

## INDEPENDENT RETRIEVAL OF AEROSOL TYPE FROM LIDAR

Doina Nicolae<sup>1\*</sup>, Jeni Vasilescu<sup>1</sup>, Camelia Talianu<sup>1</sup>, Alexandru Dandocsi<sup>1,2</sup>

<sup>1</sup>National Institute of R&D for Optoelectronics, Romania, \*Email: nnicol@inoe.ro

<sup>2</sup>University Politehnica of Bucharest, Faculty of Applied Sciences, Bucharest, Romania

### ABSTRACT

This paper presents an algorithm for aerosol typing from multiwavelength lidar data, based on Artificial Neural Networks. The aerosol model used to simulate optical properties for the training of the network is described. The algorithm is tested on real observations from ESA-CALIPSO database.

### 1 INTRODUCTION

A major step forward in Earth Observation has been made by the successful implementation of active remote sensing from space (CALIPSO, ADM-AEOLUS, ATLID). Although having multiple advantages (high dynamic range, high spatial and temporal resolution), lidar data suffer by a limited physical content. Both the problems of retrieving optical profiles and microphysical properties from lidar data are ill-posed, needing advanced mathematical algorithms [1]-[3].

Our work refers to the use of Artificial Neural Networks (ANN) combined with an extended learning process in order to retrieve the aerosol class from lidar optical data. The algorithm relies on ANN with tens to hundreds of neurons, which is trained to identify the most probable aerosol type by using aerosol intensive optical parameters as input. The advantage of ANNs compared to other methods is that can work with significant noisy data, as long as a pattern is present.

An ANN represents a mathematical projection of the human neural network. It is based on neurons, axons and synapses, the information being propagated as a neural influx. The analysis process takes part during the propagation of information through synapses, where it is adjusted to properly interpret the given data. The network complexity, architecture and dimensions depend on input data complexity and could contain hundreds of neurons.

The ANN is guided through a learning process (supervised or unsupervised) to recognize the pattern of the data. The main goal of the training

process is to find the optimal value of weights and to have the minimum errors. [4].

### 2 METHODOLOGY

More than 22 multiwavelength Raman lidars are currently operating in Europe, as part of EARLINET (the European Aerosol Research Lidar NETwork). Typical configuration includes 3 elastic backscatter channels (1064, 532 and 355 nm), 2 vibrational Nitrogen Raman channels (607 and 387 nm) and 1 depolarization channel (532 nm). Using the combination of these channels, 8 aerosol intensive parameters can be obtained: 2 lidar ratios (532 and 355 nm), 1 Ångström exponent (355/532 nm), 2 color ratios (355/532 and 532/1064 nm), 2 color indexes (355/532 and 532/1064 nm), and 1 particle depolarization ratio (532 nm). The ultimate goal of our algorithm is to be able to retrieve the aerosol type using these 8 parameters as input, without any complementary information (backtrajectory, column data, source sensitivity).

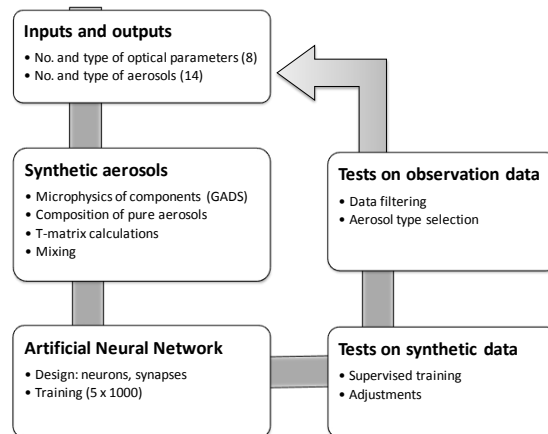


Fig. 1 Schematics of the method

The algorithm relies on a custom-designed ANN with 8 inputs and a number of outputs corresponding to the number of aerosol types to be recognized. The success of the retrieval depends on the information content of the inputs, as well as on the extent of the training program. A high number of good training cases is required to

obtain a reasonable level of confidence. Such validated database does not exist at this moment, therefore we developed an aerosol model to provide the necessary inputs for the ANN. Observations were used to adjust the aerosol model and to test the performance of the ANN.

### **2.1 Synthetic data**

The synthetic database is generated by simulating the optical properties of various aerosol types based on available information on the microphysics of basic components. The algorithm combines the GADS database (Global Aerosol Data Set) with OPAC model (Optical Properties of Aerosol and Clouds) and T-Matrix code in order to compute, in an iterative way, the intensive optical properties of each aerosol type. Four classes of humidity (50%, 70%, 80%, 90%) are considered, out of eight classes in OPAC, high humidity values being excluded in order to avoid ambiguous results related to activation of the hygroscopic particles. Each aerosol type (pure) was built as an internal mixture of typical components which do not interact physically or chemically, having different mixing ratios. Basic components were picked up from OPAC [5], and GADS database is used for the microphysical properties of each component.

The database is built in two stages: a) microphysical and optical properties of pure aerosols, as described before; b) optical properties of mixed aerosols, using the properties obtained in stage a). Six classes of pure aerosol are defined: Continental, Continental Polluted, Dust, Marine, Smoke, and Volcanic.

Particles were considered spheroids with different axis ratios, and the T-matrix code [6] was used to generate the optical properties. Axis ratios were picked up from the literature [7]. Extensive (backscatter and extinction coefficients) and intensive optical parameters (Ångström exponents, lidar ratios, color ratios, particle depolarization ratios) were computed.

Measurements have associated uncertainties (systematic and statistical errors), either due to the instrument - biases, linearity, calibration), or due to the data treatment - algorithms applied to correct / average raw signals and to calculate data products. In order to simulate values as close as possible to the ones obtained from lidar

measurements, we considered for each parameter an associated relative error of maximum 20%, and calculated the possible range based on the initially simulated value. Basically, within this range any value of the parameter is accepted.

Eight classes of mixed aerosols were considered, corresponding to the most frequent aerosol types observed by EARLINET (the European Aerosol Research Lidar Network): Continental Dust, Marine Mineral, Continental Smoke, Dust Polluted, Coastal, Coastal Polluted, Mixed Dust, and Mixed Smoke. Optical properties of mixed aerosols were computed as linear combination of corresponding optical properties of pure types.

### **2.2 ESA-CALIPSO database**

The ESA-CALIPSO database contains EARLINET ground-based retrievals which are used for the derivation of conversion factors of pure aerosol types and typical mixtures of aerosols [8]. The data are grouped in aerosol categories, more or less related to CALIPSO aerosol types. Measurements were used to identify aerosol layers and calculate extensive and intensive optical parameters. For each layer, mean lidar ratios and Ångström exponents were calculated from the backscatter and extinction profiles provided by the stations. The aerosol type and source region for each aerosol layer was identified through a number of auxiliary data, e.g. FLEXPART [9].

ESA-CALIPSO database includes a total number of 718 cases and 21 aerosol and cloud types, out of which only 13 contains all the necessary parameters (3 backscatters, 2 extinctions and 1 depolarization). To increase the number of cases to test the ANN, we added the particle depolarization, assuming typical values (from the literature) for the corresponding aerosol type. We obtained 105 cases, out of which a large number represents mixture of 3 or more pure types. In order to compromise between the performance of the algorithm and the number of output aerosol types, we made a selection of 14 representative types, including mixtures of 3.

### **2.3 Artificial Neural Networks**

A complex Jordan /Elman ANN with 6 hidden layers and 50 processing elements per layer was built to check the capability of the algorithm to

identify the aerosol type. Eight inputs were fed into the algorithm, corresponding to the aerosol intensive optical parameters which can be calculated from a  $3\beta+2\alpha+1\delta$  lidar. Fourteen outputs were considered, corresponding to the 14 aerosol types (6 pure and 8 mixed).

About 70% of the synthetic data described above were used to train the network in 5 steps, with 1000 iterations per training set. Measured data were not used to train the ANN because even a small error in the initial classification can lead to improper operation of the algorithm. Testing was performed on both synthetic and measured data. The relative error of each parameter was used to generate 19 additional values, and combination of all obtained parameters was considered for the synthetic and the observational database.

### 3 RESULTS

Comparative results of the optical parameters from synthetic data and ESA-CALIPSO are shown below.

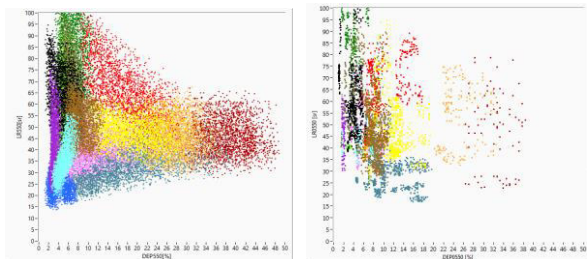


Fig. 2 Lidar ratio versus particle depolarization at 532nm for Continental (brown), ContinentalPolluted (black), Marine (blue), Dust (orange), Smoke (green), Volcanic (dark red), DustPolluted (Red), ContinentalDust (yellow), Coastal (cyan), CoastalPolluted (violet), ContinentalSmoke (olive), MixedDust (magenta), MixedSmoke (light green) and MarineMineral (steel) aerosols: left - synthetic aerosols; right – ESA-CALIPSO database [8]

As shown in Fig. 2, synthetic data agree quite well with the ESA-CALIPSO, for each of the aerosol type. With the considered accuracy and definition of types, clusters are not completely distinct in the synthetic data (left). This is also true for the observational data (right). Additionally, wide spreading of the measured parameters (right panel) is due to the presence of more complex mixtures (not considered in our simulations), incorrect calibration or underestimation of the measurement uncertainty. Although several

aerosol types are frequently observed in Europe, the observational cases available are not sufficient to describe all possible combinations of optical parameters, leading to apparently incomplete clusters, which must not be seen as a failure of the aerosol model to simulate real measurements. For example, marine aerosols are rarely present in the ESA-CALIPSO database, and the uncertainty is higher than 20% therefore are not represented in the right panel. Coastal aerosols are also poorly represented. For such types, no conclusion can be drawn.

The ANN was tested after five training cycles with synthetic data. Tests were performed considering three degrees of discrimination: 55%, 65% and 75%.

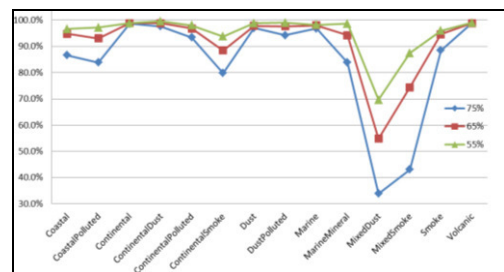


Fig. 3 Performance of the ANN after training on synthetic data; different degrees of discrimination

Twelve aerosols types are identified in more than 80% of the cases, Mixed Dust and Mixed Smoke being recognized in only 30% cases when the highest discrimination level is applied. All aerosols types are well classified in more than 70% cases when 55% discrimination level is applied (Fig. 3). Blind test was applied on ESA-CALIPSO database, with a discrimination level of 75%, and results are shown in Fig. 4.

Type ANN	Coastal Polluted	Continental	Continental Polluted	Continental Smoke	Dust	Mixed Dust	Smoke	Volcanic	Unknown
clean continent. + dust + marine aerosol						1			
clean continental		2							
clean continental + dust aerosol		1							
clean continental + marine aerosol		2							1
dust					1				3
dust + marine aerosol (= mixed dust)		1							4
poll. continent. + dust + marine aerosol		1							2
polluted contin. + dust + marine + smoke				1					
polluted continental		1	6	2					3
polluted continental + marine + smoke	1								
polluted continental + marine aerosol	5		1						
polluted continental + smoke aerosol			2						
polluted continental aerosol + dust		1							2
polluted dust (dust + smoke)				1					1
smoke					1				
volcanic aerosol						2			1

Fig. 4 Comparative results of aerosol types from ESA-CALIPSO and the ANN

Note that, although our algorithm is not always in perfect agreement with the classification in ESA-CALIPSO (marked in green), there are only small differences which are due to a different definition of mixtures (marked in blue). For example, 1 case of polluted continental is classified by the ANN as Continental, 2 are classified as Continental Smoke, and 1 as Smoke. Similarly, 2 volcanic cases are classified by the ANN as Dust, and 1 as Smoke, only 1 being classified as Volcanic. However, the type could not be retrieved for a significant number of cases (16) when applying a discrimination level of 75%. A possible cause is the underestimation of the measurement uncertainties, and/or too demanding constraints for the level of confidence of the ANN.

#### 4 CONCLUSIONS

This paper presents an algorithm for aerosol typing based on Artificial Neural Networks and lidar-retrieved aerosol intensive optical properties. Synthetic data used to train the network was generated by combining OPAC and T-matrix calculations and considering both pure and mixed aerosols. Values and ranges are within the ones published in the literature and generally agree with ESA-CALIPSO. Observations were used to adjust the aerosol model by direct comparison, and to test the performance of the algorithm.

We applied the algorithm on ESA-CALIPSO data with accuracy better than 20%. A large number of cases were correctly classified by the algorithm, although the distinction between Dust and Volcanic is still to be improved. In this case, the retrieval depends strongly on the quality of the spectral parameters (particle depolarization is similar, as well as lidar ratios), which is limited by the signal-to-noise ratio and calibration of the lidar. In case of observation data with accuracy lower than 20%, complex mixtures of 3 aerosol types cannot be retrieved without accepting a lower confidence level for the ANN.

#### ACKNOWLEDGEMENT

This work has been supported by the ESA contract no. 4000110671/14/I-LG, NATALI and by the STAR-ESA Programme Project No.55/2013- CARESSSE.

#### References

- [1] Müller, D., Wandinger, U., Ansmann, A., 1999: Microphysical particle parameters from extinction and backscatter lidar data by inversion with regularization: theory. *Appl. Opt.* **38** (12), 2346-2357.
- [2] Osterloh, L., Böckmann, C., Mamouri, R. E., Papayannis, A., 2011: An adaptive base point algorithm for the retrieval of aerosol microphysical properties. *Open Atmospheric Science Journal* **5**, 61-73.
- [3] Veselovskii, I., Kolgotin, A., Griaznov, V., Müller, D., Franke, K., Whiteman, D. N., 2004: Inversion of multiwavelength raman lidar data for retrieval of bimodal aerosol size distribution, *Appl. Opt.* **43** (5), 1180-1195.
- [4] Gupta P. and Christopher S. A., 2009: Particulate matter air quality assessment using integrated surface, satellite, and meteorological products: 2. A neural network approach, *J. Geophys. Res.*, **114**, D20205
- [5] Hess, M., P. Koepke, and I. Schult, 1998: Optical properties of aerosols and clouds: The software package OPAC, *Bull. Am. Meteorol. Soc.*, **79**, 831 – 844
- [6] Mishchenko, M. I., and L. D. Travis, 2000: Light Scattering by Nonspherical Particles, *Academic Press*, San-Diego, 690pp.
- [7] Dubovik, O., Sinyak, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., Eck, T. F., Volten, H., Munoz, O., Veihelmann, B., van der Zande, W. J., Leon, J.-F., Sorokin, M., and Slutsker, I., 2006: Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust, *J. Geophys. Res.*, **111**, D11208, doi:10.1029/2005JD006619.
- [8] Wandinger U., A. Hiebsch, I. Mattis, G. Pappalardo, L. Mona, and F. Madonna, 2011: Aerosols and Clouds, Long-term Database from Spaceborne Lidar Measurements, *Final report*, ESTEC Contract 21487/08/NL/HE.
- [9] Stohl, A., M. Hittenberger, and G. Wotawa, 1998: Validation of the Lagrangian particle dispersion model FLEXPART against large scale tracer experiment data, *Atmos. Env.*, **32**, 4245-4264.