A Neural Network Aerosol Typing Algorithm Based on Lidar Data

Doina Nicolae[1], Jeni Vasilescu[1], Camelia Talianu[1,2], Ioannis Binietoglou[1], Victor Nicolae[1,3], Simona Andrei[1], and Bogdan Antonescu[1]

[1]National Institute of R&D for Optoelectronics, 409 Atomiştilor Str., Măgurele, Ilfov, România
[2]Institute of Meteorology, University of Natural Resources and Life Sciences, 33 Gregor-Mendel Str., 1180,Vienna, Austria
[3]Faculty of Physics, University of Bucharest, Atomiştilor 405, Măgurele, Ilfov, Romania

Correspondence: Doina Nicolae (nnicol@inoe.ro)

Abstract. Atmospheric aerosols play a crucial role in the earth system, but their role is not completely understood, partly because of the large variability in their properties resulting from a large number of possible aerosol sources. Recently developed lidar-based techniques were able to retrieve the height distributions of optical and microphysical properties of fine-mode and coarse-mode particles, providing the types of the aerosols. One of such techniques is based on artificial neural networks (ANNs). In this article, a Neural Network Aerosol Typing Algorithm Based on Lidar Data (NATALI) was developed to estimate the most probable aerosol type from a set of multispectral lidar data.The algorithm has been adjusted for running on the EARLINET $3\beta + 2\alpha$ (+1δ) profiles. The NATALI algorithm is based on the ability of specialized ANNs to resolve the overlapping values of the intensive optical parameters, calculated for each identified layer in the multiwavelength Raman lidar profiles. The ANNs were trained using synthetic data, for which a new aerosol model was developed. Two parallel typing schemes were implemented in order to accommodate datasets containing (or not) the measured linear particle depolarization ratios (LPDR): a) identification of 14 aerosol mixtures (high-resolution typing) if the LPDR is available in the input data files, and b) identification of 5 predominant aerosol types (low-resolution typing) if the LPDR is not provided. For each scheme, three ANNs were run simultaneously, and a voting procedure selects the most probable aerosol type. The whole algorithm has been integrated into a Python application. The limitation of NATALI is that the results are strongly dependent on the input data, and thus the outputs should be understood accordingly. Additional applications of NATALI, e.g. to test the quality of the optical data and identify incorrect calibration or insufficient cloud screening is feasible. Blind tests on EARLINET data samples showed the capability of NATALI to retrieve the aerosol type from a large variety of data, with different quality and physical content.

1   Introduction

Aerosols represent an important component of the Earth system with a significant impact on climate (e.g. Seinfeld et al., 2016), weather (e.g. Fan et al., 2016; Gayatri et al., 2017; Marinescu et al., 2017), air quality (e.g. Fuzzi et al., 2015), biogeochemical cycles (e.g. Mahowald, 2011; Mahowald et al., 2017), and health (e.g. Trippetta et al., 2016). A wide variety of aerosols are present in the atmosphere at any time originating from multiple natural (e.g. mineral dust, sea spray, biogenic emissions, volcanic eruptions) and anthropogenic sources (e.g. traffic, industrial activities, biomass burning) and having a large variability

in space and time (e.g. Calvo et al., 2013, and their references therein). This large variety and variability of the aerosols results in uncertainties of their impact. For example, aerosols can influence the microphysical properties of clouds and hence can have an impact on the energy balance, precipitation, and the hydrological cycle.

Aerosols have different scattering and absorption properties depending on their origin, with the largest radiative contribution coming from aerosols with radii between 0.1–1 μm (Satheesh and Krishna, 2005). Seinfeld et al. (2016) indicated that the uncertainties of the radiative forcing associated with the aerosol-cloud interactions have not changed over the last four IPCC reports. Understanding the aerosol sources should reduce the uncertainties of their impact. Detailed knowledge of the aerosol sources can also be used to attribute their role to specific processes, evaluate aerosol models, and design better evidence-based air quality regulations.

Global and local properties of atmospheric aerosols have been extensively observed and measured using both space-borne and ground-based instruments, especially during the last decade. Satellite remote sensing observations have been exploited to characterize aerosol layers and to assess parameterizations for regional and global models (e.g. Amiridis et al., 2010). Global networks of sun/sky radiometers, such as Aerosol Robotic Network (AERONET, Holben et al., 1998) measure the spectral aerosol optical depth (AOD) (e.g. Dubovik et al., 2002; Cattrall et al., 2005; Hamill et al., 2016). The magnitude of the AOD together with the Ångström exponent (i.e. the AOD dependence on the wavelength) can be used to infer the aerosol type, although information about the source are required (e.g. Boselli et al., 2012; Giles et al., 2012). However, the measurements averaged over the entire atmospheric column cannot provide information regarding the vertical distribution of particles.

Active remote sensing instruments, such as lidars, have been used to distinguish between different aerosol types by providing vertical profiles of aerosol optical properties (Müller et al., 2005, 2007; Groß et al., 2013; Nemuc et al., 2013; Samaras et al., 2015; Mărmureanu et al., 2016, 2017), as well to understand the three-dimensional structure and variability in time of the aerosol field (e.g. Freudenthaler et al., 2009; Ansmann et al., 2010; Mattis et al., 2010; Gasteiger et al., 2011a, b). Even if detailed studies of aerosol optical properties have been conducted (e.g. Brock et al., 2016a, b; Palacios-Peña et al., 2018) there are no straightforward links between the optical properties and the aerosol sources given that atmospheric aerosol occurs as a mixture of types (e.g. David et al., 2013), thus difficult to characterize.

Recent advances in atmospheric aerosols measurements have helped to address some of these issues, in particular, to separate different types of aerosols and their mixtures. For example, Burton et al. (2012) analyzed lidar measurements of aerosol parameters (i.e. lidar ratio, depolarization, backscatter color ratio, spectral depolarization ratio) collected by the NASA Langley Research Center airborne High Spectral Resolution Lidar (HSRL, Hair et al., 2008) during measurements campaigns over North America. They showed that these parameters vary with location and with the aerosol type and thus can help to distinguish between different types of aerosols (e.g. HSRL measurements indicated lidar ratio can be used to discriminate between ice and dust and spectral particle depolarization to discriminate between urban and biomass burning aerosols). Another important advancement in the remote sensing of aerosols was the development of ground-based lidar networks, which provide quality assured optical profiles on a large temporal and spatial scale. One such network is the European Aerosol Research Lidar Network (EARLINET) (Pappalardo et al., 2014) established in 2000 with the goal of developing a continental database of the temporal and spatial distribution of aerosols. The EARLINET data are not only relevant for climatological studies, but also for

special events, with strong aerosol influence, such as Saharan dust outbreaks, forest-fire smoke plumes transported over large areas, photochemical smog, and volcano eruptions (Tesche et al., 2009b; Mona et al., 2012; Nicolae et al., 2013; Tesche et al., 2011; Ortiz-Amezcua et al., 2017; Vaughan et al., 2018). Recent efforts have focused on making complementary use of different instruments such as lidar and sun or sky photometry at combined EARLINET and AERONET stations (e.g. Ansmann et al.,

5 2002, 2010; Müller et al., 2010; Alados-Arboledas et al., 2011; Mamouri et al., 2012; Granados-Muñoz et al., 2016; Perone et al., 2016). Several other approaches have been developed by using the combination of ground based measurements with airbore HSRLs lidars and satellite data (e.g. Liu et al., 2002; Tesche et al., 2009b; Omar et al., 2009; Kahn et al., 2010; Burton et al., 2012, 2013; Groß et al., 2013; Burton et al., 2014, 2015; Kahn and Gaitley, 2015; Papagiannopoulos et al., 2016).

All these studies have revealed the existence of a wide variety of aerosols, difficult to be classified due to a series of draw-

10 backs (e.g. many aerosol types have similar optical properties). Another issue in aerosols classification is the difficulty in correlating their optical properties to their sources. In reality, atmospheric aerosols are mixtures from many sources, and data of pure aerosol types are sparse. To address these issues, systematic measurements and intensive measurements campaigns have been performed using different methods for aerosol typing (e.g. Tesche et al., 2009a; Burton et al., 2014), and complementary information such as trajectory and dispersion models analysis to estimate aerosols origin (e.g. Stohl et al., 2003; Stein

15 et al., 2016). Since 2000, EARLINET network has systematically measured the properties of aerosols from different sources over Europe. Intense campaigns, like ACE-Asia (Asian Pacific Regional Aerosol Characterization Experiment, Murayama et al., 2003), SAMUM-1 (Saharan Mineral Dust Experiment, Morocco, Tesche et al., 2009b), SAMUM-2 (Saharan Mineral Dust Experiment, Cape Verde Groß et al., 2011), SALTRACE (Saharan Aerosol Long-range Transport and Aerosol-Cloud interaction Experiment, Groß et al., 2015), ChArMEx/EMEP (Chemistry-Aerosol Mediterranean Experiment, Granados-Muñoz

20 et al., 2016) have helped in understanding the optical properties of aerosols (pure dust and mixtures) or anthropogenic aerosols from industrial areas. Furthermore, recent events, like the eruptions of Eyjafjallajökull in 2010 and Grimsvötn in 2011 offered a rare opportunity to perform studies on the optical properties of volcanic aerosols (e.g. Sicard et al., 2012; Mona et al., 2012; Tesche et al., 2012).

The multitude of instruments and retrievals resulted in an increasing amount of data on aerosols properties to be processed

25 and classified. One possible way of processing large amounts of data with the aim of distinguishing between different aerosols types, is to exploit artificial neural networks (ANNs). Starting from the premise that the best way to distinguish between certain data (e.g. image recognition, speech recognition, medical diagnosis) is the human experience based on learning and education, the ANNs were developed to solve problems in the same way that a human brain might. An ANN represents a mathematical projection of the brain in which the information propagates as a neural influx and it is analyzed. The ANN contain tens to

30 hundreds of neurons divided into multiple layers depending on the data to be classified. The output of the first layer of neurons representing the input to the next layer. The data for analysis must be constrained to a pattern and the ANNs needs to learn to identify this pattern. During the learning process, some weights of the connections between neurons are established. Learning in the case ANNs means to change these weights each time when training data are presented to the network. The change is based on the amount of error in the output compared to the expected result. A comprehensive description of the ANNs theory

35 can be found in Bishop (2000), Picton (2000), and Nielsen (2015).

The ANNs capabilities to classify data have been widely proved in many areas of research (e.g. Jain et al., 2000). Over the last decades, ANNs were used for remote sensing applications such as radars (e.g. Orlandini and Morlini, 2000), microwave radiometers (e.g. Roberts et al., 2010), satellite retrievals (e.g. Ali et al., 2012), multi-angle spectropolarimeters (e.g. Di Noia et al., 2015), nephelometers (e.g. Berdnik and Loikov, 2016) or multiple sources datasets (Gupta and Christopher, 2009; Taylor et al., 2014). In this article, an in-house developed ANN algorithm for aerosol typing is introduced. The algorithm relies on a set of ANNs which are trained to recognize the aerosol type based on typical lidar data products from EARLINET, i.e. three backscatter coefficients (β) at 1064, 532 and 355 nm, two extinction coefficients (α) at 532 and 355 nm, and, optional, one linear particle depolarization (δ) at 532 nm. To distinguish between different aerosol types and their mixtures, the optical data presented to the ANNs have to be characteristic (i.e. to be independent on the density of the particles). Therefore the $3\beta+2\alpha$ (+1δ) lidar data are used at first to compute the intensive properties such as Ångström exponent (AE), color ratios (CR), color indexes (CI), and lidar ratios (LR).

The ability of the ANNs to retrieve the aerosol type depends strongly on the physical content and the uncertainty of the optical inputs, as well as on the structure of the ANN and the training process, including the extent of the dataset used for this purpose. To create a consistent picture of the aerosol types, an aerosol model representing the optical properties of different aerosol was developed. This model is capable of reproducing the observed aerosol properties, and thus can be used to construct a representative and statistically relevant synthetic databases. This synthetic dataset is needed due to sparse observational data sets that are statistically relevant, well-characterized, and representative for the whole spectrum of the aerosol types. The aerosol model was constructed to simulate a large number of lidar measurements (i.e. synthetic dataset) which were then used as input data to training the ANNs. The output data from ANNs consists in the most probable aerosol type within the identified layers.

This article is organized as follows. The aerosol model that was used to generate the synthetic dataset of lidar measurements is described in section 2.1. The synthetic dataset is then used as an input for the ANNs, the core of the aerosol typing algorithm, presented in section 2.2. Sections 2.3 and 2.4 describe the neural network aerosol typing algorithm based on lidar data (NATALI). The comparison between the aerosol model output and the lidar measurements from the previous studies are discussed in section 3.1. Section 3.2 describes the performance of the ANNs. The comparison between the EARLINET-CALIPSO classification and NATALI is presented in section 3.3. Finally, section 4 summarizes this article.

2   Methodology

2.1   The aerosol model

An aerosol model was developed to calculate the optical properties of pure aerosols, those generated by a single source (e.g. dust produced by the deserts, marine particles produces by the oceans). In this article, six classes of pure aerosols are considered: continental, continental polluted, dust, marine, smoke, and volcanic (Table 1). The aerosol model combines the Global Aerosol Data Set (GADS, Koepke et al., 1997) along with the T-Matrix numerical method (Waterman, 1971; Mishchenko et al., 1996) to compute iteratively the intensive optical properties of each aerosol type. The chemical composition of each pure aerosol type

was picked up from OPAC (Optical Properties of Aerosols and Clouds) software package (Hess et al., 1998). The chemical composition of each aerosol type was varied in certain limits (the limits are detailed in Table 2 and refer to particle number density mixing ratios), in order to reproduce the large variety of particles present in the atmosphere. The synthetic database developed using the aerosol model is built for 350, 550, and 1000 nm sounding wavelengths. These wavelengths were selected from the 61 wavelengths (0.25–40 μm) of OPAC for which the microphysical characteristics of the aerosols are available from GADS. The selected wavelengths are then re-scaled to the usual lidar wavelengths (i.e. 355, 532 and 1064 nm) using an Ångström exponent equal to one. This was considered a valid assumption for all aerosol types, taking into account the small difference between the lidar and the model wavelengths. If required, the aerosol model can be extended for other wavelengths.

Each pure aerosol type is built as an internal mixture of basic components which do not interact physically or chemically, having different mixing ratios. The basic components are picked up from OPAC: water soluble, insoluble, soot, mineral (nucleation, accumulation, coarse), sulphates, sea salt (accumulation, coarse). The GADS database is used for the microphysical properties of each component (Koepke et al., 1997). However, with the current values of the complex refractive index of soot in GADS values greater than 1.2 for the Ångström exponent (550/350 nm) cannot be achieved for smoke and continental polluted types. Based on the findings of Schnaiter et al. (2003) and Henriksen et al. (2007), a typical value of 1.41 was considered for the real part of the refractive index, instead of 1.75 as it is currently in GADS.

In the aerosol model, particles were considered as spheroids with different aspect ratios (i.e. the ratio of the polar to equatorial lengths) to simulate the aerosols anisotropy (Table 2). Dust and volcanic aerosols were considered oblate (i.e. aspect ratio <1). Also, the proportion of soot was increased to counterbalance for the low hematite (iron oxide) content, consistent with Dubovik et al. (2002) and Gasteiger et al. (2011b).

Starting from the microphysical properties (i.e. mode radius, width of the log-normal distribution, number density, density and mass concentration) of each component, the microphysical properties of the pure aerosol were calculated by varying in certain limits the critical component (i.e. its number density mixing ratio), while the total mixture is normalized to one (Table 2). The mixing ratio of the aerosol components is given by

$$\mu_j = \frac{N_j}{N_t}; \;\; j = \overline{1, NC} \tag{1}$$

where NC represents the number of components, N_t is the total number of particles, N_j is the number of particles for component j, and the boundary condition is given by

$$\sum_j \mu_j = 1; \;\; j = \overline{1, NC} \tag{2}$$

For each wavelength selected in the aerosol model, the real and the imaginary parts of the complex refractive index, were determined with the Lorentz-Lorentz model

$$\frac{m_p^2 - 1}{m_p^2 + 2} = \sum_j \mu_j \cdot \frac{m_j^2 - 1}{m_j^2 + 2}; \;\; j = \overline{1, NC} \tag{3}$$

where m_p and m_j represent the complex refractive index for the particle and for the j components of the aerosol mixture, respectively. The aerosol radius (r_p) is calculated with the following equation

$$r_p = \sqrt[3]{\sum_j \mu_j \cdot \left(r_j^{mod}\right)^3}; \quad j = \overline{1, NC} \tag{4}$$

where r_j^{mod} is the radius of the component j with respect to relative humidity (RH). The aerosol size distribution ($n(r)$) as a
5   function of aerosol radius (r) assuming mono-modal log-normal distribution is given by

$$n(r) = \frac{1}{\sqrt{2 \cdot \pi} \cdot \ln(\sigma_p) \cdot r} \cdot e^{\left[-\left(\frac{\ln(r) - \ln(r_p)}{\sqrt{2} \cdot \ln(\sigma)}\right)^2\right]}, \quad j = \overline{1, NC} \tag{5}$$

where $\sigma_p = \sum_j \mu_j \cdot \sigma_j$ represents the width of the distribution for aerosols, and σ_j is the width of the distribution for component j (computed as the standard deviation of the log of the distribution with r_j^{mod} mod radius).

Using the calculated microphysical properties with the T-matrix code (Mishchenko and Travis, 1994), the effective cross
10   section for the particle scattering (C_{sca}) and extinction (C_{ext}) as well as the scattering matrix elements (phase functions) were obtained. These parameters are further used to determine (for a single particle) the aerosol optical parameters. The extinction coefficient (α) is determined from Eq. 6 and the backscatter coefficient (β) from Eq. 7, where F_{11} is the first element of the scattering matrix (phase function).

$$\alpha = \int_{R_{min}}^{R_{max}} C_{ext} \cdot n(r) \, dr \tag{6}$$

15   $$\beta = \int_{R_{min}}^{R_{max}} C_{sca} \cdot \frac{F_{11}(180°)}{4 \cdot \pi} \cdot n(r) \, dr \tag{7}$$

The integration domain ($R_{min}:R_{max}$), for which the effective radius, the extinction coefficient and scattering coefficient are calculated, covers medium size particles whose radius is between 0.1–5.0 μm that contribute to the scattering and extinction of light. The radius was not increased further due to computing time limitations and model design limitations (i.e. the code used for the calculation of the optical parameters for spheroids does not achieve the convergence for non-spherical particles).
20   However, the latter limitation is not considered critical for the range of the lidar wavelengths.

The single scattering albedo ($\overline{\omega}$) is yielded as the ratio of the scattering and extinction effective cross sections

$$\overline{\omega} = \frac{C_{sca}}{C_{ext}} \tag{8}$$

The lidar ratio (LR) is determined by the following relationship

$$LR = \frac{4 \cdot \pi}{\omega \cdot F_{11}(180°)} \tag{9}$$

while the particle linear depolarization (δ) is calculated based on the elements of the scattering matrix

$$\delta = \frac{F_{11} - F_{22}}{F_{11} + F_{22}} \tag{10}$$

5   The algorithm is iterated for each composition, wavelength and RH value until the entire selected domain is covered. The domain represents the range in which the parameters are varied (e.g. the domain for the wavelength is [350, 500, 1000 nm]; the domain for RH is [50%, 70%, 80%, 90%]; the domain for the number density mixing ratios for each component of each pure aerosol type is listed in Table 2). The algorithm generates the properties and mixing ratio of each component, the optical and microphysical properties of the aerosol, for each wavelength, each RH value, and each composition.

10   Four classes of RH (i. e., 50%, 70%, 80%, 90%) are considered, out of the eight classes in OPAC. The high RH values (i.e. above 90%) were excluded in order to avoid ambiguous results related to activation of the hygroscopic particles. Dry particles, those with 0% RH, considered as too rarely present in the ambient atmosphere. For a better representation of the particle growth, the OPAC RH classes were linearly interpolated with a 1% step for pure types and 5% for mixed types and linearly extrapolated down to 40%. Thus, within 40%–90% range the hygroscopic growth is considered linear for all pure aerosols

15   included in the model.

Even considering a certain variation of the aerosol composition and of RH, the simulated optical parameters are not covering the whole range of measured values. This is partly due to the limitations of the model itself and partly due to the various uncertainties associated to the measurements either due to the instrument (e.g. biases, calibration) or due to the data treatment (e.g. algorithms applied in pre-processing to correct or average raw signals, algorithms used to calculate data products). Optical

20   parameters calculated from lidar measurements are reported in the EARLINET database as the mean value (x_{med}) and associated uncertainty (absolute error, Δx). Optical parameters calculated from synthetic data do not carry this uncertainty, therefore a fix relative error was considered which was multiplied with the value to obtain the absolute error (uncertainty). For the actual retrieval of aerosol type was considered that any value between (x_{med}-uncertainty) and (x_{med}+uncertainty) was possible, therefore the algorithm was applied for all these values with a certain step (i.e. the finesse). The output is a "bundle" of possible

25   aerosol types, with a dimension equal to the finesse. A compromise should be made between the finesse and computing time.

Based on the values reported in the literature (e.g. Ansmann et al., 2002; Freudenthaler et al., 2009), a large uncertainty is associated with the extinction coefficient derived with the Raman method, mainly due to noisy Raman lidar signals (i.e. the relative error reported in the lidar measurements is 30%–150% and the fixed relative error considered in the synthetic data is 50%). Particle depolarization is very sensitive to the calibration, both for the raw signals of the two channels and for the

30   backscattering. Thus, the values for particle depolarization have also a significant uncertainty (i.e. the relative error reported in the lidar measurements is 2%–50% and the fixed relative error considered in the synthetic data is 30%). The backscatter

coefficient calculated from the combination of Raman-elastic channels is less sensitive (i.e. the relative error reported in the lidar measurements is 10%–50% and the fixed relative error considered in the synthetic data is 20%). Even in the case of High Spectral Resolution Lidars, were the extinction and the backscattering are independently calculated, the cross-talk between the Mie and the Rayleigh channels is still introducing systematic errors, larger for the extinction than for the backscattering.

5   The relative errors considered here are 50% for the extinction, 20% for the backscattering and 30% for the depolarization. Note that these values were assumed to be inclusive to mimic high-precision but also moderate-precision retrieved parameters. Although for the microphysical inversion the recommended maximum value for the uncertainty of the optical parameters is 20% (Müller et al., 1999a, b), this is not critical for aerosol classification, as long as a relevant number of parameters is provided (e.g. measured lidar ratios and Ångström exponent are required).

10   Table 4 shows the aerosol types considered in this study: six pure aerosols, seven mixtures of two pure aerosols, and two mixtures of three pure aerosols. The mixtures were obtained by linear combination of pure aerosols properties. The mixtures composed of only two pure types were considered not sufficient. For example, transcontinental transport involves at least three types of pure aerosols (e.g. transport from Africa to Europe can result in a mixture of continental, dust, and marine aerosols). Adding marine changes drastically the optical properties of the mixtures of two pure aerosols. Thus, mixtures of three aerosol 15   types were considered, especially those containing marine type. From the total number of possible mixtures of two and three aerosols (i.e. 35 mixtures) only those that are most frequently observed and can still be distinguished were selected (i.e. 9 mixtures, see Table 4). This selection of mixtures was also a compromise between the time-performance of the algorithm and the minimum number of output aerosol types considered significant in atmosphere.

The generated optical properties of pure aerosols and mixtures serve as a basis for the determination of the extinction 20   Ångström exponent (κ_{ext}) and the backscatter Ångström exponent (κ_{bsca}), also referred as color ratio, for each wavelength combination (Fig. 1). Thus, the Ångström exponent is given by the relationship

$$\kappa_{ext} = -\frac{\ln\left(\alpha_{\lambda_1}/\alpha_{\lambda_2}\right)}{\ln\left(\lambda_1/\lambda_2\right)} \tag{11}$$

Similarly, the backscatter Ångström coefficient (color index) can be determined using the equation

$$\kappa_{bsca} = -\frac{\ln\left(\beta_{\lambda_1}/\beta_{\lambda_2}\right)}{\ln\left(\lambda_1/\lambda_2\right)} \tag{12}$$

25   After the calculation of the spectral parameters for pure and mixed aerosols, the synthetic data are used as an input for the Artificial Neural Networks.

2.2   The architecture Artificial Neural Networks

The ANNs can be calibrated, or "trained", for a specific purpose. Here, ANNs are trained to classify aerosols using solely the lidar intensive properties as input data, without any complementary information. The ANNs used here to classify aerosols were 30   developed using *NeuroSolutions* a neural network development environment. Several ANNs architectures have been explored:

Multilayer Perceptron (MLP), Jordan/Elman Network (JE), Generalized Feed Forward Network (GFF), Self-Organizing Feature Maps (SOFM), Recurrent Neural Network (RNN). Each ANN architecture contains several hidden layers and different learning rules. Each layer is composed of a vector of processing elements of identical parameters (e.g. TanhAxon, SigmoidAxon, LinearTanhAxon) with an associated learning rule and learning parameters.

5   No significant improvement in the classification of the aerosols has been achieved for different types of processing elements on the ANN structure. Thus, TanhAxon were subsequently used. The TanhAxon applies a bias and a hyperbolic tangent function (i.e. $\tanh$) to each neuron in the layer and replaces a part of the $\tanh$ by a line having a slope β. The values of each neuron are forced to be in the interval -1 and 1. For TanhAxon the activation function is defined as

$$f\left(x_i, w_i\right) = \tanh\left[x_i^{lin}\right] \tag{13}$$

10   where $x_i^{lin} = \beta x_i$ and and w_i is bias vector.

Supervised training has been used to train ANNs. Thus, sets of input and output parameters have been being successively presented to the networks for around 1000 epochs (i.e. one forward pass and one backward pass of all the training examples) per training cycle. Backpropagation is the most common form to train ANNs with more than one hidden layer. In the case of backpropagation, the weights on input elements are changed based on their previous value and a correction term. This training

15   approach has been used also for the design of the NATALI ANN: the input data being continuously presented to the ANN and the output compared with the known aerosol type from synthetic database, in order to adjust the weights until the desired result is achieved. The optimal values of weights and the minimum errors have been the parameters taken into account for the testing process. The minimum classification errors were 75% for more than 80% of the measured data and 75% for more than 90% of the synthetic data."

20   Several learning rules have been tested: *Momentum*, *Conjugate Gradient Descent*, *Step*, and *Levenberg–Marquardt*. The *Momentum* learning rule is a simple and efficient approach in comparison with a standard gradient. It provides the gradient descent with some inertia, depending on the momentum parameter, which gives the smoothness of the gradient estimation. The momentum parameter is the same for all processing elements on a layer. The *Conjugate Gradient* has no parameters to be adjusted (e.g. learning rates, momentum parameter) and is faster and more accurate with respect to the standard backpropagation.

25   The other two rules, the *Step* rule—a type of standard gradient descent algorithm that allows the user to set a default step size for all weights within the activation component, and the *Levenberg–Marquardt* rule—which gives a numerical solution to the problem of minimizing a nonlinear function, were found inadequate for aerosol typing purpose. The *Step* rule recognized all aerosols types after first training cycle, but its active performance was low. The *Levenberg–Marquardt* algorithm blocked after several epoch.

30   The cross-validation and test set methods have been used to stop the learning process and to assess the performances. Cross-validation monitors the error for a set of data and stops training when this error begins to increase. After a full process of training, in our case 5–10 training cycles, a testing set of data is presented to the ANN and the network output is compared with the known aerosol type from the synthetic database. In total 68 ANN structures have been explored, starting from the simplest

(reduced number of hidden layers) to the complex ones, in order to compromise between the minimum possible time of training and testing, and avoiding "saturation" effects. Examples of 6 pure, 7 double-component mixtures, and 2 triple-component mixtures obtained within the 68 explored ANN are presented in Table 4. For the selection of the ANNs, the synthetic database has been split randomly into data used to train the ANN (70% of all synthetic datasets), data used to test ANN (20% of all synthetic datasets), and data used for validation (10% of all synthetic datasets). In the training process, datasets are presented to the ANN with the correct answer. The training is performed iteratively until the testing and validation classification errors are below 25% (Fig. 2). A finer adjustment of the weighting coefficients is done during the testing process. The last 10% of the data are presented to the ANN without the known result in order to validate the optimum training process and the capability of the network to classify new data inputs.

Three basic ANNs (adjusted to accommodate all data) have been chosen as appropriate to classify in parallel the multi-wavelength lidar data, for both high and low-resolution classification: the Jordan-Elman with 6 or 8 hidden layers, and the Generalized Feedforward with 10 hidden layers (Table 3). The selected types of ANNs classify the aerosols based on the response with higher confidence (i.e. the probability of having one of the aerosols types). The ANNs have been trained using 3500 samples for each aerosol type and successive training sessions until the best weights are reached (i.e. the classification process is ended, and the classification errors are low).

2.3 The typing algorithm

Following the methodology described in section 2.1 and taking into account the uncertainty threshold of each optical parameter, a bundle of inputs for each measured or simulated aerosol layer was generated. Answers with low confidence are filtered out (e.g. by using a threshold of minimum 0.7 confidence). The correct answer is selected based on a statistical approach considering two criteria: a) which answer has a higher confidence; b) which answer is more stable over the uncertainty range.

The inputs parameters for NATALI are typical data products from EARLINET database: backscatter coefficient (β) profiles at 1064, 532 and 355 nm, extinction coefficient (α) profiles at 532 and 355 nm, and, optionally, linear particle depolarization (δ) profile at 532 nm. The identification of aerosol types is not always possible, due to the dependence on the physical content (i.e. with or without δ), and do the quality of the optical data (i.e. calibration, uncertainty). For these reasons three classification schemes, with different aerosol type resolutions, are used (Table 4). First, when particle depolarization is available and all optical parameters are provided with good quality (uncertainty of the aerosol extinction coefficient $\leq$50%, uncertainty of the aerosol backscatter coefficient $\leq$20%, uncertainty of the particle linear depolarization ration $\leq$30%), the typing is performed in high resolution (AH). This means that the mixtures can be resolved and the number of outputs is fourteen (i.e. pure with minimum 90%, mixtures of two, and mixtures of three pure aerosol types). Second, when particle depolarization is available and the optical parameters have a high uncertainty (uncertainty of the aerosol extinction coefficient >50%, uncertainty of the aerosol backscatter coefficient >20%, uncertainty of the particle linear depolarization ration >30%), the typing is performed in low resolution (AL). In this case, the number of outputs is six (i.e. pure with maximum 30% traces of other types). Third, when the particle depolarization is not available, the typing is performed in low resolution, meaning again that the mixtures cannot be resolved. In this case, the predominant aerosol type is retrieved for 5 outputs (pure with maximum 30% traces of other types),

whereby if only spectral parameters are provided, the volcanic type cannot be distinguished from dust nor continental polluted, and therefore excluded as output.

The three ANNs (Table 3) were developed for three classification schemes (Table 4) to increase the confidence of the aerosol typing. A voting procedure selects the most probable answer out of the three (possibly different) individual returns. The selection is made based on the confidence level of the ANN outputs, and stability over the uncertainty range (i.e. the percentage of agreement for values between error limits).

2.4 The NATALI code

The Neural Network Aerosol Typing Algorithm based on LIdar data (NATALI) developed in the Python programming language is built on three modules: a) an input module to prepare the inputs in the specific format of the ANNs, b) a typing module to run the ANNs and decide on the most probable aerosol type and c) an output module to save the results and logs. The input module reads the lidar files in EARLINET NetCDF format, checks for the availability of all required parameters ($\beta1064$, $\beta532$, $\beta355$, $\alpha532$, $\alpha355$, and optionally $\delta532$ nm), identifies the layer geometrical boundaries, and calculates within each layer the mean intensive optical parameters (i.e. Ångström exponent, color indexes color ratios, lidar ratios, particle linear depolarization ratio) and their associated uncertainty) (Fig. 3).

The layer boundaries are calculated by applying the gradient method on the 1064 nm backscatter coefficient profile (Belegante et al., 2014). The inflexion points of the second derivative of the profile data, computed with the Savitzky-Golay filter, give the top and the bottom of the layers. The window size of the cubic Savitzky-Golay filter, which be modified by the user, has a default value of 700 m. The filter was applied twice to obtain the second derivative. A signal-to-noise ratio filter is applied at this point, making sure the said ratio is at least 5. The layer boundaries are moved towards the median height until the SNR criteria is met; if the criteria cannot be satisfied with a layer height greater than 300m, the layer is discarded. Gross or fine structure of the aerosol layers is revealed by a higher or lower value of the adjustable smoothing parameter (FINESSE). The layers with thickness of more than 300 m are considered, whereby the intensive optical properties and their uncertainties are computed for the middle of each layer in the range of at least 200 m thickness, to exclude the margins likely affected by the smoothing

Several filters are applied to the data, and only layers which pass the following criteria are further considered for typing:

  – availability of all necessary intensive optical parameters

  – values of the intensive optical parameters are between acceptable limits (Table 5)

  – the relative error of each intensive optical parameter is lower than 50%

For each layer and for each intensive optical parameter, the input module generates an adjustable number N of values x with uncertainties (Δx) in the range $x - \Delta x$ and $x + \Delta x$. Data are than scrambled considering that any combination has a similar probability to describe the reality. The cluster of possible combinations of intensive optical parameters is then converted into the ANN input format.

The typing module runs in parallel the ANNs for each dataset representing a layer, and applies the voting procedure to identify the most probable aerosol type. In case that the depolarization is available, the module runs in parallel six ANNs, three for high resolution (i.e. A1H, A2H, A3H) and three for low resolution typing (i.e. A1L, A2L, A3L). The probable aerosol type is provided by the high resolution ANNs, while the predominant type is provided by the low resolution ANNs. As such, if typing in high resolution fails due the data quality, the user has still access to information in low resolution. If the depolarization is not available, the module runs in parallel three ANNs (i.e. B1L, B2L, B3L), and returns only the most probable predominant aerosol type (volcanic overlaps, in all existing parameters, completely with dust or continental polluted type and cannot be retrieved in low-resolution). The output module prepares and saves the files in two formats csv files and human-readable (telegrams), and writes a log. The csv files and the telegrams contain the identification of the data sets for which typing is performed and provides for each identified layer the following parameters:

- identification of the datasets for which the typing was performed

- for each identified layer

  - the geometrical top and bottom

  - the intensive optical parameters and associated uncertainties

  - the aerosol type retrieved by each ANN, and the number of agreements

  - the most probable type selected with the voting procedure (in low and high resolution separately, if is the case)

  - the type of the ANN delivering the result (i.e. 1, 2, or 3)

  - comments, generally referring to situations when optical data did not passed the quality criteria, or errors in the retrieval procedure

The NATALI code additional information (e.g. run time, run parameters, network error messages) are included in the telegrams. The software structure resembles the three module approach described earlier: an input module ($nt\_input.py$), a typing module ($nt\_typing.py$), and an output module ($nt\_output.py$). The three modules are coordinated by the $natali.py$ script, which contains the high-level algorithm and calls the required module routines/codes.

3   Results

The performances of the algorithm were tested in three steps. Firstly, the outputs of the aerosol model were compared with the literature for the values of the intensive optical parameters for each aerosol type considered in this study (section 3.1). Secondly, the ANNs were selected based on their performances during the learning phase, and also by comparison with a known reference (i.e. synthetic data) (section 3.2). Thirdly, the complete NATALI algorithm was tested by comparing the retrieved aerosol types with the EARLINET-CALIPSO classification (section 3.3).

3.1 Comparison of the aerosol model with the literature

Synthetic aerosol optical properties, i.e. Ångström exponent (AE_{550_350}), color ratios (CR_{550_350} and, CR_{1000_550})), lidar ratios (LR_{350} and LR_{550}), and linear particle depolarization ratio at 550 nm (DEP_{550}) generated by the developed aerosol model have been compared with the measured intensive parameters for the six classes of pure aerosol. The comparison with the previous literature was only possible for pure types because the properties of mixed aerosols are computed base on a linear progression of the corresponding optical properties for two pure types. As shown in Table 6, the synthetic data are in general in very good agreement with the values reported in previous studies (i.e. the range of synthetic values is between the minimum and maximum values reported in the literature). Synthetic values lower than those observed are for continental rural type (AE_{550_350}), continental polluted (CR_{1000_500}), and dust (CR_{1000_500}). Synthetic values greater than those from the literature are for continental rural (LR_{350}) and volcanic (DEP_{550}) type. The reasons for these discrepancies are many-fold. In some cases, values reported in the literature have high uncertainties because of natural variability, improper calibration and retrieval. The aerosol model has also some limitations, e.g. due to spheroidal model and mono-modal log-normal distribution considered.

When comparing the aerosol model with the results from the previous studies, the changes in OPAC concerning the hygroscopic growth need be considered (e.g. Zieger et al., 2013). These changes have not been implemented here, because, at the time when this study was conducted, the new OPAC hygroscopicity was not available. However, the changes in OPAC are not expected to produce major changes in the aerosol model, considering the large uncertainties introduced in the model to simulate the observations.

In Figure 4 comparisons between the synthetic data for pure aerosol obtained from the model and the measurements obtained by Groß et al. (2013) are provided. Based on the Airborne High Spectral Resolution Lidar (HSRL) data and in situ measurements of aerosol microphysical and optical properties collected during a series of measurements campaigns in 1998 (Lindenberg Aerosol Characterization Experiment, LACE), 2006 (The Saharan Mineral Dust Experiment, Morocco, SAMUM-1), and 2008 (The Saharan Mineral Dust Experiment, Cape Verde Islands, SAMUM-2 and European integrated project on Aerosol Cloud Climate, EUCAARI), Groß et al. (2013) developed an aerosol classification scheme for six aerosol types and aerosol mixtures (i.e. Saharan mineral dust, Saharan dust mixtures, Canadian biomass burning aerosol, African biomass burning mixture, anthropogenic pollution aerosol, and marine aerosol). The aerosol typing based on the lidar ratio and the linear depolarization ratio at 550 nm, show, in general, a good agreement between the synthetic data and the observations at 532 nm from Groß et al. (2013) (Fig. 4a and d), especially for smoke/biomass burning, industrial and marine type. The continental and volcanic aerosol are not represented in the measurements, therefore were not compared. Dust presents lower values for depolarization for the synthetic data (Fig. 4b and e), but similar values for the lidar ratio (Fig. 4c and f). Clusters were identified both in synthetic and observational data, which means that for pure aerosols the combination of extinction, backscatter and depolarization at one wavelength could be sufficient for the ANN training.

Wandinger et al. (2016) provided a synthesis of ground-based observations of lidar ratio and particle linear depolarization at 355 nm for different aerosol types (i.e. dust, smoke, pollution, marine, aerosol, volcanic ash) and mixtures, collected during a series of measurements campaigns, i.e. PollyXT measurements at Cape Verde (Groß et al., 2011), at EARLINET stations

of Leipzig and Munich (Groß et al., 2012), in the Amazon Basin (Baars et al., 2012), and on board Polarstern over the North Atlantic (Kanitz et al., 2013) (Fig. 5a). The synthetic data show a wider spread because of large uncertainty accepted for the input parameters. Very high values for the linear depolarization for smoke in the Aerosol CCI (European Space Agency Aerosol Climate Change Initiative) could not be achieved in the aerosol model (Fig. 5b).

5   When the entire output of the aerosol model is considered (i.e. 14 aerosol types) there is a high overlapping between clusters, in particular for mixtures, due to the built-in uncertainty (Fig. 6a). Smoke and continental pollution almost completely overlap (Fig. 6a), which is consisted with measurements reported in literature (Table 6). This makes the typing challenging. The importance of particle depolarization shown relatively recently (e.g. Freudenthaler et al., 2009), can improve the aerosol typing (Fig. 6b). Particle depolarization contributes to the identification of complex mixtures and to the differentiation between

10 mineral and volcanic particles. The main issue for particle depolarization is calibration, recently addressed (e.g. McCullough et al., 2017; Belegante et al., 2018) and thus few datasets satisfy the depolarization ratio quality criteria for aerosol typing. However, even without particle depolarization information, the low-resolution typing can identify the aerosol predominant types in a mixture.

3.2   ANNs performance

15 Figure 7 shows the overall performances of the ANNs for the high-resolution typing (i.e. A1H, A2H, A3H) and low-resolution typing (i.e. A1L, A2L, A3L). In high-resolution typing at least 70% of the aerosol types defined (i.e. 10 out of 14) should be correctly assessed in more than 75% of the cases with a confidence higher than 0.7. In low-resolution typing at least 70% of the predominant aerosol types (i.e. 4 out of 5) defined should be correctly assessed in more than 65% of the cases with a confidence higher than 0.7.

20   The aerosol type is recognized in more than 96% of all cases in high-resolution typing (Fig. 7a). The missed cases are, in general, due to the complete overlap between the input parameters. For example, continental smoke is classified as smoke in 22% of the missed cases (i.e. 1.9% of the total number of cases); continental dust is classified as dust in 9% of the missed case (i.e. 0.3% of the total number of cases). Note that 33% of the missed cases (1.2% of the total number of cases) are in classified as unknown.

25   The predominant aerosol is recognized in more than 91% of the cases in low-resolution typing (Fig. 7b). Most of the missed cases are due to the ANNs not being able to distinguish between continental and smoke, and continental polluted (i.e 36% of the missed cases representing 3.2% of the total number of cases), and continental, smoke and marine, and continental polluted (i.e. 35% of the missed cases, 3.1% of all cases). Continental polluted and marine is sometimes identified by the ANNs as continental (i.e. 27% of all missed cases, 2.4% of the total number of cases).

30   A3H and A3L were the best performing ANNs, however not always with a high confidence level (Fig. 7). A2H and A2L have the lowest performances, but they can help in the certain cases, for example, in recognizing continental dust aerosols. The voting procedure is not always providing the right answer, as, for example, when A3H is providing the correct typing but its confidence level is low.

The dependence of the aerosol typing on RH shows that the performances of the ANNs are decreased with an increase in RH, only for continental smoke and dust continental for high-resolution typing (Fig. 8) and for continental smoke and mixed smoke for low-resolution typing (Fig. 8). Pure aerosol types are recognized for all values of RH. For coastal polluted, the relative humidity increase results in an increase of typing performance. Overall, lower performances are obtained in low-resolution

5 typing.

3.3 Comparison with EARLINET-CALIPSO classification

Observational data from EARLINET Data Base (www.earlinet.org), related to the CALIPSO (Cloud-aerosol Lidar and Infrared Pathfinder Satellite Observation) overpasses over different EARLINET observational sites, were compared with the synthetic data obtained from the aerosol model. The EARLINET-CALIPSO Data base (Pappalardo et al., 2010), covers the data of

10 2000–2018 and includes a total of 718 cases and 21 aerosol and cloud types. Only 13 of these cases contained all of the necessary parameters (i.e. 3 backscatters, 2 extinctions and 1 depolarization). In general, the missing parameter is the particle depolarization. To increase the number of cases, the particle depolarization was added assuming values reported in literature as typical for the corresponding aerosol type. This way, 105 cases containing all needed parameters were obtained. The cases for which all parameters were within 20% of relative error were selected (63 cases), whereby 57 corresponded to known aerosol

15 types.

Additionally, profiles available at the EARLINET site in Bucharest/Măgurele, established by the Romanian National Institute for Research and Development of Optoelectronics (INOE), were used to increase the validation measurements sample. The INOE database contains 464 measurement sets performed with the multi-wavelength Raman depolarization Lidar (RALI, Belegante et al., 2011) between June 2012–September 2014. About 44.6% of measurements were conducted at nighttime

20 (thus, included the Raman-derived extinction coefficient profiles). Out of these, 871 processed layers containing backscattering, extinction and particle depolarization profiles averaged over one hour. Only layers with significant loads (i.e. layers for which the uncertainty of the retrieved optical parameters is bellow the limits accepted by the algorithm) were selected, for which all intensive parameters were retrieved with an accuracy higher than 20%. Mean values within each layer were computed, excluding the edges of the layers, where the smoothing introduces large errors due to the high gradients. For each layer, the

25 Ångström exponent, color ratio, color index, lidar ratio, and linear particle depolarization ratio were computed. Thresholds were then used to estimate at first glance the type of aerosol which resulted in a dataset with 311 layers acceptable by the algorithm, out of which for only 182 layers auxiliary data was available. Auxiliary data were used to compare the results of the typing.

The time-series of lidar measurements (532 nm volume depolarization and 355, 532, and 1064 nm range corrected signals)

30 were used to identify the aerosol layers. The identification of the aerosol source was based on 96 hours backward trajectories using HYSPLIT (Stein et al., 2016). The source was assumed to originate at the region where the trajectory was closest to the ground, providing guidance for identifying possible emission sources. The rainfall along trajectory was used as an indicator for likely wet deposition. A synoptic diagnosis of the main meteorological filed (e.g. pressure, geopotential height, temperature, relative humidity, wind), based on NCEP/NCAR Reanalysis (Kalnay et al., 1996), was used as to confirm the

aerosols trajectories and also to determine the type of atmospheric circulation, weather regimes and weather phenomena along the trajectories.

Figure 9 shows the comparison between the aerosol typing based on the aerosol model (synthetic data) and the EARLINET-CALIPSO and INOE database (observed data). The large spread of the measured parameters is caused by the mixtures of three components, incorrect calibration or inappropriate estimation of aerosol type. On the other hand, the sparse observational data lead to apparently incomplete clusters. No conclusions can be drawn for marine aerosols, as they are not represented in the observational data.

Low values are observed in the Ångström exponent for several cases of dust polluted and smoke categories, as well as low values for the 532 nm lidar ratio are seen for several cases of continental and continental dust indicating a small portion of marine particles. This is most likely due to the fact that particles are transported over a short distance above the sea before reaching the target, and thus are misclassified. The high values for the Ångström exponent for some of the marine mineral aerosols indicate a mixture with smoke. High values of the Ångström exponent, greater than 1.8 as it is measured for smoke and continental polluted, failed to be simulated. Actually, in general the agreement of classification of the simulated data and the real observations is very good, given all limitations discussed.

4 Conclusions

The NATALI algorithm is based on the ability of specialized ANNs to resolve the overlapping values of the intensive optical parameters calculated for each identified layer in the multiwavelength Raman lidar profiles. The ANNs were trained using synthetic data, for which a new aerosol model was developed. Aerosols were considered spheroids and built up using OPAC-defined internal mixtures, with the associated microphysical properties retrieved from GADS. The intensive optical properties obtained from this model were compared to literature and found to be consistent with observations. Variability of the optical properties was achieved by considering different number mixing ratios and relative humidities. In addition, the uncertainty of the observations was included as a prerequisite hypothesis in order to match the lidar data. These requirements have added to the complexity of the ANNs selected to make the retrieval, because of the significant overlap of the input values for the intensive optical parameters. Although the linear particle depolarization ratio is a crucial parameter in separating aerosol types, the depolarization methodology is still maturing and only a few lidar stations provide this parameter with an acceptable accuracy. Thus, two parallel typing schemes were developed a) a high-resolution typing scheme that allows the identification of 14 aerosol mixtures if the LPDR is available in the input data files, and b) a low-resolution typing scheme that allows the identification of five predominant aerosol types when LPDR is not provided. For each scheme, three ANNs are run simultaneously. Then a voting procedure is applied to select the most probable answer. The ANNs were selected out of 68 structures tested, as having the best performances for the aerosol typing. The voting is based on the confidence of the retrieval for each of the three ANNs, and the stability of the retrieval over the uncertainty range. A series of tests showed that considering the variation with the RH from the beginning helped in making the retrieval stable for different atmospheric conditions. Also, considering the 50% uncertainty for the input data gave realistic retrieval or aerosol, making possible the retrieval of aerosol types when

using medium-quality lidar data, which is currently the case for research lidar networks. Without depolarization, the retrieval is much less certain, especially for mixtures, and questionable results were flagged. Spectral characteristics of volcanic aerosols are very similar with those of mineral dust and/or continental polluted, and this type cannot be distinguished if the LPDR is not provided.

5   The whole algorithm has been integrated into a Python code, available as source code and executable on the NATALI website (http://natali.inoe.ro/resources.html/software). The software accommodates lidar profiles—$3\beta + 2\alpha$ $(+1\delta)$—in the EARLINET data format. The NATALI is user-friendly; a user guide is provided. However, it is important that the user understands the outputs and the limitations of the algorithm, i.e. the results are strongly dependent on the quality input data, and the outputs should be understood accordingly. Although the neural network is able to recognize the pattern of noisy data, such pattern has

10   to be present and correct, otherwise, the result of the retrieval will be incorrect. The NATALI algorithm was able to

- recognize the aerosol types (high resolution, 14 types) in more than 70% of the cases for good quality optical data (i.e. the uncertainty of the intensive optical parameters of less than 20%)

- recognize the predominant aerosol types (low resolution, 6 or 5 types) in more than 70% of the cases for medium and high quality optical data (i.e. the uncertainty of the intensive optical parameters less than 50%)

15   - provide stable responses for RH up to 70%, even higher for less hygroscopic aerosols

- provide results that are comparable in high and low resolution, considering the correspondence of the types defined

Furthermore, the computing time of the algorithm is relatively short due to the optimization of the Python code. The algorithm has side-applications, for example, it can be applied to test the quality of the optical data, identify incorrect calibration or incorrect cloud screening (e.g. Nicolae et al., 2018). Blind tests on EARLINET data samples showed the capability of this

20   tool to retrieve the aerosol type from a large variety of data, with different quality and physical content. More complex datasets (e.g. availability of LPDR at 355 and/or 1064 nm) will not produce improvements with the current software because ANNs are specifically trained for $3\beta + 2\alpha + 1\delta$ datasets. However, the ANNs can be trained with more complete datasets, which potentially can lead to better scores, especially in the case of mixtures. Moreover, a similar approach could be used for any other optical instrument (e.g. photometry) as long as the physical content of the input optical parameters is sufficiently rich.

25   *Code availability.* The NATALI software together with a user guide are available from http://natali.inoe.ro/resources.html/software.

Author contributions. D. Nicolae carried out the research design and developed the aerosol typing algorithm. J. Vasilescu designed the artificial neural networks and conducted the statistical analysis of the output. C. Talianu developed the aerosol model. I. Binietoglou carried out the comparison of the results with the previous research and the testing of the typing algorithm. V. Nicolae developed the code for the aerosol typing algorithm. All the authors participated in the interpretation of the results and the writing and editing process.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The work presented in this paper was performed in the frame of the project Neural network Aerosol Typing Algorithm based on LIdar data (NATALI) funded by ESA under contract 4000110671/14/I-LG. Also, this project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 654109 ACTRIS-2, grant agreement No. 692014 ECARS, and Core National Program PN2018 33N/16.03.2018 funded by the Ministry of Research and Innovation. Camelia Talianu was also supported by the Austrian Science Fund FWF, Project M 2031, Meitner-Programm.

References

Alados-Arboledas, L., Müller, D., Guerrero-Rascado, J. L., Navas-Guzman, F., Perez-Ramirez, D., and Olmo, F. J.: Optical and microphysical properties of fresh biomass burning aerosol retrieved by Raman lidar, and star-and sun-photometry, Geophys. Res. Lett., 38, 1–5, https://doi.org/10.1029/2010GL045999, 2011.

Ali, A., Amin, S. E., Ramadan, H. H., and Tolba, M. F.: Enhancement of OMI aerosol optical depth data assimilation using artificial neural network, Neural Computing & Applications, 23, 2267–2279, https://doi.org/10.1007/s00521-012-1178-9, 2012.

Amiridis, V., Giannakaki, E., Balis, D. S., Gerasopoulos, E., Pytharoulis, I., Zanis, P., Kazadzis, S., Melas, D., and Zerefos, C.: Smoke injection heights from agricultural burning in Eastern Europe as seen by CALIPSO, Atmos. Chem. Phys., 10, 11 567–11 576, https://doi.org/10.5194/acp-10-11567-2010, 2010.

Ansmann, A., Wagner, F., Müller, D., Althausen, D., Herber, A., von Hoyningen-Huene, W., and Wandinger, U.: European pollution outbreaks during ACE 2: Optical particle properties inferred from multiwavelength lidar and star-Sun photometry, J. Geophys. Res., 107, 4259, https://doi.org/10.1029/2001JD001109, 2002.

Ansmann, A., Tesche, M., Groß, S., Freudenthaler, V., Seifert, P., Hiebsch, A., Schmidt, J., Wandinger, U., Mattis, I., Müller, D., and Wiegner, M.: The 16 April 2010 major volcanic ash plume over central Europe: EARLINET lidar and AERONET photometer observations at Leipzig and Munich, Germany, Geophys. Res. Lett., 37, L13 810, https://doi.org/10.1029/2010GL043809, 2010.

Baars, H., Ansmann, A., Althausen, D., Engelmann, R., Heese, B., Müller, D., Artaxo, P., Paixao, M., Pauliquevis, T., and Souza, R.: Aerosol profiling with lidar in the Amazon Basin during the wet and dry season, J. Geophys. Res., 117, D21 201, https://doi.org/10.1029/2012JD018338, 2012.

Belegante, L., Talianu, C., Nemuc, A., and Nicolae, D.: Detection of local weather events from multiwavelength lidar measurements during the EARLI09 campaign, Rom. J. Phys., 56(3–4), 484–494, 2011.

Belegante, L., Nicolae, D., Nemuc, A., Talianu, C., and Derognat, C.: Retrieval of the boundary layer height from active and passive remote sensors. Comparison with a NWP model, Acta Geophysica, 62, 276–289, https://doi.org/10.2478/s11600-013-0167-4, 2014.

Belegante, L., Bravo-Aranda, J. A., Freudenthaler, V., Nicolae, D., Nemuc, A., Ene, D., Alados-Arboledas, L., Amodeo, A., Pappalardo, G., D'Amico, G., Amato, F., Engelmann, R., Baars, H., Wandinger, U., Papayannis, A., Kokkalis, P., and Pereira, S. N.: Experimental techniques for the calibration of lidar depolarization channels in EARLINET, Atmos. Meas. Tech., 11, 1119–1141, https://doi.org/10.5194/amt-11-1119-2018, 2018.

Berdnik, V. V. and Loikov, V. A.: Neural networks for aerosol particles characterization, J. Quant. Spectrosc. Ra., 184, 135–145, https://doi.org/10.1016/j.jqsrt.2016.06.034, 2016.

Bishop, C.: Neural Networks for Pattern Recognition, Clarendon Press, 2000.

Boselli, A., Caggiano, R., Cornacchia, C., Madonna, F., Mona, L., Macchiato, M., Pappalardo, G., and Trippetta, S.: Multi year sunphotometer measurements for aerosol characterization in a Central Mediterranean site, Atmos. Res., 104–105, 98–110, https://doi.org/10.1016/j.atmosres.2011.08.002, 2012.

Brock, C. A., Wagner, N. L., Anderson, B. E., Attwood, A. R., Beyersdorf, A., Campuzano-Jost, P., Carlton, A. G., Day, D. A., Diskin, G. S., Gordon, T. D., Jimenez, J. L., Lack, D. A., Liao, J., Markovic, M. Z., Middlebrook, A. M., Ng, N. L., Perring, A. E., Richardson, M. S., Schwarz, J. P., Washenfelder, R. A., Welti, A., Xu, L., Ziemba, L. D., and Murphy, D. M.: Aerosol optical properties in the southeastern United States in summer–Part 1: Hygroscopic growth, Atmos. Chem. Phys., 16, 4987–5007, https://doi.org/10.5194/acp-16-4987-2016, 2016a.

Brock, C. A., Wagner, N. L., Anderson, B. E., Beyersdorf, A., Campuzano-Jost, P., Day, D. A., Diskin, G. S., Gordon, T. D., Jimenez, J. L., Lack, D. A., Liao, J., Markovic, M. Z., Middlebrook, A. M., Perring, A. E., Richardson, M. S., Schwarz, J. P., Welti, A., Ziemba, L. D., and Murphy, D. M.: Aerosol optical properties in the southeastern United States in summer–Part 2: Sensitivity of aerosol optical depth to relative humidity and aerosol parameters, Atmos. Chem. Phys., 16, 5009–5019, https://doi.org/10.5194/acp-16-5009-2016, 2016b.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements–methodology and examples, Atmos. Meas. Tech., 5, 73–98, https://doi.org/10.5194/amt-5-73-2012, 2012.

Burton, S. P., Ferrare, R. A., Vaughan, M. A., Omar, A. H., Rogers, R. R., Hostetler, C. A., and Hair, J. W.: Aerosol classification from airborne HSRL and comparisons with the CALIPSO vertical feature mask, Atmos. Meas. Tech., 6, 1397–1412, https://doi.org/10.5194/amt-6-1397-2013, 2013.

Burton, S. P., Vaughan, M. A., Ferrare, R. A., and Hostetler, C. A.: Separating mixtures of aerosol types in airborne High Spectral Resolution Lidar data, Atmos. Meas. Tech., 7, 419–436, https://doi.org/10.5194/amt-7-419-2014, 2014.

Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D. B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of linear particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar, Atmos. Chem. Phys., 15, 13453–13473, https://doi.org/10.5194/acp-15-13453-2015, 2015.

Calvo, A. I., Alves, C., Castro, A., Pont, V., Vicente, A., and Fraile, R.: Research on aerosol sources and chemical composition: Past, current and emerging issues, Atmos. Res., 120–121, 1–28, https://doi.org/10.1016/j.atmosres.2012.09.021, 2013.

Cattrall, C., Reagan, J., Thome, K., and Dubovik, O.: Variability of aerosol and spectral lidar and backscatter and extinction ratios of key aerosol types derived from selected Aerosol Robotic Network locations, J. Geophys. Res., 110, D10S11, https://doi.org/10.1029/2004JD005124, 2005.

David, G., Thomas, B., Nousiainen, T., Miffre, A., and Rairoux, P.: Retrieving simulated volcanic, desert dust and sea-salt particle properties from two/three-component particle mixtures using UV-VIS polarization lidar and T matrix, Atmos. Chem. Phys., 13, 6757–6776, https://doi.org/10.5194/acp-13-6757-2013, 2013.

Di Noia, A., Hasekamp, O. P., van Harten, G., Rietjens, J. H. H., Smit, J. M., Snik, F., Henzing, J. S., de Boer, J., Keller, C. U., and Volten, H.: Use of neural networks in ground-based aerosol retrievals from multi-angle spectropolarimetric observations, Atmos. Meas. Tech., 8, 281–299, https://doi.org/10.5194/amt-8-281-2015, 2015.

Dubovik, O., Holben, B. N., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanre, D., and Slutsker, I.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, J. Atmos. Sci., 59, 590–608, https://doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2, 2002.

Fan, J., Wang, Y., Rosenfeld, D., and Liu, X.: Review of aerosol-cloud interactions: Mechanisms, significance, and challenges, J. Atmos. Sci., 73, 4221–4252, https://doi.org/10.1175/JAS-D-16-0037.1, 2016.

Fernández, A. J., Sicard, M., Costa, M. J., Guerrero-Rascado, J. L., Gómez-Amo, J. L., Molero, F., Barragán, R., Bortoli, D., Bedoya-Velásquez, A. E., Utrillas, M. P., Salvador, P., noz, M. J. G.-M., Potes, M., Ortiz-Amezcua, P., Martínez-Lozano, J. A., mano, B. A., noz Porcar, C. M., Salgado, R., Román, R., Rocadenbosch, F., Salgueiro, V., Benavent-Oltra, J. A., Rodríguez-Gómez, A., Alados-Arboledas, L., Comerón, A., and Pujadas, M.: February 2017 extreme Saharan dust outbreak in the Iberian Peninsula: from lidar-derived optical properties to evaluation of forecast models, Atmos. Chem. Phys. Discussions, 16, https://doi.org/10.5194/acp-2018-370, 2018.

Freudenthaler, V., Esselborn, M., Wiegner, M., Heese, B., Tesche, M., Ansmann, A., Müller, D., Althausen, D., Wirth, M., Fix, A., Ehret, G., Knippertz, P., Toledano, C., Gasteiger, J., Garhammer, M., and Seefeldne, M.: Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006, Tellus B, 61, 165–179, https://doi.org/10.3402/tellusb.v61i1.16821, 2009.

Fuzzi, S., Baltensperger, U., Carslaw, K., Decesari, S., Denier van der Gon, H., Facchini, M. C., Fowler, D., Koren, I., Langford, B., Lohmann, U., Nemitz, E., Pandis, S., Riipinen, I., Rudich, Y., Schaap, M., Slowik, J. G., Spracklen, D. V., Vignati, E., Wild, M., Williams, M., and Gilardoni, S.: Particulate matter, air quality and climate: lessons learned and future needs, Atmos. Chem. Phys., 15, 8217–8299, https://doi.org/10.5194/acp-15-8217-2015, 2015.

Gasteiger, J., Groß, S., Freudenthaler, V., and Wiegner, M.: Volcanic ash from Iceland over Munich: Mass concentration retrieved from ground-based remote sensing measurements, Atmos. Chem. Phys., 11, 2209–2223, https://doi.org/10.5194/acp-11-2209-2011, 2011a.

Gasteiger, J., Wiegner, M., Groß, S., Freudenthaler, V., Toledano, C., Tesche, M., and Kandler, K.: Modeling lidar-relevant optical properties of complex mineral dust aerosols, Tellus B, 63, 725–741, https://doi.org/10.1111/j.1600-0889.2011.00559.x, 2011b.

Gayatri, K., Patade, S., and Prabha, T. V.: Aerosol–Cloud interaction in deep convective clouds over the Indian Peninsula using spectral (bin) microphysics, J. Atmos. Sci., 74, 3145–3166, https://doi.org/10.1175/JAS-D-17-0034.1, 2017.

Giannakaki, E., Balis, D. S., Amiridis, V., and Zerefos, C.: Optical properties of different aerosol types: seven years of combined Raman-elastic backscatter lidar measurements in Thessaloniki, Greece, Atmospheric Meas. Tech., 3, 569–578, https://doi.org/10.5194/amt-3-569-2010, 2010.

Giles, D. M., Holben, B. N., Eck, T. F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R. R., Thompson, A. M., and Schafer, J. S.: An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions, J. Geophys. Res., 117, D17 203, https://doi.org/10.1029/2012JD018127, 2012.

Granados-Muñoz, M. J., Navas-Guzmán, F., Guerrero-Rascado, J. L., Bravo-Aranda, J. A., Binietoglou, I., Pereira, S. N., Basart, S., Baldasano, J. M., Belegante, L., Chaikovsky, A., Comerón, A., D'Amico, G., Dubovik, O., Ilic, L., Kokkalis, P., Muñoz-Porcar, C., Nickovic, S., Nicolae, D., Olmo, F. J., Papayannis, A., Pappalardo, G., Rodríguez, A., Schepanski, K., Sicard, M., Vukovic, A., Wandinger, U., Dulac, F., and Alados-Arboledas, L.: Profiling of aerosol microphysical properties at several EARLINET/AERONET sites during the July 2012 ChArMEx/EMEP campaign, Atmos. Chem. Phys., 16, 7043–7066, https://doi.org/10.5194/acp-16-7043-2016, 2016.

Groß, S., Tesche, M., Freudenthaler, V., Toledano, C., Wiegner, M., Ansmann, A., Althausen, D., and Seefeldner, M.: Characterization of saharan dust, marine aerosols and mixtures of biomass burning aerosols and dust by means of multi-wavelength depolarization- and Raman-measurements during SAMUM-2, Tellus B, 63, 706–724, https://doi.org/10.1111/j.1600-0889.2011.00556.x, 2011.

Groß, S., Volker, F., Matthias, W., Josef, G., Alexander, G., and Franziska, S.: Dual-wavelength linear depolarization ratio of volcanic aerosols: Lidar measurements of the Eyjafjallajökull plume over Maisach, Germany, Atm. Env., 48, 85–96, https://doi.org/10.1016/j.atmosenv.2011.06.017, 2012.

Groß, S., Esselborn, M., Weinzierl, B., Wirth, M., Fix, A., and Petzold, A.: Aerosol classification by airborne high spectral resolution lidar observations, Atmos. Chem. Phys., 13, 2487–2505, https://doi.org/10.5194/acp-13-2487-2013, 2013.

Groß, S., Freudenthaler, V.and Schepanski, K., Toledano, C., Schäfler, A., Ansmann, A., , and Weinzierl, B.: Optical properties of long-range transported Saharan dust over Barbados as measured by dual-wavelength depolarization Raman lidar measurements, Atmos. Chem. Phys., 15, 11 067–11 080, https://doi.org/10.5194/acp-15-11067-2015, 2015.

Groß, S., Gasteiger, J., Freudenthaler, V., Müller, T., Sauer, D., Toledano, C., and Ansmann, A.: Saharan dust contribution to the Caribbean summertime boundary layer–a lidar study during SALTRACE, Atmos. Chem. Phys., 16, 11 535–11 546, https://doi.org/10.5194/acp-16-11535-2016, 2016.

Gupta, P. and Christopher, S. A.: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach, J. Geophys. Res., 114, D20 205, https://doi.org/10.1029/2008JD011497, 2009.

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

Hamill, P., Giordano, M., Ward, C., Giles, D., and Holben, B.: An AERONET-based aerosol classification using the Mahalanobis distance, Atm. Env., 140, 213–233, https://doi.org/dx.doi.org/10.1016/j.atmosenv.2016.06.002, 2016.

Henriksen, T., Ring, T., Call, D., Eddings, E., and Sarofim, A.: Determination of soot refractive index as a function of height in an inverse diffusion flame, in: 5th US Combustion Meeting 25–28 March 2007, pp. 1795–1803, San Diego, California, 2007.

Hess, M., Koepke, P., and Schult, I.: Optical properties of aerosols and clouds: The software package OPAC, Bull. Amer. Meteor. Soc., 79, 831–844, https://doi.org/doi:10.1175/1520-0477(1998)0792.0.CO;2, 1998.

Holben, B. N., Eck, T. F., Slutsker, I., Tanré, D., Buis, J. P., Setzer, A., Vermote, E., Reagan, J. A., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET - A federated instrument network and data archive for aerosol characterization, Rem. Sens. Environ., 66, 1–16, https://doi.org/10.1016/s0034-4257(98)00031-5, 1998.

Jain, A. K., Duin, R. P. W., and Mao, J.: Statistical Pattern Recognition: A Review, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 4–37, https://doi.org/10.1109/34.824819, 2000.

Janicka, L., S.Stachlewska, I., Veselovskii, I., and Baars, H.: Temporal variations in optical and microphysical properties of mineral dust and biomass burning aerosol derived from daytime Raman lidar observations over Warsaw, Poland, Atmos. Environ., 169, 162–174, https://doi.org/10.1016/j.atmosenv.2017.09.022, 2017.

Kahn, R. A. and Gaitley, B. J.: Analysis of global aerosol type as retrieved by MISR, J. Geophys. Res. Atmos., 120, 4248–4281, https://doi.org/10.1002/2015JD023322, 2015.

Kahn, R. A., Gaitley, B. J., Garay, M. J., Diner, D. J., Eck, T. F., Smirnov, A., and Holben, B. N.: Multiangle Imaging Spectro-Radiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res., 115, D23 209, https://doi.org/10.1029/2010JD014601, 2010.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bull. Amer. Meteor. Soc., 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Kanitz, T., Ansmann, A., Engelmann, R., and Althausen, D.: North-south cross sections of the vertical aerosol distribution over the Atlantic Ocean from multiwavelength Raman/polarization lidar during Polarstern cruises, J. Geophys. Res. Atmos., 118, 2643–2655, https://doi.org/10.1002/jgrd.50273, 2013.

Koepke, P., Hess, M., Schult, I., and Shettle, E. P.: Global Aerosol Data Set, Report No. 243, Max-Planck-Institut für Meteorologie, Hamburg, ISSN 0937-1060, 1997.

Liu, Z., Sugimoto, N., and Murayama, T.: Extinction-to-backscatter ratio of Asian dust observed with high-spectral-resolution lidar and Raman lidar, Appl. Opt., 41, 2760–2767, https://doi.org/10.1364/AO.41.002760, 2002.

Mahowald, N.: Aerosol indirect effect on biogeochemical cycles and climate, Science, 334, 794–796, https://doi.org/10.1126/science.1207374, 2011.

Mahowald, N. M., Scanza, R., Brahney, J., Goodale, C. L., Hess, P. G., Moore, J. K., and Neff, J.: Aerosol deposition impacts on land and ocean carbon cycles, Current Climate Change Reports, 3, 16–31, https://doi.org/10.1007/s40641-017-0056-z, 2017.

Mamouri, R. E., Papayannis, A., Amiridis, V., Müller, D., Kokkalis, P., Rapsomanikis, S., Karageorgos, E. T., Tsaknakis, G., Nenes, A., Kazadzis, S., and Remoundaki, E.: Multi-wavelength Raman lidar, sun photometric and aircraft measurements in combination with inversion models for the estimation of the aerosol optical and physico-chemical properties over Athens, Greece, Atmos. Meas. Tech., 5, 1793–1808, https://doi.org/10.5194/amt-5-1793-2012, 2012.

Marinescu, P. J., van den Heever, S., Saleeby, S. M., Kreidenweis, S. M., and DeMott, P. J.: The microphysical roles of lower-tropospheric versus midtropospheric aerosol particles in mature-stage MCS precipitation, J. Atmos. Sci., 74, 3657–3678, https://doi.org/10.1175/JAS-D-16-0361.1, 2017.

Mattis, I., Siefert, P., Müller, D., Tesche, M., Hiebsch, A., T. Kanitz, T., Schmidt, J., Finger, F., Wandinger, U., and Ansmann, A.: Volcanic aerosol layers observed with multiwavelength Raman lidar over central Europe in 2008–2009, J. Geophys. Res. Atmos., 115, D00L04, https://doi.org/10.1029/2009JD013472, 2010.

McCullough, E. M., Sica, R. J., Drummond, J. R., Nott, G., Perro, C., Thackray, C. P., Hopper, J., Doyle, J., Duck, T. J., and Walker, K. A.: Depolarization calibration and measurements using the CANDAC Rayleigh–Mie–Raman lidar at Eureka, Canada, Atmos. Meas. Tech., 10, 4253–4277, https://doi.org/10.5194/amt-10-4253-2017, 2017.

Mishchenko, M. I. and Travis, L. D.: T-matrix computations of light scattering by large spheroidal particles, Opt. Commun., 109, 16–21, https://doi.org/10.1016/0030-4018(94)90731-5, 1994.

Mishchenko, M. I., Travis, L. D., and Mackowski, D. W.: T-matrix computations of light scattering by nonspherical particles: A review, J. Quant. Spectrosc. Radiat. Transfer, 55, 535–575, https://doi.org/doi.org/10.1016/0022-4073(96)00002-7, 1996.

Mona, L., Liu, Z., Müller, D., Omar, A., Papayannis, A., Pappalardo, G., Sugimoto, N., and Vaughan, M.: Lidar measurements for desert dust characterization: An overview, Advances in Meteorology, 2012, ID356 265, https://doi.org/doi:10.1155/2012/356265, 2012.

Mărmureanu, L., Vasilescu, J., Nemuc, A., Nicolae, D., and Belegante, L.: Aerosol characterization based on chemical composition and optical properties, Rom. J. Phys., 61(9–10), 1635–1650, 2016.

Mărmureanu, L., Vasilescu, J., Ştefănie, H., and Talianu, C.: Chemical and optical characterization of submicronic aerosol sources, Environmental Engineering and Management Journal, 16, 2165–2172, 2017.

Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: Theory, Appl. Opt., 38, 2346–2357, https://doi.org/10.1364/ao.38.002346, 1999a.

Müller, D., Wandinger, U., and Ansmann, A.: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: Simulation, Appl. Opt., 38, 2358–2368, https://doi.org/10.1364/ao.38.002358, 1999b.

Müller, D., Mattis, I., Wandinger, U., Ansmann, A., Althausen, D., and Stohl, A.: Raman lidar observations of aged Siberian and Canadian forest fire smoke in the free troposphere over Germany in 2003: Microphysical particle characterization, J. Geophys. Res., 110, D17 201, https://doi.org/10.1029/2004JD005756, 2005.

Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., and Pisani, G.: Aerosol-type dependent lidar ratios observed with Raman lidar, J. Geophys. Res. Atmos., 112, D16 202, https://doi.org/10.1029/2006JD008292, 2007.

Müller, D., Weinzierl, B., Petzold, A., Kandler, K., Ansmann, A., Müller, T., Tesche, M., Freudenthaler, V., Esselborn, M., Heese, B., Althausen, D., Schladitz, A., Otto, S., and Knippertz, P.: Mineral dust observed with AERONET sun photometer, Raman lidar, and in situ instruments during SAMUM 2006: shape-independent particle properties, J. Geophys. Res., 115, D07 202, https://doi.org/10.1029/2009JD012520, 2010.

Murayama, T., Masonis, S. J., Redemann, J., Anderson, T. L., Schmid, B., Livingston, J. M., Russell, P. B., Huebert, B., Howell, S. G., McNaughton, C. S., Clarke, A., Abo, M., Shimizu, A., Sugimoto, N., Yabuki, M., Kuze, H., Fukagawa, S., Maxwell-Meier, K., Weber, R. J., Orsini, D. A., Blomquist, B., Bandy, A., and Thornton, D.: An intercomparison of lidar-derived aerosol optical properties with airborne measurements near Tokyo during ACE-Asia, J. Geophys. Res., 108, 8651, https://doi.org/10.1029/2002JD003259, 2003.

5 Murayama, T., Müller, D., Wada, K., Shimizu, A., Sekiguchi, M., and Tsukamoto, T.: Characterization of Asian dust and Siberian smoke with multi-wavelength Raman lidar over Tokyo, Japan in spring 2003, Geophys. Res. Lett., 31, L23 103, https://doi.org/10.1029/2004GL021105, 2004.

Nemuc, A., Vasilescu, J., Talianu, C., Belegante, L., and Nicolae, D.: Assessment of aerosol's mass concentrations from measured linear particle depolarization ratio (vertically resolved) and simulations, Atmos. Meas. Tech., 6, 3243–3255, https://doi.org/10.5194/amt-6-3243-10 2013, 2013.

Nicolae, D., Nemuc, A., Müller, D., Talianu, C., Vasilescu, J., Belegante, L., and Kolgotin, A.: Characterization of fresh and aged biomass burning events using multiwavelength Raman lidar and mass spectrometry, J. Geophys. Res. Atmospheres, 118, 2956–2965, https://doi.org/10.1002/jgrd.50324, 2013.

Nicolae, D., Talianu, C., Vasilescu, J., Nicolae, V., and Stachlewska, I. S.: Strengths and limitations of the NATALI code for aerosol typ-15 ing from multiwavelength Raman lidar observations, The 28th International Laser Radar Conference (ILRC 28), 176, 14 123–14 146, https://doi.org/10.1051/epjconf/201817605005, 2018.

Nielsen, M. A.: Neural Networks and Deep Lerning, Determination Press, 2015.

Noh, Y., Kim, Y., Choi, B., and Murayama, T.: Aerosol lidar ratio characteristics measured by a multi-wavelength Raman lidar system at Anmyeon Island, Korea, Atmos. Res., 86, 76–87, https://doi.org/D10.1016/j.atmosres.2007.03.006, 2007.

20 Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., Lee, K. L., Hostetler, C. A., Kittaka, C., Rogers, R. R., Kuehn, R. E., and Z., L.: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, J. Atmos. Ocean Tech., 26, 1994–2014, https://doi.org/10.1175/2009jtecha1231.1, 2009.

Orlandini, S. and Morlini, I.: Artificial neural network estimation of rainfall intensity from radar observations, J. Geophys. Res., 105, 24 849–24 861, https://doi.org/10.1029/2000JD900408, 2000.

25 Ortiz-Amezcua, P., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Benavent-Oltra, J. A., Böckmann, C., Samaras, S., Stachlewska, I. S., Janicka, Ł., Baars, H., Bohlmann, S., and Alados-Arboledas, L.: Microphysical characterization of long-range transported biomass burning particles from North America at three EARLINET stations, Atmos. Chem. Phys., 17, 5931–5946, https://doi.org/10.5194/acp-17-5931-2017, 2017.

Palacios-Peña, L., Baró, R., Baklanov, A., Balzarini, A., Brunner, D., Forkel, R., Hirtl, M., Honzak, L., López-Romero, J. M., Pérez, J. L., 30 Pirovano, G., San José, R., Schröder, W., Werhahn, J., Wolke, R., Zabkar, R., and Jiménez-Guerrero, P.: An assessment of aerosol optical properties from remote sensing observations and regional chemistry-climate coupled models over Europe, Atmos. Chem. Phys., 18, 5021–5043, https://doi.org/10.5194/acp-18-5021-2018, 2018.

Papagiannopoulos, N., Mona, L., Alados-Arboledas, L., Amiridis, V., Baars, H., Binietoglou, I., Bortoli, D., D'Amico, G., Giunta, A., Guerrero-Rascado, J. L., Schwarz, A., Pereira, S., Spinelli, N., Wandinger, U., Wang, X., , and Pappalardo, G.: CALIPSO climatological 35 products: evaluation and suggestions from EARLINET, Atmos. Chem. Phys., 16, 2341–2357, https://doi.org/10.5194/acp-16-2341-2016, 2016.

Papayannis, A., Amiridis, V., Mona, L., Tsaknakis, G., Balis, D., Bösenberg, J., Chaikovski, A., DeTomasi, F., Grigorov, I., Mattis, I., Mitev, V., Müller, D., Nickovic, S., Pérez, C., Pietruczuk, A., Pisani, G., Ravetta, F., Rizi, V., Sicard, M., Trickl, T., Wiegner, M., Gerding, M.,

Mamouri, R. E., D'Amico, G., and Pappalardo, G.: Systematic lidar observations of Saharan dust over Europe in the frame of EARLINET (2000–2002), J. Geophys. Res., 113, D10 204, https://doi.org/10.1029/2007JD009028, 2008.

Pappalardo, G., Wandinger, U., Mona, L., Hiebsch, A., Mattis, I., Amodeo, A., Ansmann, A., Seifert, P., Linné, H., Apituley, A., Arboledas, L. A., Balis, D., Chaikovsky, A., D'Amico, G., De Tomasi, F., Freudenthaler, V., Giannakaki, E., Giunta, A., Grigorov, I., Iarlori, M.,
5   Madonna, F., Mamouri, R., Nasti, L., Papayannis, A., Pietruczuk, A., Pujadas, M., Rizi, V., Rocadenbosch, F., Russo, F., Schnell, F., Spinelli, N., Wang, X., and Wiegner, M.: EARLINET correlative measurements for CALIPSO: First intercomparison results, J. Geophys. Res., 115, D00H19, https://doi.org/10.1029/2009JD012147, 2010.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., J., B., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable
10 European aerosol lidar network, Atmospheric Meas. Tech., 7, 2389–2409, https://doi.org/10.5194/amt-7-2389-2014, 2014.

Perone, M. and Bulizzi, P.: Mediterranean aerosol typing by integrating three-wavelength lidar and sun photometer measurements, Environ. Sci. Pollut. Res., 23, 14 123–14 146, https://doi.org/10.1007/s11356-016-6575-7, 2016.

Perone, M., , and Bulizzi, P.: Mediterranean aerosol typing by integrating three-wavelength lidar and sun photometer measurements, Environ Sci Pollut Res., 23, 14 123–14 146, https://doi.org/10.1007/s11356-016-6575-7, 2016.

15 Picton, P.: Neural Networks for Pattern Recognition, 2nd ed. Palgrave, 2000.

Roberts, J. B., Clayson, C. A., Robertson, F. R., and Jackson, D. L.: Predicting near-surface atmospheric variables from Special Sensor Microwave/Imager using neural networks with a first-guess approach, J. Geophys. Res., 115, D19 113, https://doi.org/10.1029/2009JD013099, 2010.

Sakai, T., Shibata, T., Kwon, S. A., Kim, Y. S., Tamura, K., and Iwasaka, Y.: Free tropospheric backscatter, depolarization ratio, and relative
20 humidity measured with the Raman lidar at Nagoya in 1994–1997: Contributions of aerosols from the Asian Continent and the Pacific Ocean, Atmos. Environ., 34, 431–442, https://doi.org/doi.org/10.1016/S1352-2310(99)00328-3, 2002.

Samaras, S., Nicolae, D., Böckmann, C., Vasilescu, J., Binietoglou, I., Labzovskii, L., and F. Țoancă et al., A. P.: Using Ramanlidar-based regularized microphysical retrievals and Aerosol Mass Spectrometer measurements for the characterization of biomass burning aerosols, J. Comput. Phys., 299, 156–174, https://doi.org/10.1016/j.jcp.2015.06.045, 2015.

25 Satheesh, S. K. and Krishna, K.: Radiative effects of natural aerosols: A review, Atmospheric Environment, 39, 2089–2110, https://doi.org/10.1016/j.atmosenv.2004.12.029, 2005.

Schnaiter, M., Horvath, H., Möhler, O., Naumann, K. H., Saathoff, H., and Schöck, O. W.: UV-VIS-NIR spectral optical properties of soot and soot-containing aerosols, J. Aerosol Sci., 34, 1421–1444, https://doi.org/doi.org/10.1016/S0021-8502(03)00361-6, 2003.

Seinfeld, J. H., Bretherton, C., Carslaw, K. S., Coe, H., DeMott, P. J., Dunlea, E. J., Feingold, G., Ghan, S., Guenther, A. B., Kahn, R.,
30 Kraucunas, I., Kreidenweis, S. M., Molina, M. J., Nenes, A., Penner, J. E., Prather, K. A., Ramanathan, V., Ramaswamy, V., Rasch, P. J., Ravishankara, A. R., Rosenfeld, D., Stephens, G., and Wood, R.: Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system, Proc. Natl. Acad. Sci. USA, 113(21), 5781–5790, https://doi.org/10.1073/pnas.1514043113, 2016.

Sicard, M., Guerrero-Rascado, J. L., Navas-Guzmán, F., Preißler, J., Molero, F., Tomás, S., Bravo-Aranda, J. A., Comerón, A., Rocadenbosch, F., Wagner, F., Pujadas, M., , and Alados-Arboledas, L.: Monitoring of the Eyjafjallajökull volcanic aerosol plume over the Iberian
35 Peninsula by means of four EARLINET lidar stations, Atmos. Chem. Phys., 12, 3115–3130, https://doi.org/10.5194/acp-12-3115-2012, 2012.

Stachlewska, I. S., Zawadzka, O., and Engelmann, R.: Effect of Heat Wave Conditions on Aerosol Optical Properties Derived from Satellite and Ground-Based Remote Sensing over Poland, Remote Sens., 9(11), 1199, https://doi.org/10.3390/rs9111199, 2017.

Stachlewska, I. S., Samson, M., Zawadzka, O., Harenda, K. M., Janicka, L., Poczta, P., Szczepanik, D., Heese, B., Wang, D., Borek, K., Tetoni, E., Proestakis, E., Siomos, N., Nemuc, A., Chojnicki, B. H., Markowicz, K. M., Pietruczuk, A., Szkop, A., Althausen, D., Stebel, K., Schuettemeyer, D., and Zehner, C.: Modification of local urban aerosol properties by long-range transport of biomass burning aerosol, Remote Sens., 10, 412, https://doi.org/10.3390/rs10030412, 2018.

5 Stein, A. F., Draxler, R. R., Rolph, G. D., Stunder, B. J. B., Cohen, M. D., and Ngan, F.: NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull. Amer. Meteor. Soc., 96, 2059–2077, https://doi.org/10.1175/BAMS-D-14-00110.1, 2016.

Stohl, A., Forster, C., Eckhardt, S., Spichtinger, N., Huntrieser, H., Heland, J., Schlager, H., Wilhelm, S., Arnold, F., and Cooper, O.: A backward modelling study of intercontinental pollution transport using aircraft measurements, J. Geophys. Res., 108, 4370, https://doi.org/10.1029/ 2002JD002862, 2003.

10 Taylor, M., Kazadzis, S., Tsekeri, A., Gkikas, A., and Amiridis, V.: Satellite retrieval of aerosol microphysical and optical param- eters using neural networks: a new methodology applied to the Sahara desert dust peak, Atmos. Meas. Tech., 7, 3151–3175, https://doi.org/10.5194/amt-7-3151-2014, 2014.

Tesche, M., A., A., Müller, D., Althausen, D., Engelmann, R., Freudenthaler, V., and Groß, S.: Vertically resolved separation of dust and smoke over Cape Verde by using multiwavelength Raman and polarization lidars during Saharan Mineral Dust Experiment 2008, J.

15 Geophys. Res., 114, D13 202, https://doi.org/10.1029/2009JD011862, 2009a.

Tesche, M., Ansmann, A., Müller, D., Althausen, D., Mattis, I., Heese, B., Freudenthaler, V., Wiegner, M., Esselborn, M., Pisani, G., and Knippertz, P.: Vertical profiling of Saharan dust with Raman lidars and airborne HSRL in southern Morocco during SAMUM, Tellus B, 61, 144–164, https://doi.org/10.1111/j.1600-0889.2008.00390.x, 2009b.

Tesche, M., Müller, D., Gross, S., Ansmann, A., Althausen, D., Freudenthaler, V., Weinzierl, B., Veira, A., and Petzold, A.: Optical

20 and microphysical properties of smoke over Cape Verde inferred from multiwavelength lidar measurements, Tellus B, 63, 677–694, https://doi.org/10.3402/tellusb.v63i4.16362, 2011.

Tesche, M., Glantz, P., Johansson, C., Norman, M., Hiebsch, A., Ansmann, A., Althausen, D., Engelmann, R., and Seifert, P.: Volcanic ash over Scandinavia originating from the Grímsvötn eruptions in May 2011, J. Geophys. Res., 117, D09 201, https://doi.org/10.1029/2011JD017090, 2012.

25 Trippetta, S., Sabia, S., and Caggiano, R.: Fine aerosol particles (PM1): Natural and anthropogenic contributions and health risk assessment, Air Quality, Atmosphere & Health, 9, 621–629, https://doi.org/10.1007/s11869-015-0373-0, 2016.

Vaughan, G., Draude, A. P., Ricketts, H. M. A., Schultz, D. M., Adam, M., Sugier, J., and Wareing, D. P.: Transport of Canadian forest fire smoke over the UK as observed by lidar, Atmos. Chem. Phys., 18, 11 375–11 388, https://doi.org/10.5194/acp-18-11375-2018, 2018.

Veselovskii, I., Dubovik, O., Kolgotin, A., Lapyonok, T., Di Girolamo, P., Summa, D., Whiteman, D. N., Mishchenko, M., and Tanré, D.:

30 Application of randomly oriented spheroids for retrieval of dust particle parameters from multiwavelength lidar measurements, J. Geophys. Res., 115, D21 203, https://doi.org/10.1029/2010JD014139, 2010.

Vlăduțescu, D. V., Wu, Y., Gross, B., Charles, L., Moshary, F., and Ahmed, S.: Aerosol hygroscopic models based on in situ measurements and lidar retrievals, in: 5th WSEAS Int. Conf. on Environment, Ecosystems and Development, Tenerife, Spain, 2007.

Wandinger, U., Müller, D., Böckmann, C., Althausen, D., Matthias, V., Böösenberg, J., Weiß, V., Fiebig, M., Wendisch, M., Stohl, A., and

35 Ansmann, A.: Optical and microphysical characterization of biomass–burning and industrial pollution aerosols from multiwavelength lidar and aircraft measurements, J. Geophys. Res., 107, D218 125, https://doi.org/10.1029/2000JD000202, 2002.

Wandinger, U., Baars, H., Engelmann, R., Hünerbein, A., Horn, S., Kanitz, T., Donovan, D., van Zadelhoff, G., Daou, D., Fischer, J., von Bismarck, J., Filipitsch, F., Docter, N., Eisinger, M., Lajas, D., and Wehr, T.: HETEAC: The Aerosol Classification Model for EarthCARE, EPJ Web of Conferences, 119, 01 004, https://doi.org/10.1051/epjconf/201611901004, 2016.

Wang, W., Gong, W., Mao, F., Pan, Z., and Liu, B.: Measurement and Study of Lidar Ratio by Using a Raman Lidar in Central China, Int. J. Environ. Res. Public Health, 13, 508, https://doi.org/10.3390/ijerph13050508, 2016.

Waterman, P. C.: Symmetry, unitarity, and geometry in electromagnetic scattering, Phys. Rev., 3D, 825–839, https://doi.org/doi.org/10.1103/PhysRevD.3.825, 1971.

Zieger, P., Fierz-Schmidhauser, R., Weingartner, E., and Baltensperger, U.: Effects of relative humidity on aerosol light scattering: results from different European sites, Atmos. Chem. Phys., 13, 10 609–10 631, https://doi.org/10.5194/acp-13-10609-2013, 2013.
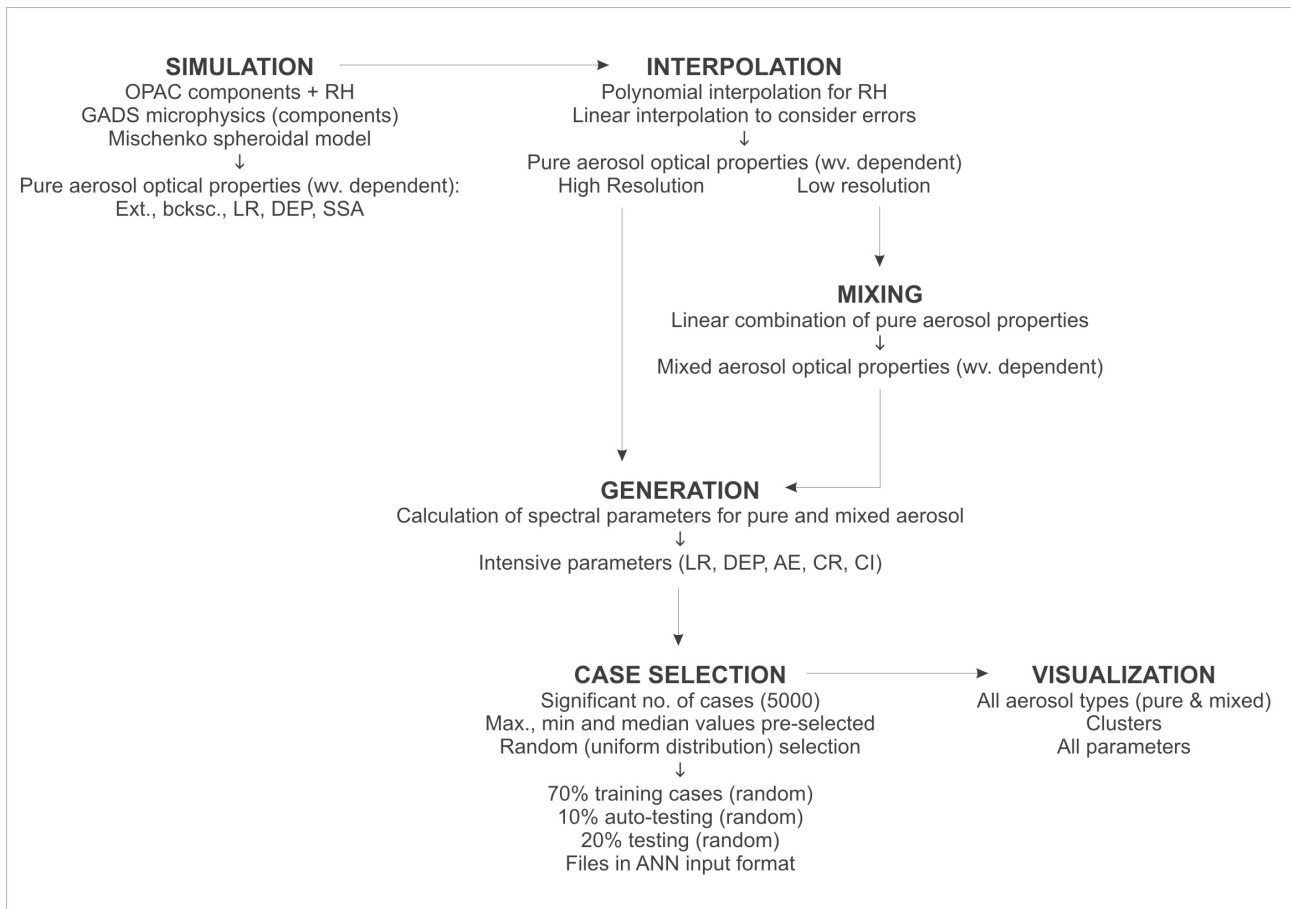
Figure 1. The generation chain of the synthetic data for the NATALI algorithm.
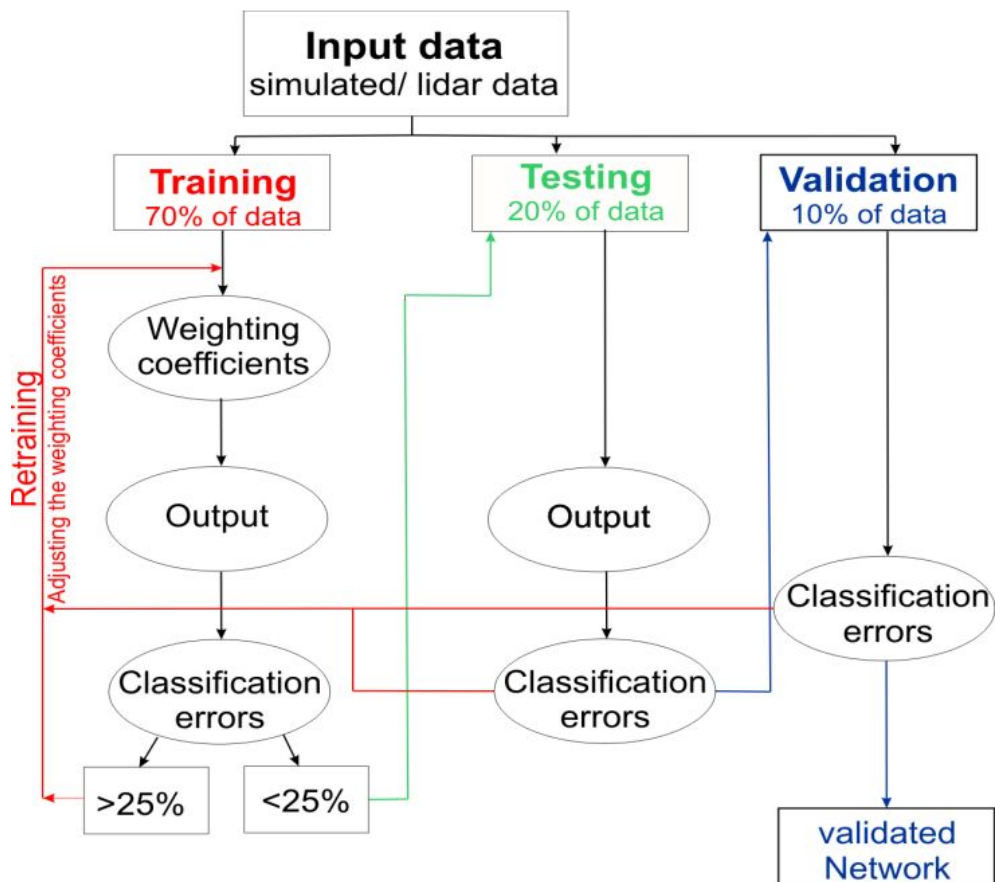
Figure 2. Artificial Neural Network logical scheme for the NATALI algorithm.
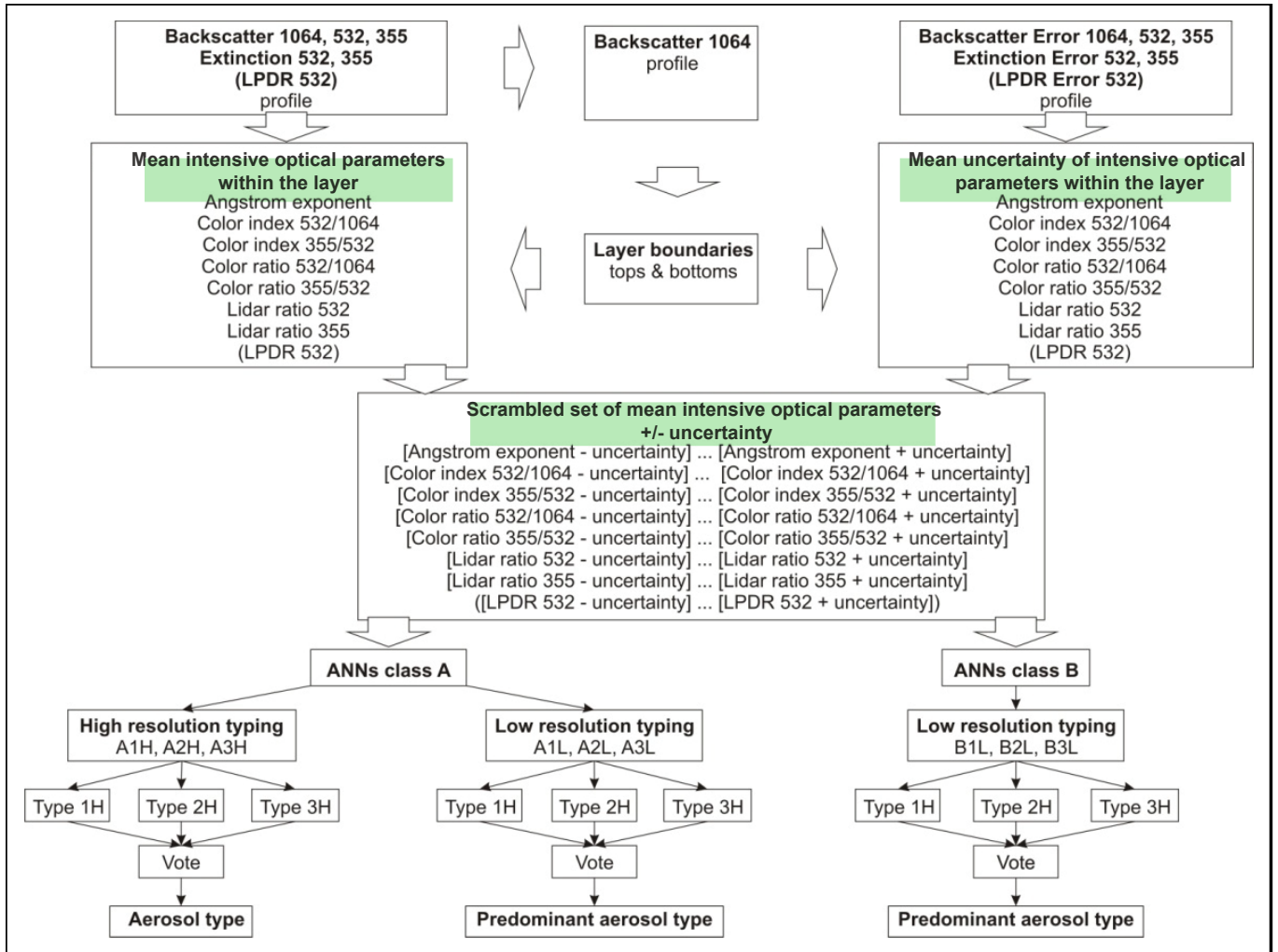
Figure 3. Schematics of the NATALI algorithm for aerosol typing.
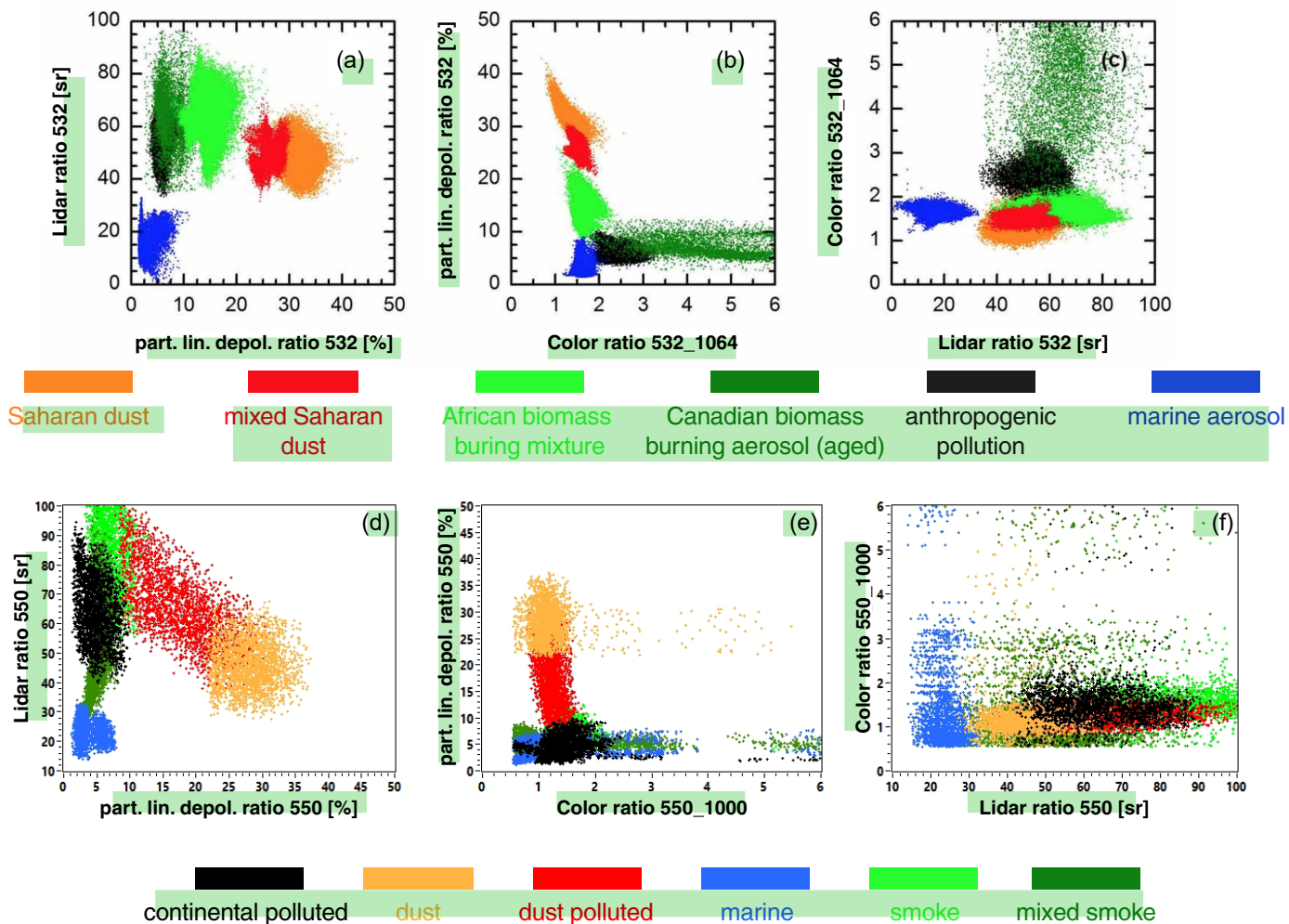
Figure 4. Characteristic quantities of various atmospheric aerosol types form lidar measurements (a–c, adapted from Groß et al. 2013, their Fig. 5) and from synthetic measurements (d–f). (a) and (d) Lidar ratio versus linear particle depolarization. (b) and (e) Linear particle depolarization versus color ratio. (c) and (f) Color ratio versus lidar ratio.
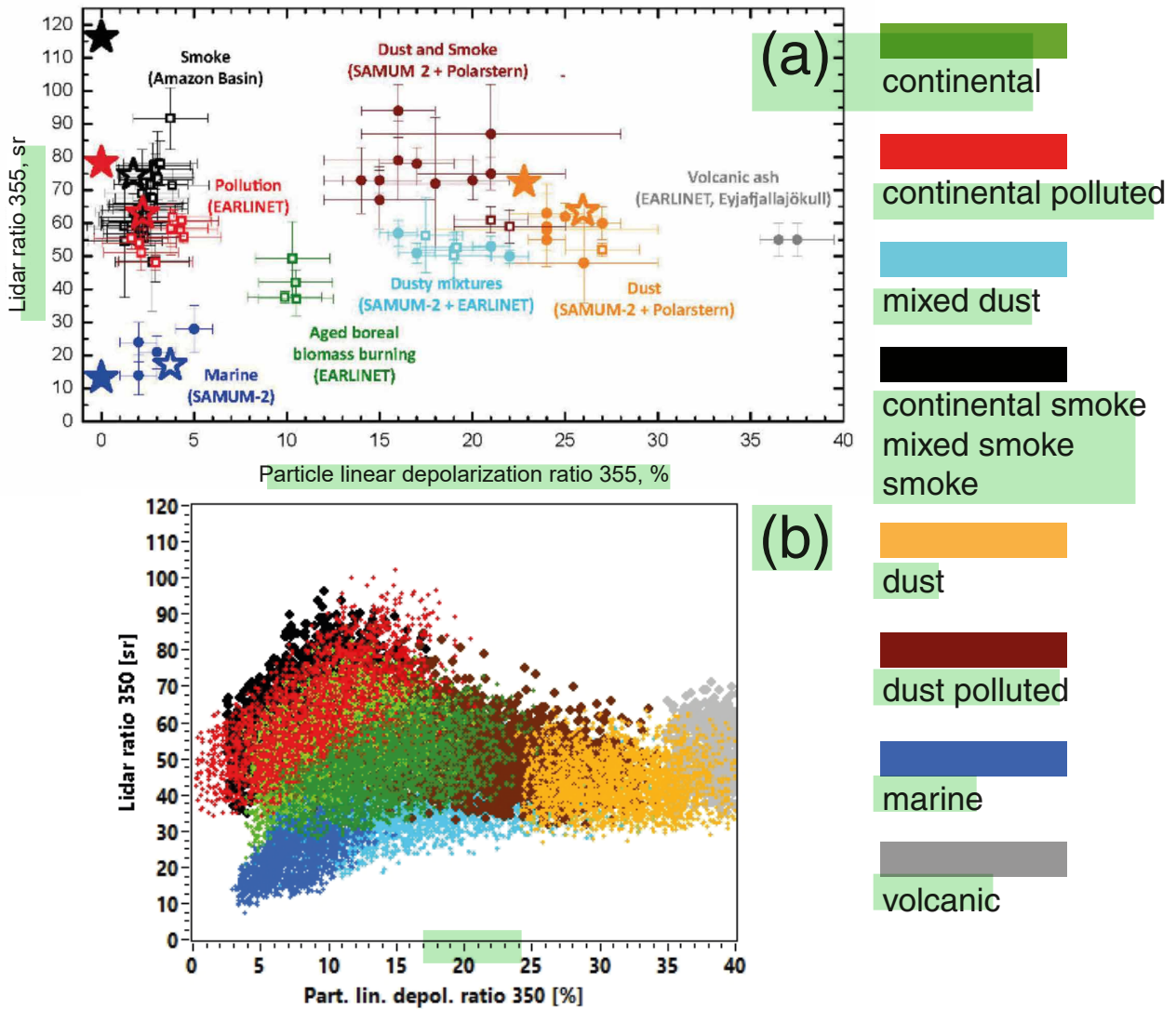
Figure 5. Lidar ratio versus particle linear depolarization ratio. (a) Synthesis of ground-based observations and simulations adapted from Wandinger et al. (2016) (their Fig. 1). Filled stars represent simulations using the components of Aerosol CCI and variations with different refractive index and shape distribution (open stars). (b) Synthetic data from the NATALI aerosol model.
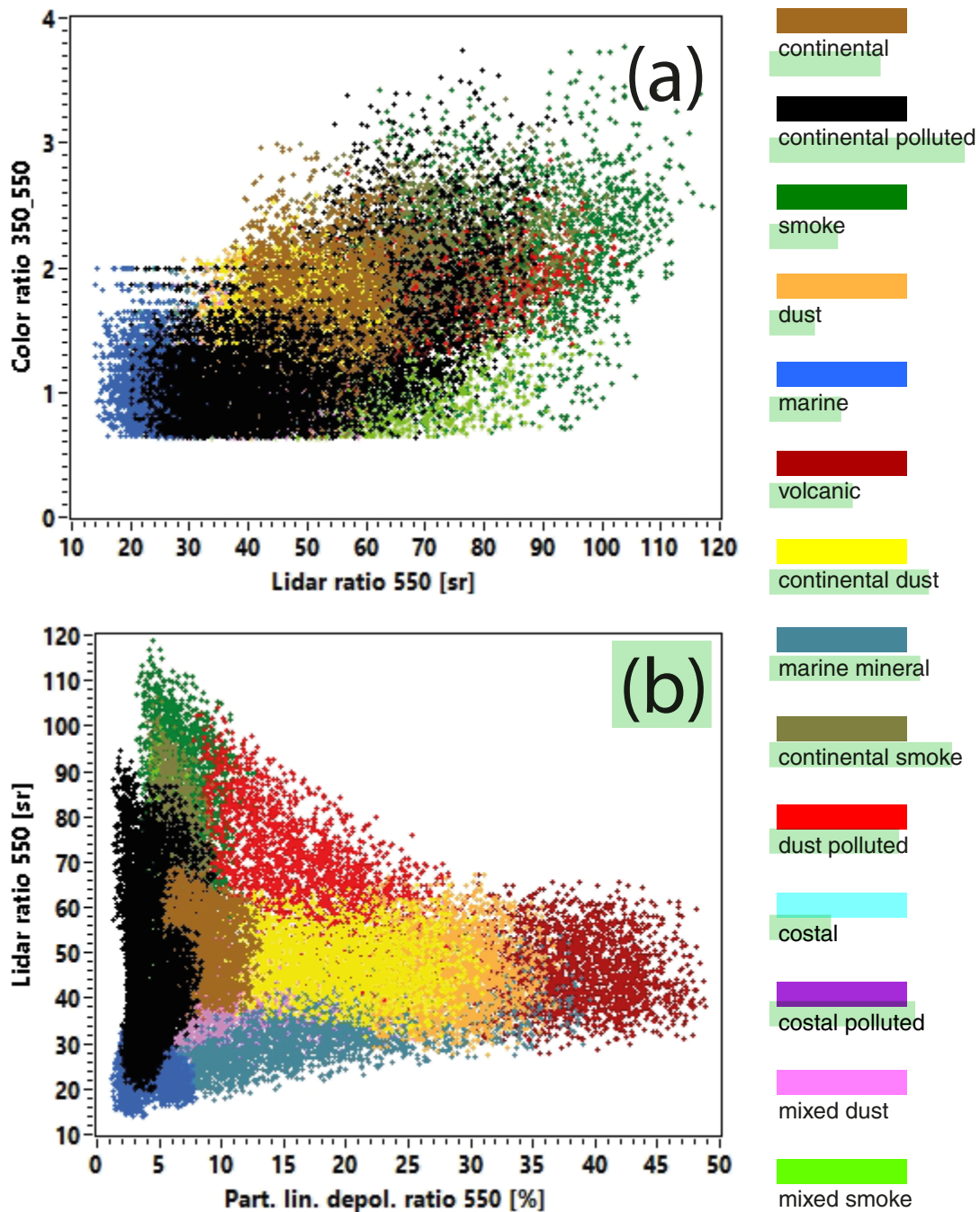
Figure 6. Synthetic dataset (a) color ratio versus lidar ratio, and (b) lidar ratio versus linear particle depolarization ratio using the NATALI classification.
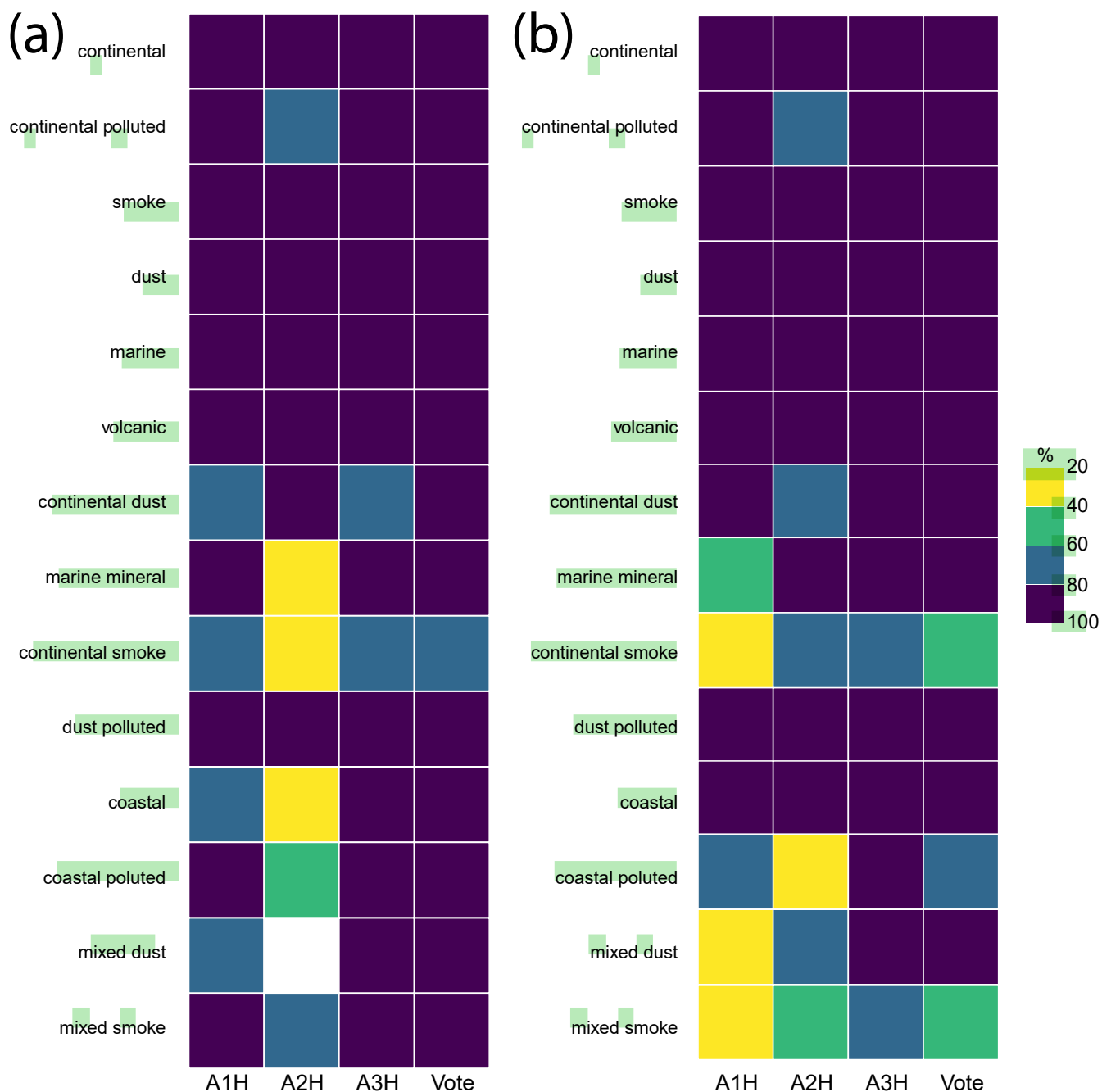
Figure 7. Performances of ANNs for (a) high-resolution typing, and (b) low-resolution typing for each ANN (i.e. A1H, A2H, A3H) and the combine results (Vote) of the three ANNs. The intervals of ANNs confidence levels are shaded according to the scale.
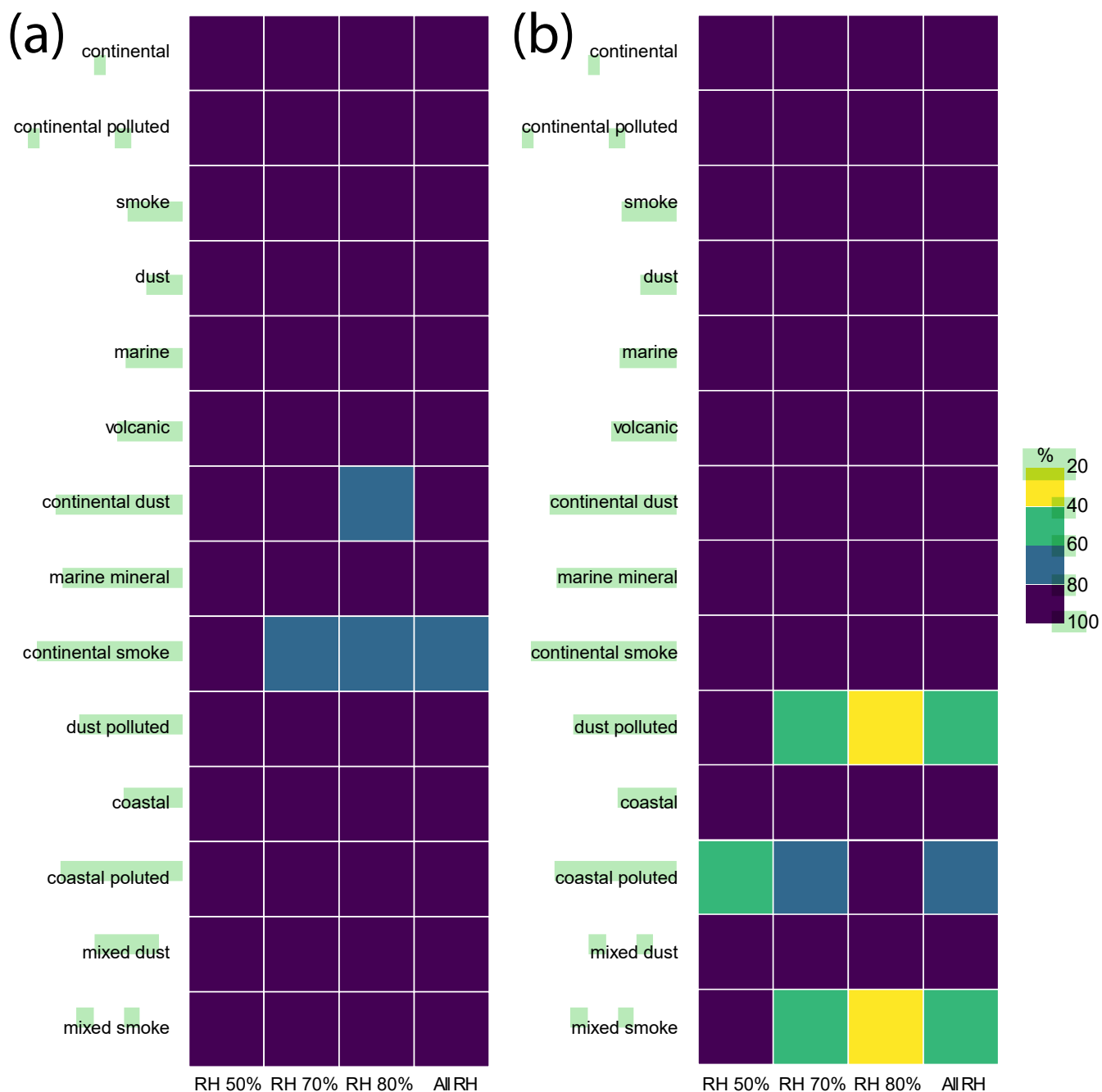
Figure 8. Performances of the ANNs for different relative humidity values (i.e. 50%, 70%, 80%) for (a) high-resolution typing, and (b) low-resolution typing. The intervals of ANNs confidence levels are shaded according to the scale.
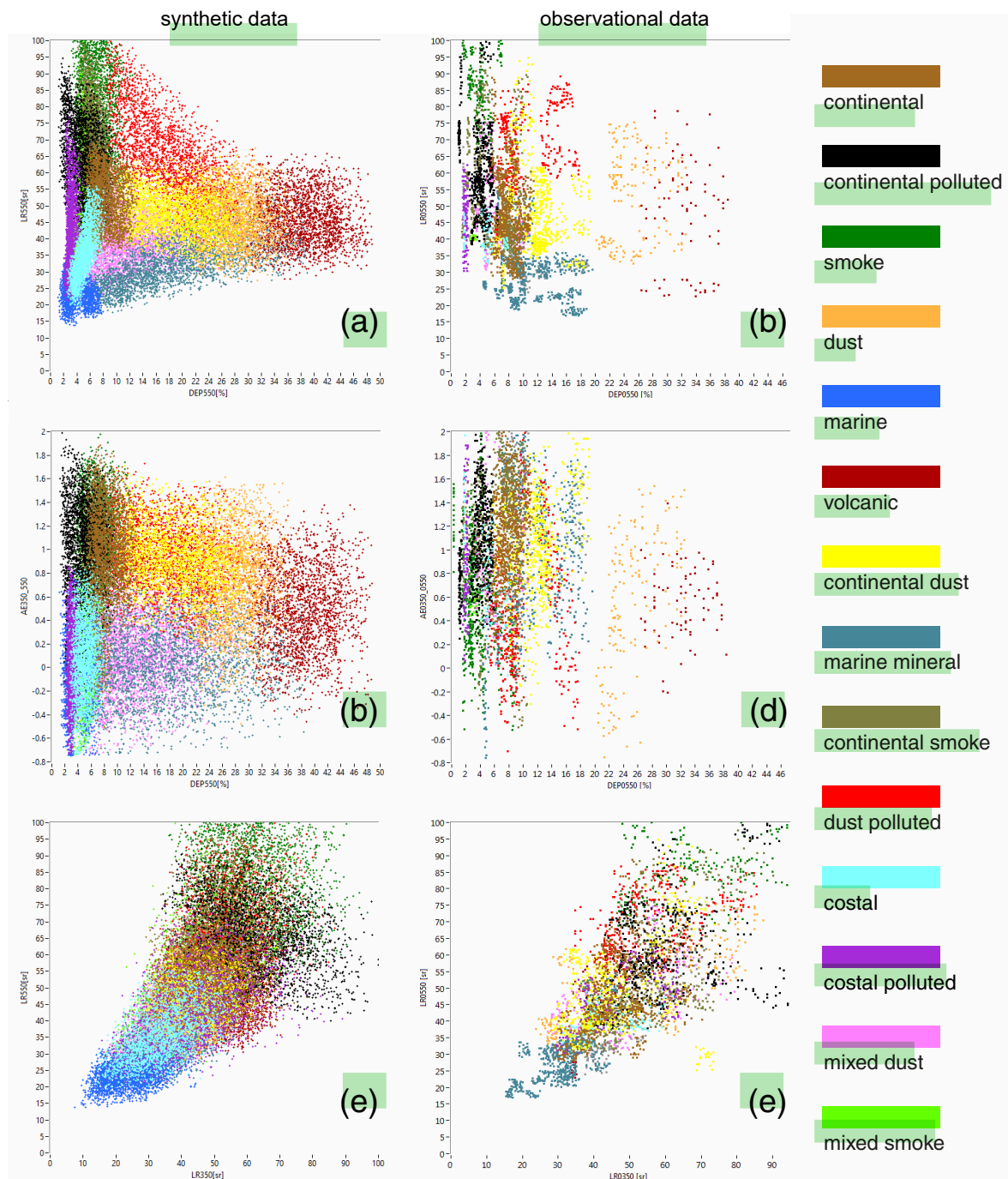
Figure 9. Results of the aerosol typing from NATALI aerosol model (synthetic data) and observations (observational data, EARLINET-CALIPSO database and additional datasets collected at the EARLINET station in Bucharest). (a) and (b) lidar ratio and particle depolarization (VIS), (c) and (d) Ångström exponent and particle depolarization (VIS), and (e) and (f) lidar ratio (VIS) and lidar ratio (UV).

Table 1. Convention names of the aerosols types.

Aerosol type	Source	Particle Characteristics
continental	land surfaces	medium size, medium spherical, medium absorbing
dust	desert surfaces	large, non-spherical, medium absorbing
continental polluted	industrial sites	small, spherical, highly absorbing
marine	sea surface	large, aspherical, non-absorbing
smoke	vegetation fires	small, spherical, highly absorbing
volcanic	volcanoes	large, non-spherical, highly absorbing
mixtures	mixed	combinations of the above

Table 2. Pure aerosol types and components.

Aerosol types	Basic components types	Range variation of the number density mixing ratios for aerosol components (limits are consistent with OPAC and literature)	Aspect ratio
continental	water soluble	0.4914–0.5914	1.100
	insoluble	0.0086–0.0086	
	soot	0.4000–0.5000	
continental polluted	water soluble	0.1998–0.2998	1.040
	insoluble	1.8e-4–1.8e-4	
	soot	0.6000–0.7000	
	sulphate	0.1000–0.1000	
smoke	water soluble	0.3900–0.4900	1.150
	soot	0.5000–0.6000	
	sulphate	0.0100–0.0100	
dust	water soluble	0.1949–0.2949	0.870
	mineral		
	nucleation mode	0.1170–0.1170	
	accumulation mode	0.0880–0.0880	
	coarse mode	0.6e-04–0.6e-04	
	soot	0.5000–0.6000	
marine	water soluble	0.1652–0.1662	1.007
	sea salt		
	accumulation mode	0.8320–0.8320	
	coarse mode	0.0e+00–0.1e-06	
	insoluble	0.5000–0.6000	
volcanic	mineral		0.850
	nucleation mode	0.0915–0.1070	
	accumulation mode	0.1470–0.1719	
	coarse mode	0.4e-04–0.5e-04	
	sulphate	0.0391–0.0457	
	soot	0.6753–0.7224	

Table 3. Selected types of Artificial Neural Networks and their structures.

ANN type	ANN architecture	Advantages	Disadvantages

JE1	6 hidden layers
	Momentum learning rule
	processing elements
	  50 for the first four layers
	  45 for the fifth layer
	  37 for the sixth layer
	tanh axons
	trained for at least 5 cycles of 1000 epochs
		Good percentage of training per aerosol class.
		Stable performances and approximatively constant for all aerosols classes.
			Slow training/time consuming.
			Reach the training limit rapidly.

JE2	8 hidden layers
	Conjugate Gradient learning rule
	processing elements
	  50 for the first four layers
	  45 for the fifth layer
	  37 for the sixth layer
	  32 for the seventh layer
	  28 for the eight layer
	tanh axons
	trained for at least 5 cycles of 1000 epochs
		Good percentage of training per aerosol class.
		Rapid training.
			Only few training cycles can be done.
			Limited performance improvement after training.

GFF	10 hidden layers
	Momentum learning rule
	processing elements
	  50 for the first four layers
	  45 for the fifth layer
	  37 for the sixth layer
	tanh axons
	trained for at least 5 cycles of 1000 epochs
		Low error of training after two training cycles.
		Rapid training.
			It trains efficiently only several cycles a further improvement of weights cannot be considered. Stable active performances per aerosol type overall but lower values for several classes.

Table 4. Correspondence between the aerosol types defined in the algorithm, as they can be retrieved by NATALI in high resolution and low resolution.

Aerosol types	High resolution type (AH)	Low resolution typing with depolarization (AL)	Low resolution typing without depolarization (BL)
continental	continental	continental	continental
continental polluted	continental polluted	continental polluted	continental polluted
smoke	smoke	smoke	smoke
dust	dust	dust	dust
marine	marine	marine	marine
volcanic	volcanic	volcanic	dust or continental
continental + dust	continental dust	continental or dust	continental or dust
dust + marine	marine mineral	dust or marine	dust or marine
volcanic + marine			
continental + smoke	continental smoke	continental polluted or smoke	continental polluted or smoke
dust + smoke	dust polluted	dust or smoke	dust or smoke
continental + marine	coastal	continental or marine	continental or marine
continental polluted + marine	coastal polluted	continental polluted or marine	continental polluted or marine
continental + dust + marine	mixed dust	continental or dust	continental or dust
continental + smoke + marine	mixed smoke	continental polluted or smoke	continental polluted or smoke

Table 5. Acceptable limits for the layer average intensive optical parameters.

Intensive parameter	Minimum acceptable value	Maximum acceptable value
Ångström exponent	-1±1	3±3
color ratio	-1±1	3±3
color index	-1±1	3±3
lidar ratio	2.5±2.5	100±100
linear particle depolarization ratio (%)	0±0	30±25

Table 6. Optical properties of aerosols from the synthetic dataset and from measurements.

Aerosol type	Parameter	Synthetic	Measured	Reference
continental(rural)	AE_{550_350}	1.17–1.29	2.2±0.5	Giannakaki et al. (2010)
	CR_{550_350}	1.56–2.07	–	–
	CR_{1000_550}	1.37–1.85	–	–
	$LR_{350}(sr)$	43–54	29±7	Giannakaki et al. (2010)
	$LR_{550}(sr)$	52–53	–	–
	$DEP_{550}(\%)$	7.23–10.7	–	–
continental polluted/industrial	AE_{550_350}	1.17–1.34	1.4±1.0	Giannakaki et al. (2010)
			≈1.1–1.6	Perone and Bulizzi (2016)
			1.17–1.19	Stachlewska et al. (2017)
	CR_{550_350}	1.34–2.29	0.68–0.85	Vlăduţescu et al. (2007)
	CR_{1000_550}	1.33–1.65	1.7–2.1	Burton et al. (2013)
			2.43±0.27	Groß et al. (2013)
	$LR_{350}(sr)$	55–75	58±12	Müller et al. (2007)
			65–100	Vlăduţescu et al. (2007)
			56±23	Giannakaki et al. (2010)
			56±1	Perone and Bulizzi (2016)
	$LR_{550}(sr)$	62–74	71±10	Cattrall et al. (2005)
			53±11	Müller et al. (2007)
			53–70	Burton et al. (2013)
			56±6	Groß et al. (2013)
			55±1	Perone and Bulizzi (2016)
			57±7	Wang et al. (2016)
	$DEP_{550}(\%)$	2.47–4.97	<5	Müller et al. (2007)
			6±1	Groß et al. (2013)
			3–7	Burton et al. (2013)

smoke	AE$_{550\_350}$	1.15–1.31	1.0±0.5	Müller et al. (2007)
			1.7±0.7	Giannakaki et al. (2010)
			0.9±0.26	Tesche et al. (2011)
			0.3–0.7	Janicka et al. (2017)
			1.0–1.5	Stachlewska et al. (2018)
	CR$_{550\_350}$	1.90–2.59	–	–
	CR$_{1000\_550}$	1.52–1.61	2.1–2.5	Burton et al. (2013)
			1.63±0.13	Groß et al. (2013)
			4.70±0.30	
	LR$_{350}$(sr)	56–72	37.9±13.1	Müller et al. (2005)
			46±13	Müller et al. (2007)
			69±17	Giannakaki et al. (2010)
			87±17	Tesche et al. (2011)
			60±20	Janicka et al. (2017)
			55–70	Stachlewska et al. (2018)
	LR$_{550}$(sr)	81–92	40–80	Wandinger et al. (2002)
			26–87	Müller et al. (2005)
			53±11	Müller et al. (2007)
			63±7	Noh et al. (2007)
			60–65	Alados-Arboledas et al. (2011)
			79±17	Tesche et al. (2011)
			63±7	Groß et al. (2013)
			69±17	
			33–46	Burton et al. (2013)
			100±30	Janicka et al. (2017)
			50–62	Stachlewska et al. (2018)
	DEP$_{550}$(%)	5.04–7.12	5–8	Murayama et al. (2004)
			2–3	Müller et al. (2005)
			<2–5	Burton et al. (2013)
			14±2	Groß et al. (2013)
			7±2	
			3–6	Burton et al. (2013)
			0.93±0.01	Burton et al. (2015)

			0.19±0.20	Tesche et al. (2009b)
			0.62±0.15	Veselovskii et al. (2010)
			1.5±1.0	Giannakaki et al. (2010)
	AE$_{550\_350}$	0.88–0.92	0.06±0.21	Groß et al. (2011)
			0.22±0.27	
			0.9–0.5	Janicka et al. (2017)
			0.0–0.3	
			0.01–0.18	Fernández et al. (2018)
	CR$_{550\_350}$	1.51–1.55	–	–
	CR$_{1000\_550}$	1.1–1.14	1.4–1.6	Burton et al. (2013)
			1.30±0.15	Groß et al. (2013)
dust	LR$_{350}$(sr)	43–46	55±6	Müller et al. (2007)
			30–80	Papayannis et al. (2008)
			53±7	Tesche et al. (2009b)
			52±18	Giannakaki et al. (2010)
			65±10	Veselovskii et al. (2010)
			58±7	Groß et al. (2011)
			53±5	Groß et al. (2016)
			42±10	Janicka et al. (2017)
			40–55	Fernández et al. (2018)
	LR$_{550}$(sr)	44–49	46±5	Sakai et al. (2002)
			42–55	Liu et al. (2002)
			55±5	Müller et al. (2007)
			55±7	Tesche et al. (2009b)
			62±9	Veselovskii et al. (2010)
			62±5	Groß et al. (2011)
			49±9	Burton et al. (2012)
			45–51	Burton et al. (2013)
			48±5	Groß et al. (2013)
			56±7	Groß et al. (2015)
			32±10	Janicka et al. (2017)
			38–61	Fernández et al. (2018)

			10–35	Müller et al. (2007)
dust	DEP$_{550}$(%)	27.22–30.97	10–25	Papayannis et al. (2008)
			32±2	Freudenthaler et al. (2009)
			31–33	Burton et al. (2013)
			24–27	Groß et al. (2011)
			31±25	Groß et al. (2013)
			26±3	Groß et al. (2015)
			32.7±001	Burton et al. (2015)
marine	AE$_{550\_350}$	-0.26–0.21	–	–
	CR$_{550\_350}$	0.77–1.35	–	–
	CR$_{1000\_550}$	0.7–2.91	1.3–1.6	Burton et al. (2013)
			1.64±0.10	Groß et al. (2013)
	LR$_{350}$(sr)	13–32	18±4	Groß et al. (2011)
			20±3	Groß et al. (2016)
	LR$_{550}$(sr)	19–25	28±5	Cattrall et al. (2005)
			23±3	Müller et al. (2007)
			18±2	Groß et al. (2011)
			15–25	Burton et al. (2012)
			17–27	Burton et al. (2013)
			18±5	Groß et al. (2013)
			22±5	Groß et al. (2016)
	DEP$_{550}$(%)	1.9–3.73	2–3	Groß et al. (2011)
			<10	Burton et al. (2012)
			4–9	Burton et al. (2013)
			3±1	Groß et al. (2013)
			2±1	Groß et al. (2016)

volcanic	AE_{550_350}	-0.21–1.07	0.03±0.4	Ansmann et al. (2010)
			-0.11±0.4	
			0.68±0.63	Sicard et al. (2012)
	CR_{550_350}	0.82–1.29	–	–
	CR_{1000_550}	0.74–2.57	–	–
	$LR_{350}(sr)$	50–54	60±5	Ansmann et al. (2010)
			30–60	Mattis et al. (2010)
			39±10	Sicard et al. (2012)
			60±11	Mona et al. (2012)
	$LR_{550}(sr)$	41–49	60±5	Ansmann et al. (2010)
			30–45	Mattis et al. (2010)
			32±4	Sicard et al. (2012)
			78±12	Mona et al. (2012)
	$DEP_{550}(\%)$	37.27–41.8	0.33±0.03	Ansmann et al. (2010)
			16±7	Mona et al. (2012)