{\lambda_2}{\lambda_1}\right)^{-\text{å}(\lambda_2,\lambda_1)}. \tag{1}$$

These interpolated AERONET AOD(470), AOD(550), AOD(660) values were also appended to the AERONET data set. This data set therefore contains both ground retrievals of the satellite inputs (aligned to the central wavelengths provided by MODIS) plus the output parameters which the NN model is built to retrieve.

2.1.3 Aerosol typing

In order to isolate suitable desert dust data for this study, we developed a qualitative two-step approach. In the first step, the AERONET data set was ranked by the number of complete records available at each site (without data gaps in the input parameters AODs, H$_2$O, AAOD, and the output parameters AVSD, CRI-R, CRI-I, SSA and ASYM). The requirement for records to be complete caused the number

of sites to drop from 807 to 623 sites. It was found, for example, that the top-ranked site in the study region (Northern Africa) is Banizoumbou (Niger) which contains 2283 complete data records. The second step of our approach aims to answer the question: how many of these daily averaged records are dust dominated? For this, the GOCART model AOD extinction per aerosol type was used. GOCART provides 3-hourly measurements of the total extinction AOD as well as the contribution to total extinction AOD of sulphate (SO_2), organic carbon (OC), black carbon (BC), desert (mineral) dust and sea salt. GOCART data was downloaded for the first 155 AERONET sites ranked in step 1 by the number of complete records. This list accounts for 75 % of all available Level 2.0 Version 2 inversions. GOCART provides eight 3-hourly measurements per day, which were averaged to produce daily averages and expressed as a percentage of the total extinction AOD for each aerosol type. The percentage of dust was then used as a basis for re-ranking the list of high data volume sites. Table 1 shows the AERONET complete-record ranked-sites, ranked by dust contribution (according to GO-CART data) for the study region (Northern Africa).

In Table 1, data set A comprises AERONET sites that operate the older CIMEL model I sun photometers which lie on the peak of dust AOD extinction as extracted from the mean global GOCART model output, and which are verified via cross reference with the strongest TOMS dust sources (shown in Fig. 1). Data set B comprises those sites that operate the newer CIMEL model II sun photometers which, in addition, also contribute measurements of near-UV AOD at 380 and 500 nm. This separation of the Northern Africa data was made so as to investigate the possible effect of UV AOD inputs on NN model performance. Dakar was selected as the testing site since: (i) it has the largest number of days of co-located synchronous satellite measurements, (ii) it is also located on the peak of dust AOD extinction, and (iii) it operates the newer model II CIMEL sun photometer.

2.1.4 Handling of outliers

While it is generally not good practice to remove outliers since they often correspond to interesting phenomena, in relation to NNs, it is important that infrequently occurring, extreme data that can significantly bias the data-fitting procedure is removed. This led us to investigate various methods of outlier detection and to study the distribution of the data for each of the input and output parameters. Histograms were produced that partitioned the data into 20 bins and it was found that many of the parameters presented near-normal distributions in quantile–quantile plots (the H_2O, the volume concentration in each radial bin, the CRI-R and the ASYM), but that AODs and the CRI-I presented positive skew-normal distributions, and the SSA presented negative skew-normal distributions. We elected to apply the Grubbs' test (Grubbs, 1969) to remove outliers. Grubb's Test consists of testing one data point at a time and finding and removing

the value furthest from the sample mean (usually applied to normally distributed data). Since the median is more statistically robust when analyzing data that is skew-normal, Grubb's Test was applied with reference to the sample median rather than the sample mean. This procedure was applied iteratively to data sets A and B (used to train the NNs) until outliers were removed at the 68 % confidence level of the entire two-tailed data distribution. Outliers were deliberately not removed from the inputs used in testing the NN so that the ability of the NNs to extrapolate on raw, unseen data could be properly tested. The data selection scheme produced dust-typed input output data that (a) is homogeneous (does not contain parameter data gaps), (b) is wavelength-matched and (c) is free of biasing values (at the 68 % level of confidence).

2.2 The NN model

Feed-forward NNs having at least one layer of hidden neurons whose activation functions, are nonlinear hyperbolic tangent (Tanh) functions (or other general nonlinear sigmoidal functions), are able to operate as universal function approximators (Cybenko, 1989; Hornik et al., 1989). This means that, given enough hidden neurons and training data, such networks are capable, in principle, of learning the mathematical relation between inputs and outputs. The input and output parameters used in this work were connected via two network layers – the first layer containing hidden neurons with Tanh activation functions and the second layer containing output neurons having linear activation functions. We also tested three-layer models that used two layers of hidden neurons but the results were worse than those obtained here. The relation between input and output parameters for the type of NN used in this study is presented in Sect. 3.1, together with details of the methodology adopted for evaluating network training (Sect. 3.2) and network validation (Sect. 3.3). Here, we describe the operation of the NN model which was coded using MATLAB's object-oriented scripting language in conjunction with its neural network toolbox (Demuth and Beale, 2004).

NN models require specification of (1) how the performance error associated with the network model is to be measured and (2) the architecture used. We measure the performance error of the network using the mean squared error (MSE) calculated from the difference between its outputs and target output data. The details of the macro-statistical approach we adopt are presented in Sect. 3.2 in the context of NN training. The NN architecture is a more complex entity. It involves not only the number of hidden neurons and their activation functions, but also the proportion of data used to train and validate the NN as well as the learning algorithm used. The perception that NN models are somewhat subjective is due to what is often seen as an arbitrary choice of some or all of these elements. In order to try to make the choice of architecture more objective, we developed a new procedure to

Table 1. Selection of desert-dust dominated AERONET sites for this work. N is the number of complete AERONET daily averaged Level 2.0 Version 2 inversion records available (up to 7 April 2012). For each site, the total mean extinction AOD and the percentage composition of the total is given for GOCART-modelled aerosol types.

AEROSOL TYPING	SITE	N	GOCART Mean AOD & aerosol composition					
			$<$ AOD $>$	% SO$_2$	% OC	% BC	% Sea Salt	% Dust
TRAINING	Tamanrasset INM	407	0.793	4.54 %	1.39 %	0.63 %	0.13 %	93.44 %
(Data set A)	Agoufou	1028	0.973	3.70 %	2.47 %	0.82 %	0.10 %	92.91 %
	Banizoumbou	2283	0.920	4.57 %	3.48 %	1.09 %	0.11 %	90.76 %
	DMN Maine Soroa	680	0.967	5.27 %	3.52 %	1.14 %	0.10 %	90.07 %
	IER Cinzana	1469	0.823	4.86 %	4.62 %	1.22 %	0.12 %	89.19 %
	Ouagadougou	966	0.776	6.06 %	7.47 %	1.93 %	0.13 %	84.41 %
TRAINING	Niamey	310	0.920	4.57 %	3.48 %	1.09 %	0.11 %	90.76 %
(Data set B)	IER Cinzana	1469	0.823	4.86 %	4.62 %	1.22 %	0.12 %	89.19 %
	Dahkla	299	0.629	8.43 %	1.91 %	0.79 %	0.95 %	88.08 %
	Santa Cruz Tenerife	660	0.405	15.06 %	2.96 %	1.23 %	4.20 %	76.79 %
	Izana	563	0.358	17.32 %	3.07 %	1.40 %	5.59 %	72.63 %
SIMULATION	Dakar	1583	0.705	7.38 %	5.53 %	1.42 %	0.71 %	84.82 %

detect optimal NN architectures. We began by creating a list of candidate input–output combinations (see below). Then, we trained the corresponding NNs by following these four steps:

1. normalize all input and output variables,

2. apply principal components analysis (PCA) to inputs and outputs separately so as to exclude redundant variability (it is required that the PCs account for 98 % of the total variance),

3. loop through a grid of 100 NNs of varying numbers of hidden neurons (4–24 in steps of 2) and proportions of training data (40–90 % in steps of 5 %),

4. select the NN that has the minimum total training and validation MSE.

This procedure can be automated and was found to avoid the bias and underfitting that can result from having too few neurons on the one hand, and the high variance and overfitting that can result from having too many on the other (see Sect. 3.2). It also avoids arbitrary partitioning of the data into training and validation proportions, and the use of PCA helps exclude redundant variability which can adversely affect training efficiency (Jolliffe, 2002). Normalization of the input and output variables was achieved as follows. For each input and output variable data vector X, we calculated the mean μ_X and standard deviation σ_X. The vector means and standard deviations were then used to map (or shift and scale) the input and output data vectors onto their z score values: $z_X = (X - \mu_X)/\sigma_X$ (i.e. standard normal values having a mean $= 0$ and a standard deviation $= 1$). In this study, we consider the min–max values to be those

available in our training data set, and as characteristic of dust in the Northern Africa region. The application of PCA in our study was done to reduce the redundancy in the input and output variables. PCA is an effective procedure for removing this redundancy and has two effects: it orthogonalizes the components of the data vectors (so that they are uncorrelated with each other), and it orders the resulting orthogonal components (principal components or PCs) so that those with the largest variation come first – allowing us to eliminate the components that contribute the least to the variation in the data set. The application of PCA requires normalization of the variables prior to application of the method due to the fact that different variables have very different value ranges and bias the measurement of the variance (Abdi and Williams, 2010). PCA was applied separately to the input and output variables and the extracted PCs were ordered. Best results were obtained by retaining the top ranked PCs that accounted for 98 % of the total variation in the input and output data. The components calculated from PCA are a mixture of the original variables. We also did some trials applying PCA on groups of variables of the same type (e.g. AVSD bins and spectral parameters separately) so as to retain physical characteristics within variable clusters – but the results were worse than those presented here. We wish to emphasize that the methodology presented here is a first attempt at objectivizing the choice of NN architecture, and is not ideal. For example, the discrete steps in (neuron, proportion of training data [%]) space could be made finer (i.e. instead of steps of 5 % in the training proportion (in what follows, we denote the proportion of training and validation data used as "training %" and "validation %", respectively) we could have used a 1 % step size). In addition, a bootstrapping approach could be adopted that would allow several different instances
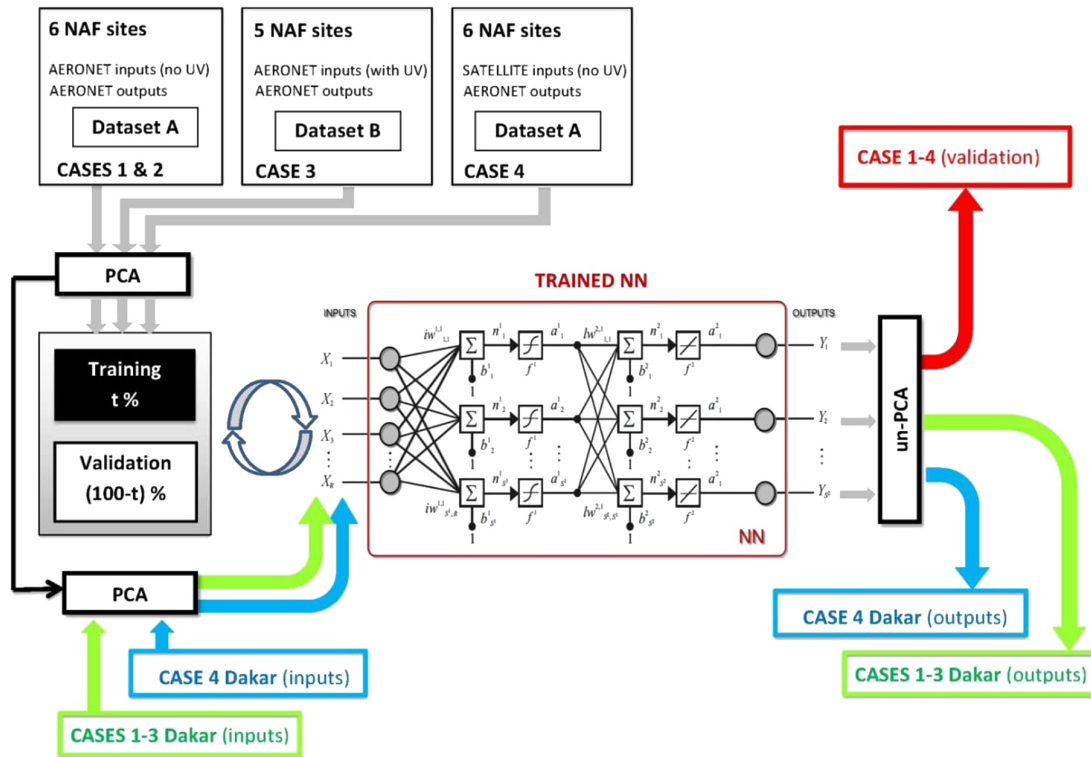
Figure 2. Schematic of the NN model used in this work. Principal components obtained from PCA applied to the case 1–4 data are formed and used to train the central engine NN shown in the centre. The training cycle is repeated for the grid of NNs and the optimal trained NN is found. The outputs of the trained NN are then transformed back to the full parameter space using the reverse principal components (un-PCA). The outputs from the trained NN are used to validate the interpolation potential of the optimal NN. Principal components obtained during the data pre-processing step of network training are used to transform new case 1–4 inputs at Dakar which are fed to the trained NN to simulate case 1–4 outputs at Dakar.

at the same training % / validation % ratio to be evaluated. It should also be noted that it is customary to optimize an NN on the validation data rather than the training data (Bishop, 1995). This was also our initial approach. However we found that the performance of the resultant NN on unseen data at the test site (see Sect. 4) was maximized when we coupled the training and validation MSE. We recognize that the NN is not built to work in the general case (i.e. to retrieve dust properties worldwide), but it works well for the Northern Africa region where we performed our study. We hope to address the generalization problem in a future publication. For a thorough description of data handling in the context of constructing and testing function approximating NNs, we refer the reader to Bishop (1995).

Aiming to perform an empirical sensitivity analysis with respect to candidate input combinations, we drew up a list of aerosol parameters which are provided by the two satellites globally at 1×1 degree spatial resolution, leading to the following set: AOD(470), AOD(550), AOD(660) and H_2O from MODIS, and AOD(380), AOD(500) and AAOD (500) from OMI. Since it has been suggested that there is high sensitivity to particle absorption in the near-UV (Torres et al., 2002),

it was decided that this effect would be studied separately by constructing an input combination that depended on the near-UV AOD at 380 and 500 nm – which are provided by the new CIMEL (model II) AERONET sun photometers comprising data set B. Note that the AAOD(500) provided by OMI is a modelled parameter obtained by using a look-up table of expected SSA values that depend on the aerosol type and the geographical location (Torres et al., 2007). Conversely, in the case of AERONET, the value of AAOD (at the central wavelengths: 440, 675, 870 and 1020 nm) is calculated from retrieved aerosol microphysical properties (Dubovik et al., 2000). In all, the following four distinct scenarios were identified and used in this study:

– Case 1 inputs = AERONET: AOD(470), AOD(550), AOD(660)

– Case 2 inputs = AERONET: AOD(470), AOD(550), AOD(660) + H_2O

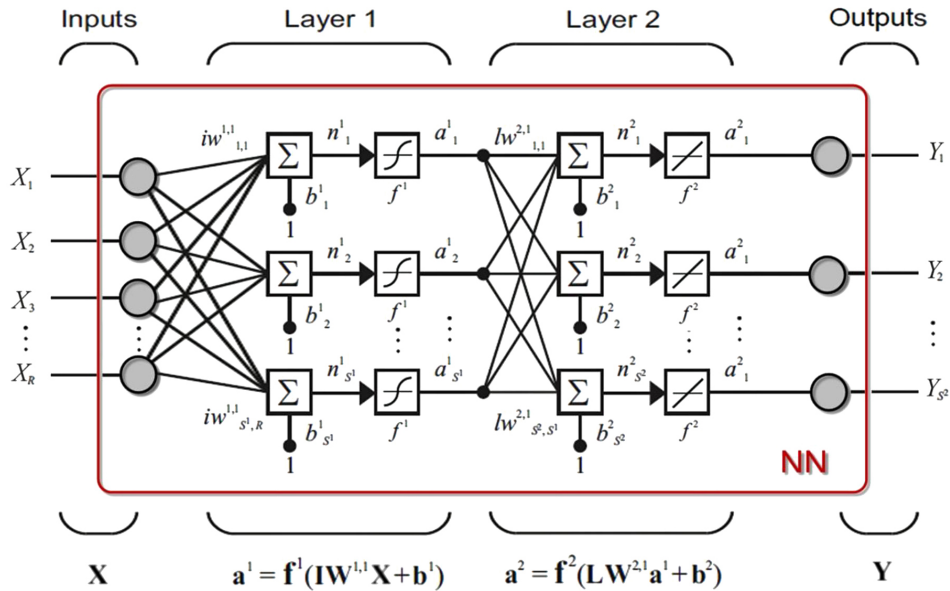– Case 3 inputs = AERONET: AOD(470), AOD(550), AOD(660) + H_2O + AOD(380), AOD(500)

Figure 3. Schematic showing the neural connectivity between input and output parameters.

– Case 4 inputs = MODIS: AOD(470), AOD(550), AOD(660) + H_2O and OMI: AAOD(500).

This approach is essentially a form of empirical sensitivity analysis applied to the input data. In each case, the set of output variables comprises: the AERONET microphysical AVSD (calculated at 22 equidistant logarithmic radial bins spanning the range 0.05 to 15 µm), the spectral refractive index and the optical parameters SSA and ASYM centred at 440, 675, 870 and 1020 nm. Cases 1 and 2 use daily averaged records drawn from data set A, case 3 uses daily averaged records drawn from data set B and case 4 uses co-located satellite data synchronous with data set A (see Table 1). The NN model then proceeds as follows. PCA is applied to the input and output data separately for each of the cases 1–4 and a grid of 100 NNs of differing (hidden neuron, training %) architecture is produced, trained and validated. The optimal NN is then identified using the minimum total training and validation MSE between the NN outputs and target AERONET data. The PCA is inverted back to parameter space and comparative (linear regression) statistics are calculated for the outputs of the optimal trained NN in relation to the AERONET training output data. In order to test each optimally trained NN, new and unseen case 1–4 data at the coastal dust site Dakar is transformed into PCA space and fed to the corresponding NNs. In each case, the network's output is transformed back from PCA space to parameter space where comparative statistics are again applied to the NN outputs in relation to AERONET ground-truth data.

A schematic of the overall NN model is shown in Fig. 2. In Sect. 3, the functional relation between network outputs and inputs is presented together with details of the methods used to train and validate the performance of the NNs.

3 Theory

3.1 The NN input–output function approximation

As we discussed in Sect. 2.2, the motor behind the NN model is the multiple input, multiple output two-layer feed-forward NN at the centre of Fig. 2. The NN has the following input–hidden layer–output layer connectivity shown in detail in Fig. 3.

The NN has a vector X of R-input PCs and a vector Y of s^2-output PCs (grey circles). For case 4 for example, PCA applied to the inputs generated $R = 3$ PCs, and PCA applied to the outputs produced $s^2 = 7$ PCs (see Sect. 2.2 for details). The NN has 2 layers of neurons connecting the inputs to the outputs. The first layer (the hidden layer) has s^1 neurons with nonlinear activation functions $f^1 = $ Tanh and the output layer has s^2 neurons with linear activation functions f^2. Each neuron has a single bias $[0, 1]$ and so the hidden layer has a vector b^1 of s^1 biases while the output layer has a vector b^2 of s^2 biases. The vector of R-inputs X is connected to the s^1-neurons of the hidden layer via a matrix of $[s^1 x R]$ input weights $\mathbf{IW}^{1,1}$ while the vector a^1 of s^1-outputs is connected to the s^2-output neurons via a matrix of $[s^2 x s^1]$ layer weights $\mathbf{LW}^{2,1}$. Finally, the vector a^2 of s^2-outputs is the vector Y of NN outputs. The exact mathematical equation relating the NN outputs to the NN inputs is then the matrix equation:

$$Y = f^2\left(\mathbf{LW}^{2,1} f^1\left(\mathbf{IW}^{1,1} X + b^1\right) + b^2\right). \tag{2}$$

The multiplication of the matrix $\mathbf{IW}^{1,1}$ and the vector X is a dot product equivalent to the summation of all input connections to each neuron in the hidden layer. Equation (2)

above is the nonlinear functional approximation N that relates the output parameters to the input parameters:

$$\mathbf{Y} = N(\mathbf{X}). \tag{3}$$

As we described in Sect. 2.2, the input vector X contains a combination of the satellite input parameters while the output vector Y contains the sought-after retrievals. Traditionally, an NN is assessed by dividing available data into 3 proportions: a training set, a validation set, and a testing set. However, since the data reduction scheme described in Sect. 2 led to a substantial loss of available data records, it was decided that all available data should be put to use in NN training and validation, with none reserved for testing. During the testing phase, the NNs therefore are presented with unseen (not used for the NN training) input data at a new site (Dakar) in the same region (Northern Africa), and used to simulate the outputs – i.e. they are blind to the expected outputs. In this way, all available aerosol-typed data for the region of interest is used (apart from Dakar) in the training and validation process and testing is able to shine light on the potential of the trained and validated NNs to work properly with unseen data. The results of the NN training and validation phase are presented below. In Sect. 4 the results of the NN testing phase are presented.

3.2 NN training

In the training phase batch runs are performed on a grid of 100 NNs permuting through a range of architectures such that the number of hidden neurons ranged from 4–24 (in steps of 2) and so that the training proportion ranged from 40 to 90 % (in steps of 5 %). The NN connection weights and biases are updated (i.e. trained) using an optimization learning algorithm. Initial tests were made with both a single layer of hidden neurons and also with two layers of hidden neurons. For each of these tests, 4 different optimization learning algorithms were also investigated: (i) the Levenberg–Marquardt (LM) back-propagation optimisation learning algorithm (Levenberg, 1944; Marquardt, 1963) (MATLAB flag "trainlm"), (ii) Bayesian regularization (MATLAB flag "trainbr"), (iii) resilient back-propagation (MATLAB flag "trainrp"), and (iv) scaled conjugate-gradient back-propagation (MATLAB flag "trainscg"). The best results were obtained with the LM algorithm applied to a single layer of hidden Tanh neurons. During each iteration of the learning process, the weights and biases are tuned so as to minimize the MSE cost function:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{t}_i - y_i)^2. \tag{4}$$

Note that the MSE is calculated from N output vectors y_i against N AERONET target vectors $\mathbf{t}_i$. Training proceeds through a number of epochs until the MSE between NN outputs and AERONET targets (expected outputs) is minimised. In particular, the MSE obtained from the training data and

the MSE obtained from the validation data were summed for each NN in the grid. The optimal NN was identified as the one whose architecture had the smallest total MSE.

Table 2 shows the results of applying this optimisation process to cases 1–4. One thing to note from Table 2 is that the training error in cases 1–4 is substantially larger than the validation error (having percentage fractional errors of +15.8, +17.2, +27.8 and +12.1 %, respectively). This can be due to outliers in the data set, although we attempted to implement a strict quality filter via aerosol typing and the exclusion of outliers at the 68 % level of confidence with Grubb's Test. While the percentage fractional error does not appear to depend on the size of the sample (the case 1–4 NNs have $N = 3808, 3808, 353$ and 213 training data records, respectively), we cannot exclude the possibility (even in cases 2–4) that there may be data vectors that are associated with input–output values that occur less frequently and which are therefore not learned well by the NN. The second thing to note is that for the case 1 NN, convergence was achieved very rapidly (2 epochs), suggesting, as expected, that the input vector is clearly not containing the information needed to recover the target vector.

The optimal case 4 NN, trained with data from satellite inputs and outputs from the AERONET stations comprising data set A, has 22 neurons in the hidden layer, 7 neurons in the output layer, and used 90 % of data set A for training and 10 % for validation. This NN has three inputs: the three principal components (PCs) of AOD(470), AOD(550), AOD(660), H_2O and AAOD(500). It also has seven outputs: 7 PCs of the 22 logarithmically equidistant radial bins of the AVSD and the CRI-R, CRI-I, SSA and ASYM spectral parameters centred at 440, 675, 870 and 1020 nm. The evolution of the optimization process as well as the statistics associated with this optimal case 4 NN found are shown in Fig. 4.

Figure 4a shows, as expected, that the training MSE tends to decrease as the number of hidden neurons is increased. Furthermore, it shows that as the number of neurons increases, a positive gradient emerges in the training MSE with training % (most clearly visible in the lower panel of Fig. 4a when the number of neurons is greater than about 12 neurons) – i.e. for a fixed number of neurons the training MSE is increasing with training %. While this may be somewhat counter-intuitive, it is possible that by increasing the training data sample, we increase the likelihood of including a couple of records from the long tail of the parameter distributions which are not easily retrieved, resulting in larger MSEs. Figure 4b shows that the validation MSE increases slowly with the number of hidden Tanh neurons. Figure 4b shows that the validation MSE increases slowly with the number of hidden Tanh neurons. Two sharp peaks at (10, 60 %) and (20, 45 %) are probably due to the fact that over-fitting is occurring at these points due to the small size of the data set. The total training time is seen in Fig. 4c to increase sharply and non-linearly when the number of neurons is > 22. In relation to the evolution of NN performance with epoch in Fig. 4d,

Table 2. In this study, four distinct optimal NN architectures were constructed corresponding to cases 1–4. NNs for cases 1–3 are trained on AERONET-only inputs and outputs. Case 4 is trained on satellite inputs and AERONET outputs.

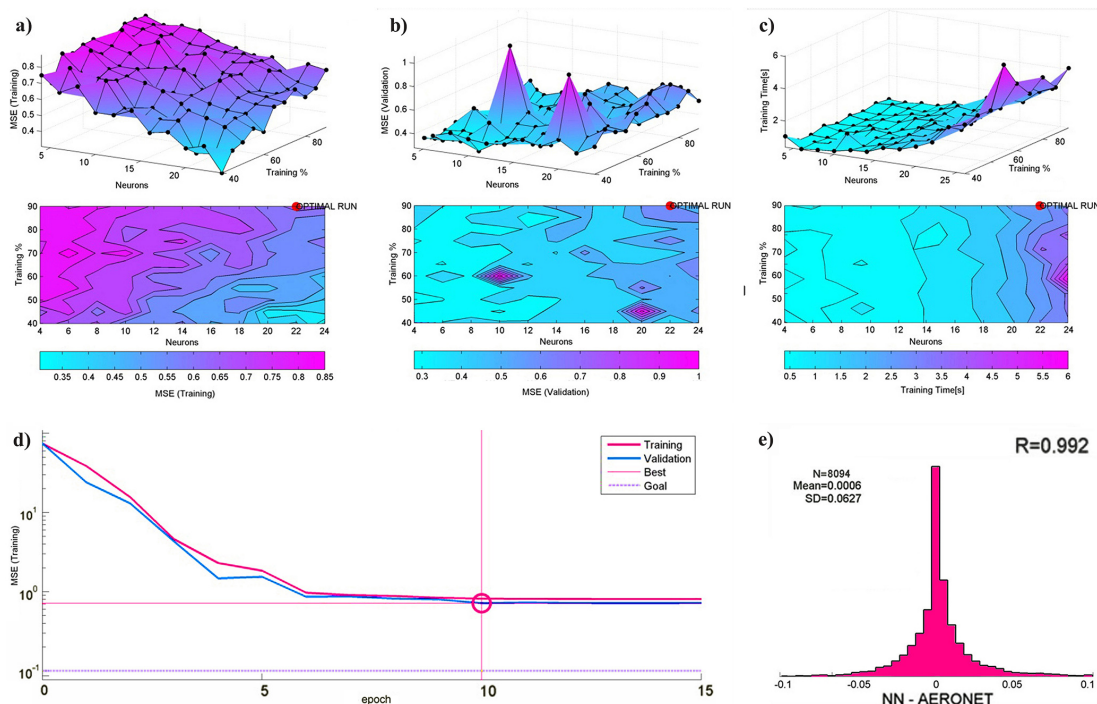| NN OPTIMISATION	PARAMETER	AERONET			SATELLITE
		Case 1	Case 2	Case 3	Case 4
INPUTS	AOD (Visible)	AERONET AOD(470)	AERONET AOD(470)	AERONET AOD(470)	MODIS AOD(470)
		AERONET AOD(550)	AERONET AOD(550)	AERONET AOD(550)	MODIS AOD(550)
		AERONET AOD(660)	AERONET AOD(660)	AERONET AOD(660)	MODIS AOD(660)
	Columnar Water Vapour		AERONET H$_2$O	AERONET H$_2$O	MODIS H$_2$O
	Absorption AOD			AERONET AOD(380)	OMI AAOD(500)
	AOD (UV)			AERONET AOD(500)	
OUTPUTS	Microphysics	AVSD(22 bins): 0.05–15 μm	AVSD(22 bins): 0.05–15 μm	AVSD(22 bins): 0.05–15 μm	AVSD(22 bins): 0.05–15 μm
		CRI-R(440, 675, 870, 1020)	CRI-R(440, 675, 870, 1020)	CRI-R(440, 675, 870, 1020)	CRI-R(440, 675, 870, 1020)
		CRI-I(440, 675, 870, 1020)	CRI-I(440, 675, 870, 1020)	CRI-I(440, 675, 870, 1020)	CRI-I(440, 675, 870, 1020)
	Optics	SSA(440, 675, 870, 1020)	SSA(440, 675, 870, 1020)	SSA(440, 675, 870, 1020)	SSA(440, 675, 870, 1020)
		ASYM(440, 675, 870, 1020)	ASYM(440, 675, 870, 1020)	ASYM(440, 675, 870, 1020)	ASYM(440, 675, 870, 1020)
	Training data set	A	A	B	A
	X = PCA (inputs): 98 % variance	1	2	2	3
	Y = PCA (outputs): 98 % variance	7	7	6	7
ARCHITECTURE	Number of Layers	2	2	2	2
	Back-propagation Method	LM	LM	LM	LM
	Cost Function	MSE	MSE	MSE	MSE
	Activation Functions (Layer 1/Layer 2)	Tanh/Linear	Tanh/Linear	Tanh/Linear	Tanh/Linear
	Optimal Neurons (Layer 1/Layer 2)	10/7	14/7	10/6	22/7
	Optimal Train %	90%	85 %	85 %	90%
	N (samples)	3808	3808	353	213
	N (training)	3427	3237	300	181
	N (validation)	381	571	53	32
	Train time [s]	539	1079	86	100
OUTCOME	Best Epoch	2	13	24	10
	Best training MSE	0.848	0.760	0.765	0.818
	Best validation MSE	0.714	0.629	0.552	0.719
	Pearson R ($Y = aX + b$)	0.998	0.998	0.998	0.992
	a	0.994	0.995	0.994	0.985
	b	0.000	0.000	0.000	0.005

Figure 4. Optimization of the NN for case 4. The upper panels show the training MSE surface (left), the validation MSE surface (middle), and the total training time surface [s] (right) for the grid of 100 NNs. The MSE of the training data and validation data (100-training %) with back-propagation iteration (epoch) is shown for the optimal NN (22, 90 %) in the lower left panel, while the errors calculated from the difference between the NN PC outputs and the AERONET PC outputs for the same NN together with the value of their regression coefficient R, is shown in the lower right panel.

convergence has clearly been reached after 10 epochs (iterations) at the horizontal asymptote where the best validation MSE = 0.719. For all NNs, the goal for the back-propagation cost function is set to 1/100th of the variance of the targets (for the optimal case 4 NN this is equal to 0.12). In this case, the goal is very stringent and is unlikely to be reached with an increase in the number of iterations – suggesting that a much larger and uniform training data set is required to improve the training performance further. We base our interpretations in this work mostly on macro-scale statistics so as not to distract from the main goal of the study. We will consider intrinsic NN errors and uncertainty in more detail in a future paper. The Pearson product–moment correlation coefficient calculated from NN PC outputs and AERONET training PC targets for the optimal case 4 NN is $R = 0.992$ (see Fig. 4e) and is suggestive of an excellent NN fit. This is further backed up by the histogram of the differences between NN PC outputs and AERONET training PC targets (Fig. 4e) which presents a sharply peaked Gaussian having a near-zero mean error = 0.0006 and a standard deviation (SD) = 0.0627. These macro-statistics suggest that the optimal NN is generally well trained and properly performs the function approximation between inputs and outputs. More transparency can be gained by performing comparative macro-statistics on the output parameters separately, as described in the next section.

3.3 NN validation

The results of NN training along with the training data size for each of the cases 1–4, are shown in Table 3. The columns "Target" and "NN output" present the mean value of each parameter. In Table 3, the daily averaged coarse mode peak is measured by the volume concentration in "Radial Bin 15" ($\approx 2.241\,\mu m$), the entry $< AVSD >$ is the mean value of all correlations between the NN-derived AVSD and the AERONET target AVSD, and AAOD(440 V 500) represents the regression of the satellite (from OMI) AAOD at 500 nm against the AERONET AAOD at 440 nm.

3.3.1 Microphysical outputs

The training of all NN cases showed that only the AVSD related to the coarse mode of dust is accurately retrieved from the AOD information. In particular, cases 1–3 retrieved the daily averaged coarse volume concentration $V(c)$ and its modal peak ("Radial bin 15") to a very high level of precision: $0.967 \leq R(d) \leq 0.970$ and $0.956 \leq R(d) \leq 0.983$, respectively. The satellite input case 4 also retrieved the daily averaged coarse volume concentration $V(c)$ and its modal peak, but with $R(d) = 0.365$ and $R(d) = 0.375$, respectively. Only case 4 was able to retrieve the coarse mode geometric radius $r(c)$ with $R(d) = 0.346$. The AERONET input cases

Table 3. Training results obtained for the optimal NN found for each of the cases 1–4. The mean AERONET "Target" values are presented along with the mean NN outputs and the Pearson product–moment correlation coefficient obtained at the daily timescale $R(d)$. The outputs are divided into microphysical parameters derived from the AVSD and the CRI, and the optical parameters SSA and ASYM.

Training		AERONET									Satellite		
		Case 1			Case 2			Case 3			Case 4		
		Target	NN output	$R(d)$	Target	NN output	$R(d)$	Target	NN output	$R(d)$	Target	NN output	$R(d)$
N (samples)			2099			1985			353			134	
AOD(470)			0.721			0.695			0.624			0.609	
AOD(550)			0.704			0.678			0.604			0.583	
AOD(660)			0.684			0.659			0.582			0.556	
H_2O						2.265			2.357			2.289	
AAOD(440 V 500)												0.061	
AOD(380)									0.621				
AOD(500)									0.650				
Microphysics	$V(f)$	0.022	0.022	-0.119	0.022	0.023	0.177	0.023	0.026	0.290	0.029	0.030	0.461
	$V(c)$	0.434	0.432	0.969	0.420	0.417	0.967	0.328	0.320	0.970	0.383	0.342	0.365
	η	0.058	0.058	0.473	0.059	0.061	0.520	0.079	0.087	0.387	0.090	0.085	0.404
	$r(f)$	0.109	0.109	0.031	0.109	0.110	0.209	0.118	0.120	0.288	0.111	0.112	0.243
	$r(c)$	2.051	2.028	0.007	2.055	2.046	0.114	1.855	1.871	0.255	2.018	1.994	0.385
	var(f)	1.139	1.135	-0.007	1.139	1.136	0.053	1.153	1.178	0.046	1.137	1.135	-0.194
	var(c)	1.560	1.562	0.158	1.563	1.569	0.166	1.584	1.572	0.346	1.534	1.560	0.268
	Radial Bin 15			0.982			0.983			0.956			0.375
	$<$ AVSD $>$			0.958			0.961			0.964			0.944
	CRI-R(440)	1.472	1.466	0.068	1.472	1.467	0.447	1.475	1.472	0.476	1.448	1.446	0.532
	CRI-R(675)	1.487	1.482	0.204	1.488	1.484	0.512	1.492	1.490	0.480	1.472	1.472	0.528
	CRI-R(870)	1.471	1.467	0.276	1.473	1.470	0.546	1.481	1.481	0.484	1.464	1.465	0.521
	CRI-R(1020)	1.458	1.453	0.326	1.459	1.457	0.565	1.469	1.469	0.493	1.452	1.454	0.521
	CRI-I(440)	0.005	0.005	0.426	0.005	0.005	0.437	0.006	0.007	0.292	0.006	0.006	0.338
	CRI-I(675)	0.003	0.003	0.450	0.003	0.003	0.461	0.004	0.006	0.258	0.004	0.004	0.354
	CRI-I(870)	0.003	0.003	0.461	0.003	0.003	0.465	0.003	0.006	0.230	0.003	0.004	0.346
	CRI-I(1020)	0.003	0.003	0.473	0.003	0.003	0.473	0.003	0.006	0.220	0.003	0.004	0.331
Optics	SSA(440)	0.900	0.900	0.404	0.899	0.898	0.409	0.898	0.895	0.363	0.900	0.896	0.262
	SSA(675)	0.948	0.945	0.509	0.947	0.944	0.511	0.939	0.931	0.387	0.938	0.934	0.347
	SSA(870)	0.954	0.951	0.508	0.953	0.950	0.508	0.950	0.940	0.381	0.949	0.945	0.354
	SSA(1020)	0.957	0.954	0.512	0.956	0.952	0.510	0.953	0.942	0.379	0.952	0.948	0.351
	ASYM(440)	0.770	0.770	0.435	0.769	0.768	0.643	0.761	0.756	0.530	0.763	0.761	0.322
	ASYM(675)	0.742	0.742	0.456	0.741	0.740	0.657	0.735	0.726	0.422	0.736	0.733	0.335
	ASYM(870)	0.743	0.743	0.399	0.743	0.741	0.630	0.733	0.723	0.382	0.736	0.733	0.370
	ASYM(1020)	0.748	0.748	0.386	0.747	0.746	0.627	0.737	0.727	0.384	0.742	0.739	0.410

1–3 failed here. As described in Appendix A, this is most likely due to the fact that the AVSD of desert dust does not have a clearly defined minimum to separate the coarse and fine modes. This leads to a lot of variation in the location of the mode separation point rs. A lack of correlation in rs then translates into a lack of correlation in the secondary microphysical parameters like the modal geometric radii and variances that depend sensitively on it. For AVSD outputs related to the fine mode, the NN performances were moderately accurate with a maximum $R(d) = 0.461$ for the daily averaged fine mode volume $V(f)$ (case 4 NN). The lack of correlation with the AERONET targets for both $r(f)$ and var(f) for all NNs is due to the fact that for desert dust AVSDs, $V(f)$ is a small proportion of the total volume concentration ($\leq 9\%$). The pre-dominance of the coarse mode in this region meant that all four models retrieved the fine fraction (η) to a similar (poor to moderate) degree: $0.404 \leq R(d) \leq 0.560$. The variation of $R(d)$ across the entire AVSD (not just at radial bin 15) and the daily averaged time series of the retrieved $V(c)$ in case 4 are presented in Fig. 5.

The NN trained with satellite inputs in case 4, retrieved CRI-R with $0.521 \leq R(d) \leq 0.532$, excelling over the AERONET-input NNs. This is likely to be due to the inclusion of the modelled AAOD from OMI in the NN inputs.

3.3.2 Optical outputs

In case 1, all optical parameters (SSA and ASYM) are retrieved with regression coefficients in the range: $0.386 \leq R(d) \leq 0.512$, with the best result being obtained for SSA(1020). The addition of columnar water vapour (H_2O) in case 2, while hardly impacting on the retrieval accuracy of the SSA, led to a significant improvement in the retrieval of the asymmetry factor (ASYM) at all wavelengths: $0.630 \leq R(d) \leq 0.657$. Once again, the case 3 training results, despite having four inputs in common with case 2 underperforms even the case 1 optics outputs (with the exception of ASYM at 440 nm which is slightly better than the case 1 result but still worse than the case 2 retrieval). The addition of the 2 UV AODs in case 3 does not appear then to offer an improvement for dust in Northern Africa. The optical parameter retrievals of SSA and ASYM from the case 4 NN
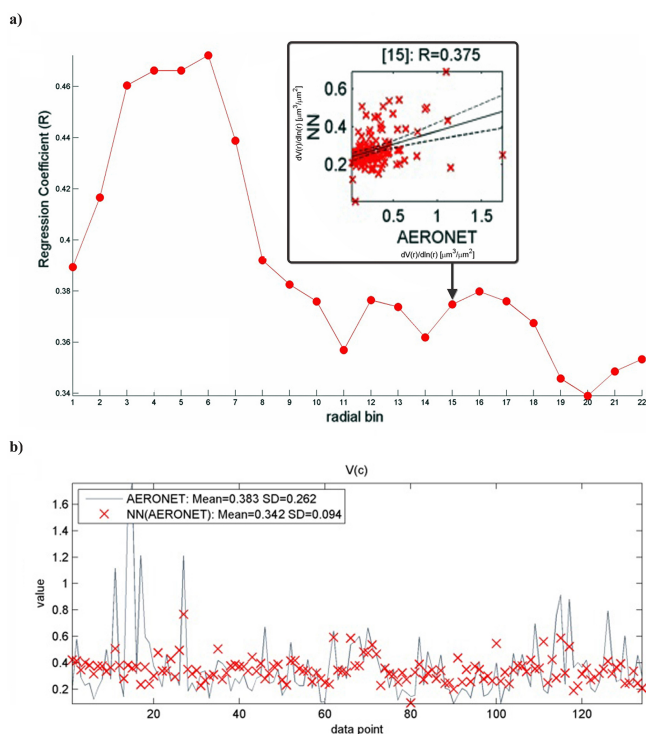
Figure 5. Aerosol microphysical parameter training results obtained for case 4: **(a)** regression per radial bin of the AVSD (inset: Radial bin 15) and **(b)** daily averaged time series for the volume concentration of the coarse mode $V(c)$.

are, in general in the range: $0.322 \leq R(d) \leq 0.410$ (with the exception of SSA(440) where $R(d) = 0.262$). There appears to be a play-off between the ability of the NN to recover all microphysical parameters and simultaneously all optical parameters. This is expected, since the information content of the input parameters is low for retrieval of the complete set of aerosol parameters. The best training and validation results are associated with case 2 NN. In the next section we report on the performance of the case 1–4 trained NNs by feeding them with unseen input data, i.e. NN testing.

4 Results

The performance of the trained NNs was tested by feeding them with unseen case 1–4 input data at the coastal dust site Dakar in Northern Africa (or in the pixel containing the site in the case of satellite inputs). The test outputs are compared with the daily averaged target AERONET microphysical AVSD, the CRI and the optical parameters SSA and ASYM at 440, 675, 870 and 1020 nm. The test results are collected in Table 4 following the same general format as the training results of Table 3.

In addition to the regression coefficient for daily averages $R(d)$, regression coefficients are also calculated for weekly averages $R(w)$ and monthly averages $R(m)$ so as to assess

the behaviour of the NN results at other timescales. It is important, at this point, to make a comment about NN generalization and the potential for extrapolation. While Dakar has a distinct spatial geolocation with respect to the training sites used in the Northern Africa region, Fig. S1 of the Supplement, shows clearly that the range of values of the AERONET targets at Dakar (with the exception of the minimum value of the spectral SSA) can be seen to fall within the range of values of the AERONET targets used to train the NN. As such, the trained case 4 NN is not expected to be able to extrapolate outside this range and to have general extrapolation potential.

4.1 Inputs

As for the training inputs described in Sect. 3.3, for cases 1–2 the number of AERONET Level 2.0 Version 2 inversion products daily averages at Dakar is substantially larger (862–942 records) than the 149 records available in case 3, and the 167 records obtained in case 4 due to the co-location and synchronization (the same day) of AERONET data with the satellite data. The fewer records for case 3 is due to the fact that relatively fewer UV measurements of AOD(380) and AOD(500) exist at Dakar. Another thing to be noted about the input data for case 1–4 is that outliers were deliberately not removed in the testing data sets so as to provide a more stringent test of the NN retrieval. In particular, it is important to compare the CASE 4 satellite inputs with their co-located and synchronous AERONET counterparts. This is especially important for the AAOD which is modelled from OMI, whereas from AERONET is calculated (see discussion in Sect. 3.3.2). With reference to Table 4, the regression of satellite values for AOD(470), AOD(550) and AOD(660) on their AERONET co-located and synchronous counterparts spans the narrow range: $0.421 \leq R(d) \leq 0.442$. A similar level of correlation is found for the AAOD(500): $R(d) = 0.450$. However a strong positive correlation is evident in the case of columnar H_2O: $R(d) = 0.834$. Figure 6 shows the daily averaged time series of AOD(660) (as a representative measure of the aerosol optical thickness), H_2O and AAOD(500) satellite inputs overlaid on the time series of co-located and synchronous AERONET counterparts (note that the AERONET AAOD used for comparison is at 440 nm).

The MODIS and OMI data appear to be systematically lower than AERONET, particularly at higher values. This is explainable by the difference in the way AERONET's ground-based and MODIS's space-based remote sensing instruments measure the AOD. AERONET's sun photometers perform almucantar scans of light radiation based around the pointing direction to the sun (zenith angle) whereas MODIS's spectro-radiometers measure the intensity of solar radiation reflected vertically by the earth's system (the planetary surface and the atmosphere). As a result, the light paths are usually different and sample different angular variations

Table 4. Test results obtained from the optimised trained NNs for cases 1–4 using inputs from the Dakar AERONET site and satellite inputs from MODIS and OMI over Dakar.

Testing (Dakar)		AERONET														
		Case 1					Case 2					Case 3				
		Target	NN output	$R(d)$	$R(w)$	$R(m)$	Target	NN output	$R(d)$	$R(w)$	$R(m)$	Target	NN output	$R(d)$	$R(w)$	$R(m)$
N (samples)			942					931					149			
AOD(470)			0.649					0.640					0.674			
AOD(550)			0.626					0.618					0.650			
AOD(660)			0.603					0.594					0.626			
H_2O								2.698					2.244			
AAOD(440 V 500)																
AOD(380)													0.710			
AOD(500)													0.669			
MICROPHYSICS	$V(f)$	0.026	0.026	0.115	0.279	0.581	0.026	0.026	0.209	0.290	0.504	0.033	0.027	−0.016	0.063	0.390
	$V(c)$	0.357	0.360	0.965	0.950	0.959	0.353	0.355	0.967	0.950	0.961	0.344	0.361	0.969	0.940	0.949
	η	0.092	0.079	0.474	0.599	0.762	0.093	0.085	0.491	0.600	0.812	0.115	0.082	0.400	0.446	0.706
	$r(f)$	0.115	0.118	0.043	0.227	0.344	0.115	0.117	0.029	0.152	−0.048	0.127	0.115	−0.060	−0.237	−0.703
	$r(c)$	1.928	1.909	−0.028	−0.257	0.000	1.929	1.923	0.009	0.103	0.098	1.934	1.888	−0.035	−0.151	0.151
	$var(f)$	1.529	1.514	−0.041	0.065	0.067	1.142	1.171	−0.065	0.018	−0.146	1.165	1.175	0.227	0.248	0.601
	$var(c)$	3.056	2.650	0.239	0.312	0.049	1.529	1.530	0.176	0.273	0.125	1.514	1.501	0.102	0.460	0.466
	Radial Bin 15		0.956					0.943					0.963			
	$< AVSD >$		0.912					0.913					0.906			
	CRI-R(440)	1.472	1.457	0.209	0.235	0.375	1.457	1.458	0.374	0.370	0.542	1.462	1.463	0.307	0.289	0.368
	CRI-R(675)	1.488	1.479	0.048	−0.058	−0.240	1.480	1.481	0.335	0.347	0.473	1.482	1.482	0.318	0.337	0.303
	CRI-R(870)	1.472	1.471	0.175	0.120	0.034	1.471	1.473	0.383	0.396	0.491	1.471	1.471	0.350	0.396	0.300
	CRI-R(1020)	1.459	1.460	0.244	0.220	0.176	1.460	1.461	0.410	0.432	0.529	1.457	1.457	0.379	0.390	0.264
	CRI-I(440)	0.005	0.006	0.406	0.437	0.651	0.006	0.006	0.395	0.427	0.585	0.007	0.007	0.208	0.214	−0.048
	CRI-I(675)	0.003	0.003	0.436	0.464	0.675	0.003	0.003	0.427	0.458	0.617	0.004	0.004	0.169	0.207	−0.004
	CRI-I(870)	0.003	0.003	0.445	0.466	0.665	0.003	0.003	0.433	0.461	0.613	0.004	0.003	0.154	0.199	−0.001
	CRI-I(1020)	0.003	0.003	0.453	0.473	0.661	0.003	0.003	0.439	0.465	0.608	0.004	0.003	0.147	0.195	−0.005
OPTICS	SSA(440)	0.901	0.897	0.336	0.360	0.559	0.896	0.895	0.314	0.291	0.462	0.883	0.885	0.203	0.233	−0.105
	SSA(675)	0.948	0.947	0.472	0.519	0.708	0.947	0.946	0.463	0.506	0.665	0.938	0.941	0.263	0.282	0.035
	SSA(870)	0.954	0.957	0.477	0.509	0.698	0.956	0.956	0.464	0.499	0.655	0.948	0.951	0.275	0.280	0.024
	SSA(1020)	0.956	0.959	0.481	0.507	0.692	0.959	0.958	0.466	0.493	0.643	0.952	0.955	0.277	0.276	0.004
	ASYM(440)	0.769	0.764	0.425	0.555	0.734	0.764	0.763	0.504	0.525	0.489	0.763	0.764	0.544	0.547	0.557
	ASYM(675)	0.742	0.731	0.440	0.507	0.680	0.731	0.730	0.512	0.479	0.452	0.731	0.732	0.451	0.496	0.399
	ASYM(870)	0.744	0.731	0.405	0.471	0.648	0.731	0.730	0.504	0.454	0.425	0.732	0.733	0.395	0.453	0.305
	ASYM(1020)	0.748	0.736	0.393	0.445	0.610	0.736	0.736	0.516	0.446	0.409	0.738	0.740	0.382	0.446	0.276

of aerosol (this effect is likely to be minimized when the sun is overhead but tends to increase close to sunrise and sunset). Furthermore, in the case of measurements from orbit, the separation of the effect of surface reflectance and the effect of aerosol extinction on the total measured radiance is a much more difficult task, especially over deserts which can have bright surface pixels. AOD data from 132 global AERONET stations over a two-year period were regressed against MODIS-derived AOD values, and revealed MODIS values to be systematically lower than AERONET values (Remer et al., 2005).

4.2 Microphysical outputs

For AVSD outputs related to the coarse mode, the AERONET input cases 1–3 were able to retrieve the daily averaged coarse volume concentration $V(c)$ at Dakar to a very high level of precision: $0.965 \leq R(d) \leq 0.969$, and also the location of the coarse mode peak ("Radial bin 15"): $0.943 \leq R(d) \leq 0.963$. This level of accuracy is also maintained at the weekly and monthly timescales. The case 4 NN with satellite inputs was able to retrieve the daily averaged coarse volume concentration $V(c)$ and its modal peak with the degree of correlation: $R(d) = 0.514$ and $R(d) = 0.486$ respectively. The other aerosol parameters are not retrieved accurately enough, as shown also in the NN training/validation performance. More specifically, none of the NNs could retrieve

the daily averaged coarse mode geometric radius $r(c)$ or its variance $var(c)$ for reasons described in Sect. 3.3.1 regarding the problematic determination of the mode separation point for dust AVSDs. The same is true for the daily averaged fine mode volume $V(f)$. The satellite case 4 NN could only retrieve $V(f)$ with $R(d) = 0.261$ (with some improvement at the monthly timescale: $R(m) = 0.388$). Cases 1–4 present unacceptable correlations for $r(f)$ and $var(f)$ which, as described in Sect. 3.3.2, is explained by the fact that for desert dust fine particles contribute only a small proportion to the total volume concentration. Future work will present results of an NN retrieval scheme for regions dominated by other aerosol types such as urban pollution or the products of biomass burning that have a more clearly defined fine mode. With respect to the fine fraction (η), the dominance of the coarse mode meant that cases 1–4 were able to retrieve this daily averaged parameter with an accuracy: $0.400 \leq R(d) \leq 0.491$ with the satellite case 4 NN retrieving η with $R(d) = 0.413$. Much stronger correlations for this parameter are evident at the monthly timescale: $R(m) \geq 0.541$. Satellite retrievals for $V(c)$, its peak at radial bin 15 and η all show some correlation with co-located and synchronous AERONET outputs at the daily timescale: $R(d) = 0.514$, 0.486 and 0.413, respectively. Finally, with respect to the AVSD, in this section the effect of increasing aerosol load (using AOD(470) as a proxy), on the AVSD regression is also briefly investigated. Low values of AOD correspond to

Table 4. Continued.

TESTING (Dakar)		Satellite				
		Case 4				
		Target	NN output	$R(d)$	$R(w)$	$R(m)$
N (samples)		167	167			
AOD(470)		0.590	0.357	0.421		
AOD(550)		0.562	0.316	0.442		
AOD(660)		0.532	0.301	0.439		
H_2O		2.419	2.779	0.834		
AAOD(440 V 500)		0.067	0.048	0.450		
AOD(380)						
AOD(500)						
MICROPHYSICS	$V(f)$	0.030	0.030	0.261	0.214	0.388
	$V(c)$	0.305	0.315	0.514	0.438	0.487
	η	0.112	0.093	0.413	0.329	0.541
	$r(f)$	0.115	0.112	-0.117	-0.296	-0.117
	$r(c)$	1.906	1.891	0.105	-0.060	0.042
	$\text{var}(f)$	1.137	1.134	-0.115	-0.076	-0.096
	$\text{var}(c)$	1.529	1.525	0.114	0.092	0.000
	Radial Bin 15			0.486		
	$<\text{AVSD}>$			0.918		
	CRI-R(440)	1.449	1.450	0.344	0.228	0.294
	CRI-R(675)	1.474	1.475	0.228	0.104	0.162
	CRI-R(870)	1.469	1.470	0.153	0.057	0.179
	CRI-R(1020)	1.460	1.461	0.139	0.036	0.162
	CRI-I(440)	0.007	0.007	0.381	0.347	0.550
	CRI-I(675)	0.004	0.004	0.372	0.288	0.482
	CRI-I(870)	0.004	0.004	0.373	0.274	0.486
	CRI-I(1020)	0.004	0.004	0.368	0.249	0.469
OPTICS	SSA(440)	0.887	0.887	0.440	0.506	0.710
	SSA(675)	0.936	0.935	0.395	0.347	0.562
	SSA(870)	0.947	0.947	0.383	0.314	0.546
	SSA(1020)	0.951	0.950	0.373	0.283	0.521
	ASYM(440)	0.757	0.756	0.159	0.120	0.331
	ASYM(675)	0.725	0.723	0.149	0.084	0.348
	ASYM(870)	0.726	0.724	0.094	0.010	0.304
	ASYM(1020)	0.732	0.731	0.067	-0.031	0.261

small volume concentrations and are important to inspect due to the fact that spurious retrieval effects are known to exist at low number densities (Dubovik and King, 2000). The reason for this is that AERONET's Level 2.0 Version 2 inversion products are obtained following certain constraints: (i) aerosol loads should be moderate (AOD > 0.4), (ii) the sky should not have strong cloud contamination, (iii) solar zenith angles should be high (> 50 degrees) so that the air mass factor is high, and (iv) simultaneous measurements of AOD(440), AOD(675), AOD(870) and AOD(1020) should be available within ±15 min of the almucantar measurement. When these conditions are not satisfied, inversions are less reliable or absent from the AERONET data record. Assessment of the dependence of AVSD on AOD(470) is done as follows: (1) the NN-derived AVSDs were individually regressed on co-located and synchronous AERONET AVSD targets for days sorted by AOD(470), and (2) the 20 % quantiles of AOD(470) were identified and used to calculate the mean AVSD from a sample of AVSDs corresponding to days where the AOD(470) is 10 % above and below the quantile point. Figure 7 looks into this behaviour in more analytical detail.

In the left panel of Fig. 7 showing the variation of the regression coefficient (R) with AOD(470), it is clear that the variation in the value of R decreases with increasing AOD(470). There is much greater variance in the value of R when AOD(470) $\leq$ 0.4. This is expected since, as mentioned above, AERONET retrievals are not as reliable for low aerosol loads. In the right panels of Fig. 7, the mean AVSD is calculated at 20, 40, 60 and 80 % of the min–max range (0.01 to 1.43) of AOD(470) values. The mean NN-derived and AERONET AVSD at each quantile is calculated from a 20 % sample (10 % above and below) in the AOD(470) domain. It can be seen that for the satellite NN of case 4, a substantial difference is observable at the 60 % quantile level where AOD(470) = 0.865 and also at the 80 % level where AOD(470) = 1.15. However, the number of AVSDs used to calculate the mean AVSD at these quantile points is small ($N = 7$ and $N = 3$, respectively) and are not likely to be statistically representative. There is a strong resemblance between the mean AVSD obtained at the more populated 20 and 40 % AOD(470) quantiles. Figure 8 shows a visual comparison of the entire record of daily averaged NN versus AERONET retrievals at Dakar.
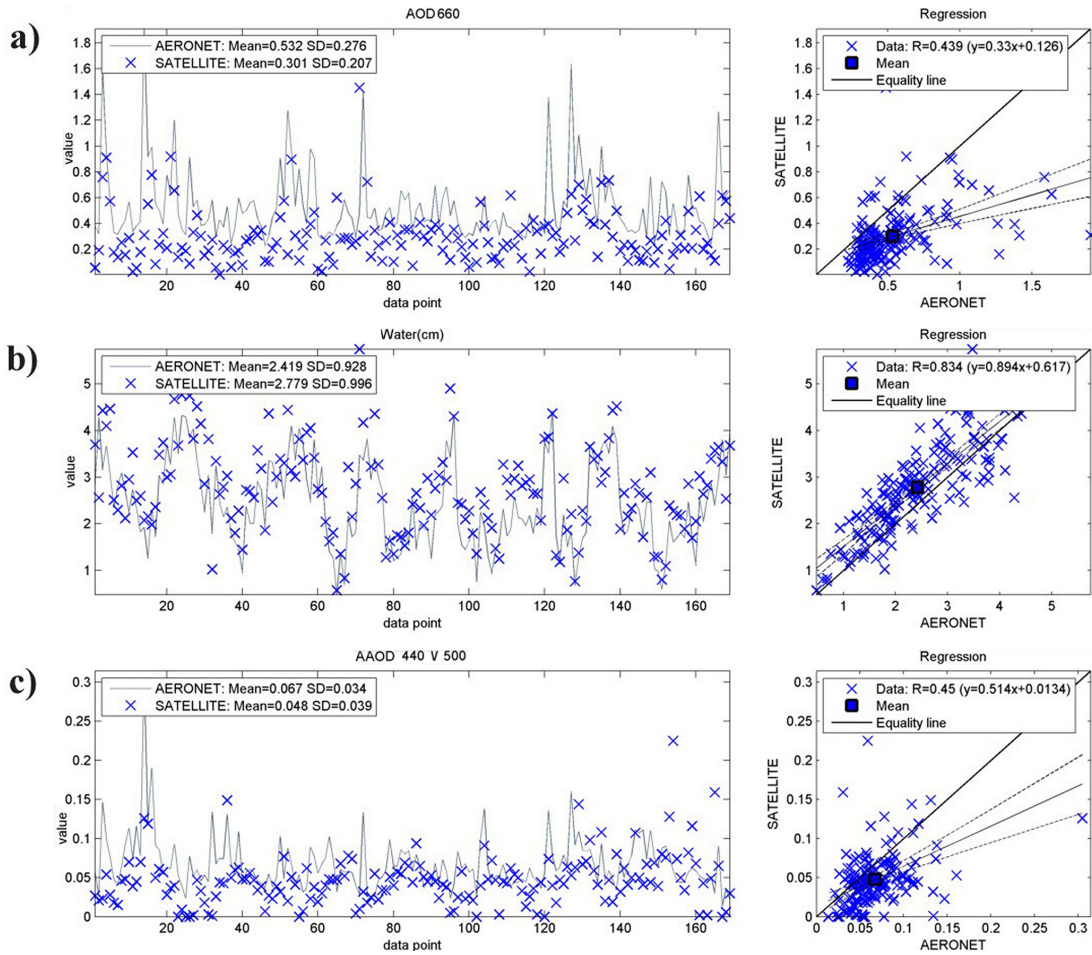
Figure 6. Comparison of representative case 4 satellite inputs with co-located and synchronous AERONET values for the representative parameters: AOD(660), H_2O and AAOD(500). Mean values and standard deviations are shown for each time series together with the results of performing a linear regression.
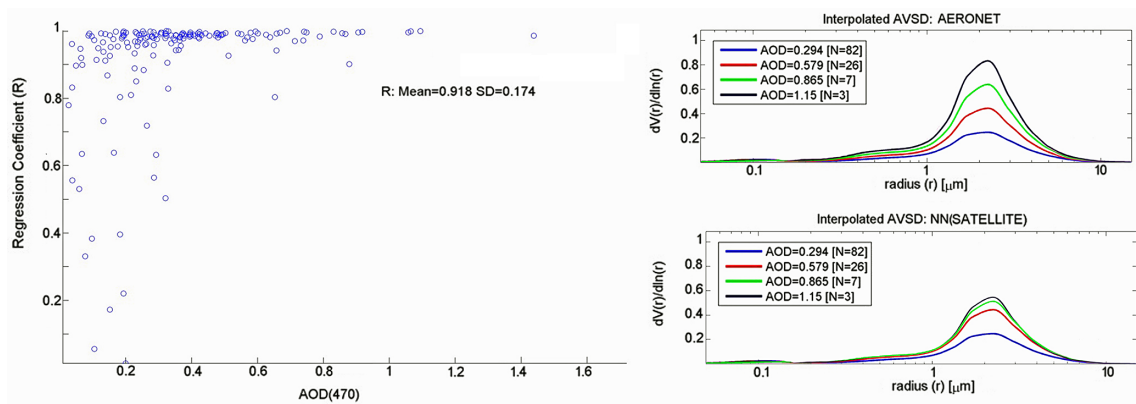

Figure 7. Test results for the dependence of the AVSD regression on aerosol load using AOD(470) as a proxy are shown for the satellite inputs NN of case 4. (Left panel): each point is the regression of the 22 radial bins of the AERONET AVSD on the NN AVSD. Also shown is the AOD $= 0.4$ suggested limit for the validity of the results of the AERONET Level 2.0 Version 2 inversion products. (Right panels): the median AVSD at 20, 40, 60 and 80 % quantile values of AOD(470).

Figure 8. Test results for the daily averaged AVSD at Dakar. The panels at left show the AERONET target distributions (upper) and their corresponding contour plot (lower). The panels at right show the NN output distributions (upper) with their corresponding contour plot (lower).

For more detail, we refer the reader to Sect. S2 of the Supplement where the NN retrieval of the daily averaged AVSD is compared with the AERONET AVSD for each of the test days at Dakar individually. The results suggest that for the most cases, the case 4 NN appears to return an AVSD close to the climatological mean of the training data set. With regard to the complex refractive index, Table 4 shows that none of the NNs were able to retrieve the CRI-R. The results for CRI-I from the case 1 NN are improved substantially at the monthly timescale: $0.651 \le R(m) \le 0.675$ (with CRI-R(440) having the value $R(m) = 0.375$). As described in Sect. 3.3, the addition of H_2O (i.e. the case 2 simulation) improves the regression for CRI-R: $0.335 \le R(d) \le 0.410$ (with even more pronounced positive correlations at the monthly timescale). The retrieval of CRI-I is relatively unaffected by the addition of H_2O to the inputs. These test results validate our claim that H_2O is indeed an important input parameter and should be added to the base set: AOD(470), AOD(550) and AOD(660) for satellite-based retrievals. In particular, H_2O is required for moderate retrieval of CRI-R. This effect is shown in Fig. 9.

The further addition of UV AOD inputs in case 3 did not lead to an increase in the ability of the NN to retrieve the complex refractive index. To the contrary, the correlations were systematically worse. For the satellite inputs case 4 NN, the retrievals of the absorption-related CRI-I are in the range: $0.368 \le R(d) \le 0.381$. The correlation strengthens substantially at the monthly timescale and especially at

shorter wavelengths: $0.469 \le R(m) \le 0.550$. The maximum correlation observed for CRI-R(440) at the daily timescale is $R(d) = 0.344$.

4.3 Optical outputs

Referring to Table 4, for the absorption-related parameter SSA (as noted above for the CRI-I), the retrieval improves with increasing wavelength and also substantially at the monthly timescale: $0.559 \le R(m) \le 0.734$. The addition of H_2O (i.e. the case 2 simulation) leads to a minor improvement in the retrieval of the asymmetry factor (ASYM): $0.504 \le R(d) \le 0.516$. The correlations for SSA are relatively unaffected by the addition of H_2O. Once again, the further addition of UV AOD inputs in case 3 muddied the waters and failed to improve the retrieval of the optical parameters (with the exception of ASYM(440) which showed a slight improvement over the case 1–2 NNs at the daily timescale. For the case 4 NN (satellite inputs), the retrievals of the absorption-related SSA are in the range: $0.373 \le R(d) \le 0.440$. The correlation strengthens substantially at the monthly timescale and especially at shorter wavelengths: $0.521 \le R(m) \le 0.710$. A positive correlation is also observed for ASYM (440–870) at the monthly timescale: $0.304 \le R(m) \le 0.348$. A visual overview of the retrieval performance of the spectrally dependent microphysical (CRI) and optical parameters (SSA and ASYM) at the
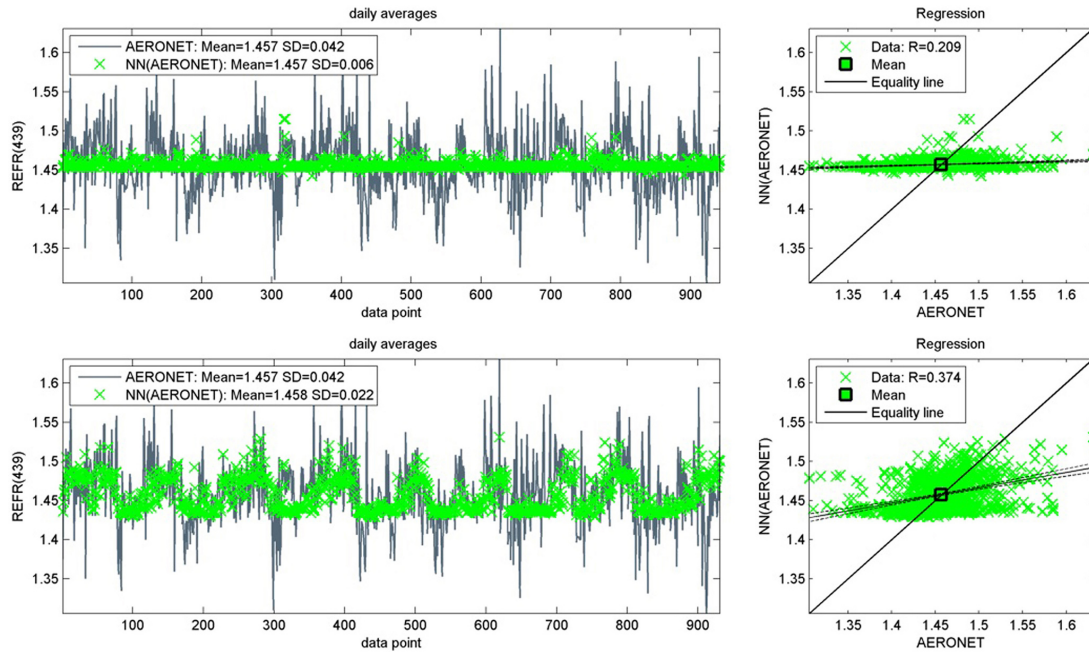
Figure 9. Test results for the daily averaged CRI-R illustrating the effect of adding columnar water vapour (H$_2$O [cm]) as a input (case 1 → case 2).
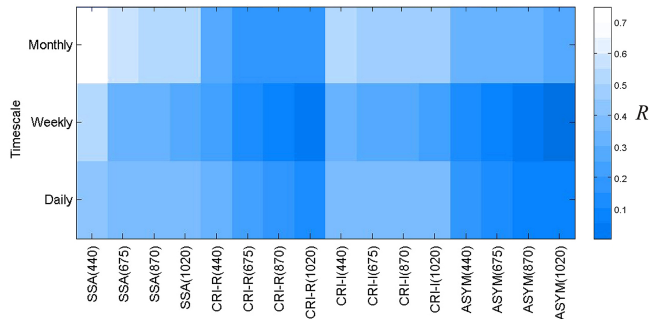

Figure 10. Test results obtained for all spectrally dependant microphysical and optical parameters with the satellite input case 4 NN at the daily, weekly and monthly timescale.

daily, weekly and monthly timescale for the satellite case 4 is shown in Fig. 10.

When tabulated in this micro-array format, one can see at a glance that the satellite input trained NN of case 4 retrieves the spectral behaviour of the absorption-related SSA and CRI-I parameters better than the shape-related CRI-R and ASYM parameters at all timescales. More detail is revealed by looking at the time series of the daily averaged retrievals. For example, in Fig. 11, daily averaged retrievals of SSA(440) at Dakar are shown for the case 4 NN.

Figure 11 shows that the satellite retrieval at Dakar, while insufficiently fitting the magnitude of peaks and troughs in the SSA(440) time series, does echo them to some degree.

4.4 Evaluation of the results with respect to AERONET data variability and errors

In this section, we investigate the ability of the NN to capture the variability of the target data. Also, we evaluate the information content of the NN results taking into account the uncertainties in the AERONET data. Figure 12 shows the case 4 NN retrieval of the coarse mode volume concentration $V(c)$ at the daily and seasonal (3-monthly) timescales compared with AERONET data at Dakar.

While the mean values are almost indistinguishable, the standard deviation of the NN retrieval is approximately 50 % of the standard deviation of the AERONET data at both timescales. This suggests that, while the input information used to train the NN is enough for retrieval of the climatologically expected value, it is not sufficient to fully retrieve the variability in the target data at the daily timescale. In order to test this, we checked to see whether or not the median absolute error (MAE) for each NN output (NN-AERONET) is significantly lower than the MAE of the difference between the AERONET target values of that parameter and their mean value over the training set (AERONET-mean) at Dakar. The percentage fractional error (PFE) between the two MAEs was found to be negative but small for the majority of the parameters ($-9.2\% \leq$ PFE(CRI-R) $\leq -17.0\%$, $-8.3\% \leq$ PFE(SSA) $\leq -13.2\%$ and $-10.4\% \leq$ PFE(ASYM) $\leq -12.3\%$). Two exceptions to this were the coarse mode volume which was strongly negative (-321.9%) and the CRI-I(440) which was very weakly positive ($+1.0\%$). This tends to support the view
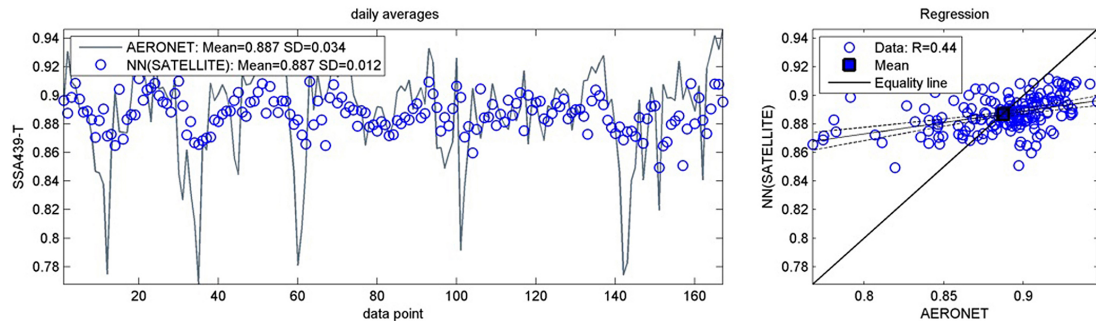
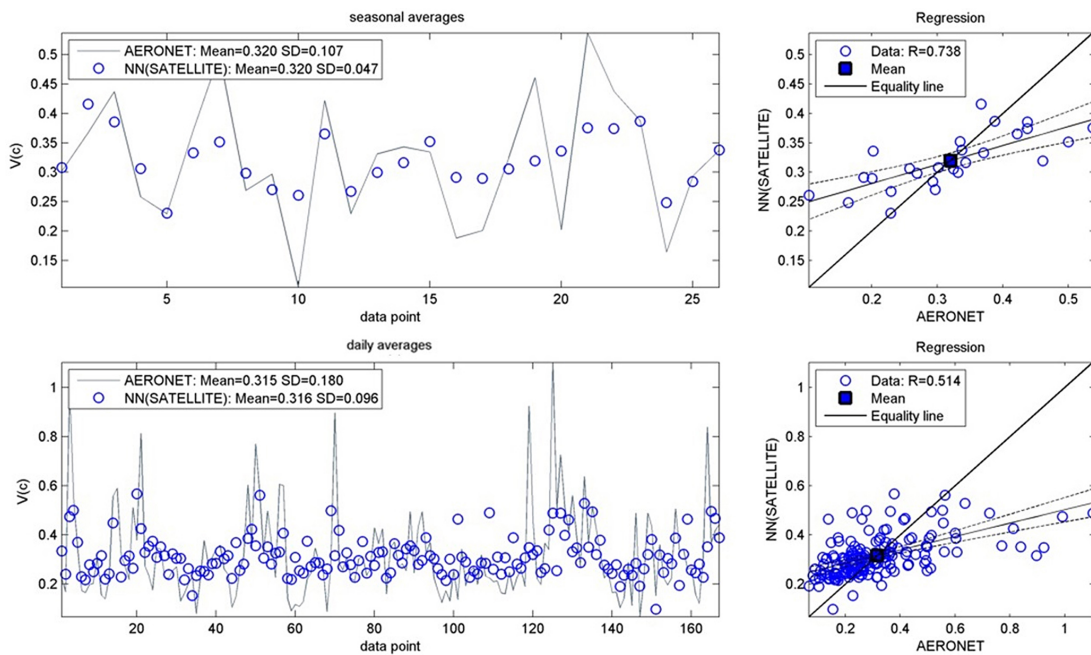**Figure 11.** Test results at Dakar for SSA(440) with the case 4 NN.

Figure 12. Test results at Dakar for the volume concentration of the coarse mode $V(c)$ at the seasonal (3-monthly) timescale (upper panels) and the daily timescale (lower panels). Note that while the mean value of the NN retrieval and the AERONET target data are almost equal, the standard deviation of the NN retrieval is approximately half of that associated with the AERONET target data.

that the NN is capturing only the daily variability for the coarse mode. For all other output parameters, the NNs only learn to return the average value of the parameter over the training set or, in other words, the climatological mean. This can be seen for the case 4 NN in Figs. S3.1 and S3.2 of the Supplement.

Finally, in this section, in Table 5 we present the values of the MAE and the MARE (the median absolute relative error expressed as a percentage) for $V(c)$ as a proxy for the AVSD, the spectral CRI and the spectral optical parameters SSA and ASYM at the daily, weekly and monthly timescales.

Figures S4.1–S4.4 of the Supplement show the trend in the MAE as a function of timescale for case 4 NN at Dakar over the range of scales: 1-dy to 1-yr. It should be borne in mind that, while the MAE and the MARE are a measure of the NN retrieval with respect to the AERONET target values, the

AERONET values themselves are not error-free. AERONET aerosol parameters themselves also often have non-negligible uncertainties (Dubovik et al., 2000). A formal evaluation of the uncertainty of the NN with respect to true values is beyond the scope of this paper and we refer the reader to recent work on NN uncertainty in the retrieval of AOD by Ristovski et al. (2012). It is hoped that further validation studies using a cohort of larger data sets will be able to provide a more clear assessment of NN performance.

5 Discussion

A new methodology has been developed, based on an NN model, with the aim of retrieving aerosol microphysical and optical parameters from satellite remote sensing data at the

Table 5. Test results at Dakar obtained for cases 1–4 for the median absolute difference (MAE) and median absolute relative error (MARE) of output parameters at the daily, weekly and monthly timescale. $V(c)$ is the volume concentration of the coarse mode calculated from the AVSD.

Timescale	Case 1						Case 2					
Analysis	Daily (d)		Weekly (w)		Monthly (m)		Daily (d)		Weekly (w)		Monthly (m)	
	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %
$V(c)$	0.081	11.6	0.093	15.6	0.072	11.5	0.079	11.6	0.090	15.2	0.062	12.0
CRI-R(440)	0.022	1.5	0.022	1.6	0.021	1.5	0.021	1.4	0.020	1.4	0.018	1.2
CRI-R(675)	0.018	1.2	0.019	1.3	0.017	1.1	0.018	1.2	0.016	1.1	0.013	0.9
CRI-R(870)	0.018	1.2	0.018	1.3	0.018	1.2	0.016	1.1	0.017	1.1	0.013	0.9
CRI-R(1020)	0.018	1.3	0.017	1.2	0.015	1.1	0.017	1.1	0.017	1.2	0.014	0.9
CRI-I(440)	0.002	34.4	0.002	35.1	0.002	36.7	0.002	42.2	0.002	42.2	0.002	40.6
CRI-I(675)	0.001	53.2	0.001	49.4	0.001	50.4	0.002	67.7	0.002	55.5	0.001	50.9
CRI-I(870)	0.001	49.1	0.001	45.5	0.001	46.9	0.001	62.9	0.001	50.2	0.001	47.7
CRI-I(1020)	0.001	47.0	0.001	44.2	0.001	39.8	0.001	59.2	0.001	49.9	0.001	44.8
SSA(440)	0.014	1.6	0.022	2.4	0.019	2.1	0.017	1.9	0.023	2.6	0.023	2.6
SSA(675)	0.014	1.5	0.017	1.8	0.017	1.8	0.017	1.8	0.020	2.0	0.019	2.0
SSA(870)	0.012	1.3	0.014	1.5	0.015	1.5	0.016	1.6	0.017	1.7	0.016	1.6
SSA(1020)	0.012	1.2	0.013	1.4	0.015	1.5	0.015	1.6	0.016	1.6	0.015	1.5
ASYM(440)	0.011	1.4	0.010	1.3	0.007	1.0	0.011	1.4	0.012	1.5	0.010	1.3
ASYM(675)	0.013	1.8	0.011	1.5	0.008	1.0	0.013	1.8	0.012	1.7	0.010	1.4
ASYM(870)	0.013	1.8	0.011	1.5	0.008	1.1	0.013	1.7	0.012	1.7	0.010	1.4
ASYM(1020)	0.012	1.6	0.011	1.5	0.007	1.0	0.011	1.5	0.011	1.5	0.010	1.3

Timescale	Case 3						Case 4					
Analysis	Daily (d)		Weekly (w)		Monthly (m)		Daily (d)		Weekly (w)		Monthly (m)	
	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %	MAE	MARE %
$V(c)$	0.131	13.4	0.165	23.8	0.127	11.6	0.268	27.6	0.265	32.5	0.255	24.8
CRI-R(440)	0.022	1.5	0.026	1.8	0.024	1.7	0.031	2.2	0.028	2.0	0.028	1.9
CRI-R(675)	0.017	1.2	0.024	1.6	0.018	1.2	0.022	1.5	0.022	1.5	0.023	1.5
CRI-R(870)	0.019	1.3	0.027	1.8	0.020	1.3	0.022	1.5	0.024	1.7	0.022	1.5
CRI-R(1020)	0.019	1.4	0.027	1.9	0.021	1.4	0.024	1.6	0.024	1.6	0.022	1.5
CRI-I(440)	0.002	38.7	0.002	28.0	0.003	41.9	0.002	40.8	0.002	42.1	0.002	34.6
CRI-I(675)	0.002	52.8	0.002	46.1	0.002	55.4	0.002	60.0	0.002	63.3	0.002	52.9
CRI-I(870)	0.001	42.9	0.001	44.4	0.002	52.6	0.002	60.2	0.002	63.7	0.001	59.5
CRI-I(1020)	0.001	40.8	0.001	48.8	0.002	54.5	0.002	57.0	0.002	64.1	0.001	58.2
SSA(440)	0.013	1.4	0.018	2.1	0.022	2.4	0.018	2.0	0.020	2.2	0.021	2.3
SSA(675)	0.017	1.8	0.019	1.9	0.033	3.4	0.021	2.2	0.022	2.3	0.023	2.4
SSA(870)	0.017	1.8	0.017	1.7	0.024	2.7	0.020	2.1	0.018	1.9	0.021	2.2
SSA(1020)	0.018	1.8	0.015	1.5	0.023	2.4	0.020	2.1	0.018	1.9	0.017	1.8
ASYM(440)	0.013	1.7	0.017	2.2	0.011	1.5	0.020	2.6	0.017	2.3	0.017	2.2
ASYM(675)	0.016	2.2	0.016	2.2	0.013	1.8	0.024	3.2	0.021	2.8	0.020	2.7
ASYM(870)	0.016	2.1	0.017	2.3	0.013	1.8	0.022	3.0	0.022	3.0	0.019	2.6
ASYM(1020)	0.013	1.7	0.016	2.2	0.013	1.7	0.020	2.6	0.020	2.8	0.016	2.2

daily timescale and to an acceptable degree of accuracy. Through the use of different input scenarios we performed an empirical sensitivity analysis of the available measurements from satellite sensors for retrieving the properties of dust aerosol in the Northern Africa region. The NNs were regularised and trained with AERONET Level 2.0 Version 2 inversion products at sites centred on the peak of dust extinction (according to the GOCART model averaged over a 10-year period) in Northern Africa, and have been shown to be capable to some degree of learning the relationship between satellite inputs and the desired output parameters. The trained NNs have the added benefit that they retrieve the entire time series of all output parameters simultaneously. We were also able to demonstrate a technique for objectively deducing optimal NN architectures by minimizing the back-propagation cost function over a grid of runs. While such an approach is well established in the scientific literature (Gorr et al., 1994; Lawrence et al., 1996; Curry and Morgan, 2006; Stathakis, 2009), this is the first time it has been applied in the development of an atmospheric measurement technique. Since, in regression schemes like NN models, possible redundancies in both the data and the NN model space can lead to ill-posed problems, we have tried to eliminate these problems by carefully selecting data of the same aerosol type (predominantly dust as flagged up by the GOCART model global average), by constructing representative test scenarios, and by removing missing values and outliers. Furthermore, PCA was used to extract components from the variables in the NN model space, to eliminate redundancies and to increase the performance of the NN-based retrievals.

Table 6. Overall assessment of the simulation performance of trained NNs fed with AERONET inputs (cases 1–3) and satellite inputs (case 4) at the daily timescale.

ASSESSMENT	AERONET			SATELLITE
	Case 1	Case 2	Case 3	Case 4
$V(f)$	Very Poor	Poor	Very Poor	Poor
$V(c)$	Very Good	Very Good	Very Good	Moderate
η	Moderate	Moderate	Moderate	Moderate
Radial Bin 15	Very Good	Very Good	Very Good	Moderate
CRI-R(440)	Poor	Poor	Poor	Poor
CRI-R(675)	Very Poor	Poor	Poor	Poor
CRI-R(870)	Very Poor	Poor	Poor	Very Poor
CRI-R(1020)	Poor	Moderate	Poor	Very Poor
CRI-I(440)	Moderate	Poor	Poor	Poor
CRI-I(675)	Moderate	Moderate	Very Poor	Poor
CRI-I(870)	Moderate	Moderate	Very Poor	Poor
CRI-I(1020)	Moderate	Moderate	Very Poor	Poor
SSA(440)	Poor	Poor	Poor	Moderate
SSA(675)	Moderate	Moderate	Poor	Poor
SSA(870)	Moderate	Moderate	Poor	Poor
SSA(1020)	Moderate	Moderate	Poor	Poor
ASYM(440)	Moderate	Moderate	Moderate	Very Poor
ASYM(675)	Moderate	Moderate	Moderate	Very Poor
ASYM(870)	Moderate	Moderate	Poor	Very Poor
ASYM(1020)	Poor	Moderate	Poor	Very Poor
Very Poor	$R(d) < 0.2$			
Poor	$0.2 \leq R(d) < 0.4$			
Moderate	$0.4 \leq R(d) < 0.6$			
Good	$0.6 \leq R(d) < 0.8$			
Very Good	$R(d) \geq 0.8$			

With regard to testing the ability of the NNs to perform well on unseen data at Dakar, it is important to bear in mind that while an estimate of output uncertainties is provided with reference to known a priori target values, the test results presented in Sect. 4 incorporate also a network-induced error and are therefore only approximate. Having said this, the histogram of the differences between NN PC outputs and PC targets at the training stage was found to present a sharply peaked Gaussian having a near-zero mean error. The NN test results at the daily timescale are presented in a qualitative way in Table 6 with a categorization based on a linear scale of values of the regression coefficient.

More specifically, assessing the performance of NNs trained on different combinations of inputs in the context of Northern African desert dust data revealed the following.

1. AERONET-measured AOD inputs (470, 550 and 600 nm) alone (case 1) are insufficient to retrieve the daily spectral behaviour of the CRI-R simultaneously with the AVSD, CRI-I and secondary microphysical parameters, together with the optical parameters SSA and ASYM.

2. The inclusion of AERONET H_2O in the inputs improves the retrieval potential of the NN, especially with respect to the daily spectral behaviour of CRI-R and ASYM.

3. The further inclusion of AERONET UV AOD (380 and 500 nm) in the inputs led to a deterioration in the performance particularly with respect to the absorption-related parameters SSA and CRI-I at the daily timescale.

4. The NN trained with MODIS AOD (470, 550 and 660 nm) and H_2O, and OMI-modelled AAOD(500 nm) was able to retrieve with a good to very good degree of accuracy the daily averaged coarse mode of the AVSD at the daily timescale.

6 Conclusions

The results show that it is possible to deduce an optimal NN architecture and to train it to retrieve the daily averaged volume concentration of the coarse mode of dust aerosol to a high degree of accuracy from satellite inputs. The potential

of the NN for retrieving size distribution information is interesting as this may open up the possibly of adding size distribution data to the arsenal of satellite products currently available. The climatological mean retrieval of the complex refractive index and the optical parameters, although unable to provide information regarding daily variability, nevertheless can provide important information on these key parameters over regions where no ground-truth data exists. In essence, the NN model applied to satellite inputs, may allow for the creation of a virtual space-based AERONET climatology centred at 1×1 degree resolution grid points over the earth's surface.

The results presented here are appropriate to dust-dominated data over Northern Africa and further studies will assess whether or not the same methodology can be applied to other dust regions, as well as to regions dominated by other key aerosol types such as marine aerosol and the products of biomass burning and urban pollution. The NN model developed appears to offer some potential for obtaining daily retrievals from satellite data, and it is hoped, will contribute to efforts currently under way for globally monitoring aerosols from space and hence improving assessments of global climate forcing.

Appendix A: Calculation of secondary microphysical parameters from the AVSD

As mentioned in Sect. 2.1.2, the secondary microphysical parameters: $r(f)$, $r(c)$, var(f), var(c), $V(f)$ and $V(c)$ need to be calculated so as to compare NN outputs with AERONET outputs. In AERONET's retrieval algorithm, all of these parameters are calculated from the AVSD by specifying a mode separation point r_s that divides the distribution into 2 volume concentrations $V(f)$ and $V(c)$. From the retrieved AVSD: $dV(r)/d\ln r$, the volume concentration V occupied by particles spanning the range of radial sizes $[r_1, r_2]$ is then calculated by integrating over the distribution,

$$V = \int_{r_1}^{r_2} \frac{dV(r)}{d\ln r} \, d\ln r. \tag{A1}$$

In principle, the aerosol number size distribution (ANSD): $dN(r)/d\ln r$ or $dN(r)/dr$, could equally well be used instead of the AVSD (King et al., 1978), since the conversion between the AVSD and ANSD parameters is straight-forward (see for example Appendix A of Sayer et al., 2012). Note that the AVSD is preferable to the ANSD as it is more accurate when inverting scattering properties that are more sensitive to aerosol particle volume, than number (Dubovik et al., 2011). The AERONET inversion algorithm estimates V integration by means of the trapezium rule (Dubovik and King, 2000), and the same approach was adopted in this work. The volume concentration of the fine mode $V(f)$ is obtained by setting $r_1 = 0.05$ μm and $r_2 = r_s$ while the volume concentration of the coarse mode $V(c)$ is obtained by setting $r_1 = r_s$ and $r_2 = 15$ μm. The ratio of the area of the AVSD contributed to by the fine mode to the total area over the whole distribution constitutes the fine fraction η. The logarithmic volume geometric radius (mean logarithm of radius) measures the characteristic size of typical aerosol particles in the atmospheric column sampled, and is given by the following:

$$\ln r_V = \frac{\int_{r_1}^{r_2} \ln r \frac{dV(r)}{d\ln r} \, d\ln r}{\int_{r_1}^{r_2} \frac{dV(r)}{d\ln r} \, d\ln r}. \tag{A2}$$

The geometric radius of the fine mode $r(f)$ is obtained by setting $r_1 = 0.05$ μm and $r_2 = r_s$ in this expression and then exponentiating, while the geometric radius of the coarse mode $r(c)$ is obtained by setting $r_1 = r_s$ and $r_2 = 15$ μm and then exponentiating. The geometric standard deviation which measures the spread of particle modes is given by the following:

$$\sigma_V = \sqrt{\frac{\int_{r_1}^{r_2} (\ln r - \ln r_V)^2 \frac{dV(r)}{d\ln r} \, d\ln r}{\int_{r_1}^{r_2} \frac{dV(r)}{d\ln r} \, d\ln r}}. \tag{A3}$$

The geometric variance of the fine mode var(f) is obtained by setting $r_1 = 0.05$ μm and $r_2 = r_s$ in this expression and then squaring, while the geometric variance of the coarse mode var(c) is obtained by setting $r_1 = r_s$ and $r_2 = 15$ μm and then squaring. From the above, it is clear that all secondary microphysical parameters depend on a precise determination of the fine mode/coarse mode separation point r_s. At present, AERONET estimates this by finding the minimum within the size interval $0.439 \le r \le 0.992$ μm (Dubovik et al., 2000). The same approach was used in this study although there are signs (see Sects. 3.3.2 and 4) that this is perhaps problematic for aerosol distributions like those for desert dust that do not have clearly separated fine and coarse modes.

Table A1. Acronyms.

AERONET	aerosol robotic network
AAOD	absorption aerosol optical depth
AOD	aerosol optical depth
ASYM	asymmetry factor
AVSD	aerosol volume size distribution
CRI-R	complex refractive index-real part
CRI-I	complex refractive index-imaginary part
GOCART	global ozone chemistry aerosol radiation and transport model
MAE	median absolute error
MARE	median absolute relative error
MODIS	moderate-resolution imaging spectro-radiometer
MSE	mean squared error
NN	neural network
OMI	ozone measuring instrument
PCA	principal components analysis
SSA	single scattering albedo

References

Abdi, H. and Williams, L. J.: Principal component analysis, Wiley Interdisciplinary Reviews, Comput. Stat., 2, 433–459, doi:10.1002/wics.101, 2010.

AERONET: Level 2.0 Version 2 daily averaged almucantar inversion products, available at:f http://aeronet.gsfc.nasa.gov/cgi-bin/combined_data_access_inv, last access: 7 April 2012.

Albayrak, A., Wei, J., Petrenko, M., Lynnes, C., and Levy, R. C.: Global bias adjustment for MODIS aerosol optical thickness using neural network, J. Appl. Remote Sens., 7, 073514, 1–16, 2013.

Bishop, C. M.: Neural Networks for Pattern Recognition, Oxford University Press, New York, NY, USA, 1995.

Chin, M., Rood, R. B., Lin, S. J., Müller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties, J. Geophys. Res., 105, 24671–24687, 2000.

Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B. N., Duncan, B. N., and Nakajima, T.: Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements, J. Atmos. Sci., 59, 461–483, 2002.

Curry, B. and Morgan, P. H.: Model selection in neural networks: some difficulties, Euro. J. Operat. Res., 170, 567–577, 2006.

Cybenko, G.: Approximation by super-positions of a sigmoidal function, Math. Control Signals, 2, 303–314, 1989.

Demuth, H. and Beale, M.: Neural Network Toolbox For Use with MATLAB, Mathworks, 2004.

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.-Atmos., 105, 15329–15346, doi:10.1029/2000JD900148, 2000.

Deuzé, J. L., Bréon, F. M., Devaux, C., Goloub, P., Herman, M., Lafrance, B., and Tanré, D.: Remote sensing of aerosols over land surfaces from POLDER-ADEOS-1 polarized measurements, J. Geophys. Res.-Atmos., 106, 4913–4926, doi:10.1029/2000JD900364, 2001.

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, 20673–20696, 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.-Atmos., 105, 9791–9806, 2000.

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci. 59, 590–608, 2002.

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.

Gobbi, G. P., Kaufman, Y. J., Koren, I., and Eck, T. F.: Classification of aerosol properties derived from AERONET direct sun data, Atmos. Chem. Phys., 7, 453–458, doi:10.5194/acp-7-453-2007, 2007.

GOCART: Version 4 model 3 hourly yearly model output, available at: http://aeronet.gsfc.nasa.gov/cgi-bin/bamgomas_interactive, last access: 7 April 2012.

Gorr, W. L., Nagin, D., and Szczypula, J.: Comparative study of artificial neural network and statistical models for predicting student grade point averages, Int. J. Forecast., 10, 17–34, 1994.

Grubbs, F. E.: Procedures for detecting outlying observations in samples, Technometrics, 11, 1–21, 1969.

Hansen, J., Nazarenko, L., Ruedy, R., Sato, M., Willis, J., Del Genio, A., and Tausnev, N.: Earth's energy imbalance: confirmation and implications, Science, 308, 1431–1435, 2005.

Hasekamp, O. P., Litvinov, P., andButz, A.: Aerosol properties over the ocean from PARASOL multiangle photopolarimetric measurements. J. Geophys. Res., 116, D14204, doi:10.1029/2010JD015469, 2011.

Holben, B. N., Eck, T. F., Slutsker, I., Tanre, D., Buis, J. P., Setzer, A., and Smirnov, A.: Aeronet – a federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Hornik, J., Stinchombe, M., and White, H.: Mutlilayer feedforward networks are universal approximators, Neural Networks, 2, 359–366, 1989.

Intergovernmental Panel on Climate Change (IPCC): Climate change 2007, the Physical Science Basis: Contribution of the Working Group I to the Fourth Assessment Report of the IPCC, Cambridge University Press, New York, 2007.

Intergovernmental Panel on Climate Change (IPCC): Climate change 2013 The Physical Science Basis: Contribution of the Working Group I to the Fifth Assessment Report of the IPCC, Cambridge University Press, New York, 2013.

Jolliffe, I. T.: Principal Component Analysis, 2nd Edn., Springer, New York, 2002.

King, M. D., Byrne, D. M., Herman, B. M., and Reagan, J. A.: Aerosol size distributions obtained by inversion of spectral optical depth measurements, J. Atmos. Sci., 21, 2153–2167, 1978.

Lawrence, S., Giles, C. L., and Tsoi, A. C.: What size neural network gives optimal generalization? Convergence properties

of backpropagation, Technical Report UMIACS- TR-96-22 and CS-TR-3617, Institute for Advanced Computer Studies, University of Maryland, College Park, MD, United States, 1996.

Levenberg, K.: A method for the solution of certain non-linear problems in least squares, Q. Appl. Math., 2, 164–168, 1944.

Marquardt, D.: An algorithm for least-squares estimation of nonlinear parameters, Siam J. Appl. Math., 11, 431–441, 1963.

Mishchenko, M. I., Cairns, B., Kopp, G., Schueler, C. F., Fafaul, B. A., Hansen, J. E., Hooker, R. J., Itchkawich, T., Maring, H. B., and Travis, L. D.: Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the Glory mission, B. Am. Meteorol. Soc., 88, 677–691, 2007.

MODIS: Terra & Aqua Level 3 Collection 5.1 daily global 1_×1_products, available at: http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=MODIS_DAILY_L3, last access: 7 April 2012.

OMI: Aura Level 3 Version 3 daily global 1_×1_gridded aerosol products, available at: http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=omi, last access: 7 April 2012.

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.

Prospero, J. M., Ginoux, P., Torres, O., Nicholson, S. E., and Gill, T. E.: Environmental characterization of global sources of atmospheric soil dust identified with the Nimbus 7 Total Ozone Mapping Spectrometer (TOMS) absorbing aerosol product, Rev. Geophys., 40, 1–31, 2002.

Radosavljevic, V., Vucetic, S., and Obradovic, Z.: A data-mining technique for aerosol retrieval across multiple accuracy measures, IEEE Geosci. Remote S., 7, 411–415, 2010.

Remer, L. A., Kaufman, Y. J., Tanre, D., Mattoo, S., Chu, D. A., Martins, J. V., Li, R. R., Ichoku, C., Levy, R. C., Kleidman, R. G., Eck, T. F., Vermote, E., and Holben, B. N.: The MODIS aerosol algorithm, products, and validation, J. Atmos. Sci., 62, 947–973, 2005.

Ristovski, K., Vucetic, S., and Obradovic, Z.: Uncertainty analysis of neural-network-based aerosol retrieval, Geosci. Remote Sens., 50, 409–414, 2012.

Samet, J. M., Dominici, M. D. F., Curriero, F. C., Coursac, I., and Zeger, S. L.: Fine particulate air pollution and mortality in 20 U.S. Cities, 1987–1994, New. Engl. J. Medicine, 343, 1742–1749, doi:10.1056/NEJM200012143432401, 2000.

Sayer, A. M., Smirnov, A., Hsu, N. C., and Holben, B. N.: A pure marine aerosol model, for use in remote sensing applications, J. Geophys. Res., 117, 1–25, 2012.

Schwartz, S. E.: Uncertainty requirements in radiative forcing of climate change, J. Air Waste Manage., 54, 1351–9, 2004.

Sellitto, P., Di Noia, A., Del Frate, F., Burini, A., Casadio, S., and Solimini, D.: On the role of visible radiation in ozone profile retrieval from nadir UV/VIS satellite measurements: An experiment with neural network algorithms inverting SCIAMACHY data, J. Quant. Spectr. Radiat. Trans., 113, 1429–1436, 2012.

Stathakis, D.: How many hidden layers and nodes?, Int. J. Remote Sens., 30, 2133–2147, 2009.

Tanré, D., Herman, M., and Kaufman, Y. J.: Information on aerosol size distribution contained in solar reflected spectral radiances, J. Geophys. Res., 101, 19043–19060, 1996.

Torres, O., Bhartia, P. K., Herman, J. R., Syniuk, A., Ginoux, P., and Holben, B.: A long term record of aerosol optical depth from TOMS observations and comparison to AERONET measurements, J. Atmos. Sci., 59, 398–413, 2002.

Torres, O., Tanskanen, A., Veihelmann, B., Ahn, C., Braak, Rz., Bhartia, P. K., and Levelt, P.: Aerosols and surface UV products from Ozone Monitoring Instrument observations: an overview, J. Geophys. Res.-Atmos., 112, D24S47, doi:doi:10.1029/2007JD008809, 2007.

Waquet, F., Peers, F., Goloub, P., Ducos, F., Thieuleux, F., Derimian, Y., Riedi, J., Chami, M., and Tanré, D.: Retrieval of the Eyjafjallajökull volcanic aerosol optical and microphysical properties from POLDER/PARASOL measurements, Atmos. Chem. Phys., 14, 1755–1768, doi:10.5194/acp-14-1755-2014, 2014.