

Aerosol climatology from lidar observations using deep learning

Camelia Talianu^{(a),(b)}, Doina Nicolae^(a), Victor Nicolae^(a)

^(a) Remote Sensing Department, National Institute of Research and Development for Optoelectronics, 409 Atomistilor Street, Magurele, RO77125, Romania

^(b) Institute of Meteorology and Climatology, University of Natural Resources and Life Sciences, 33 Gregor-Mendel Street, 1180 Vienna, Austria
 Lead Author e-mail address: camelia@inoe.ro

Abstract: This paper presents a methodology for deriving aerosol climatology with artificial intelligence using lidar observations combined with atmospheric modeling and unsupervised deep learning. The method can be applied to lidar measurements and modeled data, or their combination, if observed and modeled data are consistent. The advantages of the method are: ability to automatically retrieve the optical properties of aerosols, repeatability and scalability.

1 Introduction

One of the biggest strengths of lidar systems is the ability to provide real-time information on the height and the structure of cloud and aerosol layers. By applying inversion algorithms [1–4], the vertical distributions of the optical properties of the aerosol can be obtained. Based on an “expert analysis”, the aerosol optical properties are further used for the characterization and classification of aerosols present in the atmosphere. In general, expert analyzes are dependent on the analyzer choices and experience (thresholds, correction factors, etc) , therefore they are subjective and slow. An analysis based on deep learning could be faster than expert analysis. Machine learning techniques have emerged as a very promising approach to extract and process more information from synergistic measurements in order to estimate the aerosol vertical distributions [5,6] or to classify the types of aerosols [7,8].

In this paper, a novel synergistic methodology for aerosol characterization and classification using aerosol remote sensing observations combined with atmospheric modeling and unsupervised deep learning (DL) is presented.

2 Methodology

For this study, data collected from lidar, ceilometer and photometer measurements, performed at the Magurele Centre for Atmosphere and Radiation Studies (MARS), combined with Copernicus Atmosphere Monitoring Services (CAMS) products (<https://atmosphere.copernicus.eu/>), NATALI

aerosol-typing model [7] and FLEXPART (FLEXible PARTicle dispersion model) atmospheric transport model [9] together with TRACE “Automated time–height-resolved air mass source attribution tool” [10] are used. Data sets including simultaneous lidar, ceilometer and photometer measurements from 2018–2023, selecting transport events of different aerosol types, were considered for testing the DL analysis method.

Figure 1 shown the block diagram for the DL-based methodology used in atmospheric aerosol climatology. The main aerosol products used in this climatology are: backscatter and extinction coefficients, aerosol optical depth (AOD), lidar ratios (LR), Angstrom exponent (AE), and color ratio (CR).

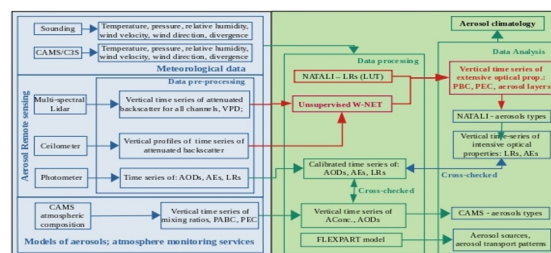


Figure 1. DL-based methodology block diagram

Meteorological data and atmospheric composition products were selected from CAMS for pressure levels corresponding to altitudes between 0.2–10 km with a temporal resolution of one hour. They were synchronized in time with the lidar and ceilometer measurements and interpolated for the MARS location.

The vertical profiles of attenuated backscatter retrieved from lidar and ceilometer are averaged for a temporal resolution of five minutes and interpolated by altitude for a spatial resolution of 15 m.

The aerosol layers heights are obtained from the lidar and ceilometer data using unsupervised DL W-Net models [5]. For each identified segment, the TRACE with FLEXPART model is run to estimate source-receiver sensitivity as well as aerosol type.

For each identified aerosol type, the corresponding lidar ratio is extracting from the optical properties of the NATALI aerosol model. The lidar ratios are used further to compute the extinction profiles.

3 W-Net model

The W-Net architecture has been adapted for training on 1D signals (lidar and ceilometer signals) and not on 2D images. The W-Net network was built with 14 modules to identify all types of aerosols defined in NATALI. For each module, the network consists of two U-type architectures, each containing an encoder block and a decoder block. On the first U-Net, segmentation of the signals is obtained. For each profile of lidar and ceilometer measurements, the segmentation task involves indicating areas with similar characteristics where the main type of aerosols is the same. The first U-Net outputs are further used as input to the second U-Net to construct the backscatter profiles.

Figure 2 illustrates the W-Net flow diagram.

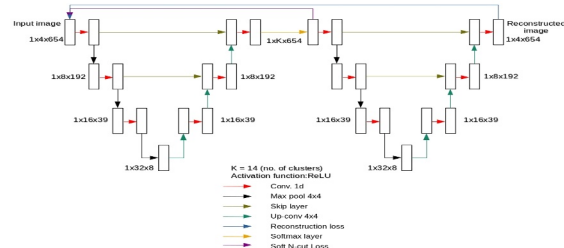


Figure 2. W-Net architecture adapted for segmentation aerosol layers

Each U-Net has 3 convolution blocks for encoding and 3 convolution blocks for decoding. Each such block in turn has a 1D convolution block defined by the number of input channels, a number of output channels, the size of the convolution kernel, a 1D Max

Pooling block of size 4, and the ReLu activation function.

The number of input channels in the coding/decoding blocks is doubled at the output.

To define the segments, the U-Net uses the latent information from the hidden layer in the middle of the W-Net model (i.e. the layer with size $1 \times K \times 654$; where K represents the number of clusters, see Figure 2). Each of the K vertical vectors (of size 654) from the output of this layer include the probabilities that one of the K types of aerosols is found at the altitude indicated by the vertical index (from 0 to 653). In our case $K = 14$.

Three attenuated backscatter profiles retrieved from lidar (1064 nm, 532 nm, 355 nm) and one from ceilometer measurements were used as input data in W-Net model.

4 Case study: 26 July 2021

Below the results of an DL-analysis for the July 26, 2021 data sets are presented.

For the aerosol analysis, we consider the combination of lidar and CAMS aerosol products correlated with photometer and ceilometer measurements. The optical properties retrieved from the photometer are re-scaled to the lidar wavelengths using an Angstrom exponent equal to 1 for the comparison with optical properties obtained from lidar measurements. The values of the attenuated backscatter profiles retrieved from CAMS are at the same wavelengths as the lidar system, no re-scaling is needed.

The attenuated backscatter profiles obtained from lidar at 1064 nm for two time periods: (a) in the time interval 8:30–11:30 UTC (morning) and (b) in the time interval 17:45–21:15 UTC (evening) are shown in Figure 3.

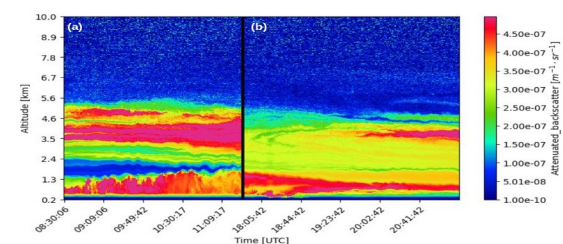


Figure 3. Lidar attenuated backscatter profiles The attenuated backscatter profiles from ceilometer are shown in Figure 4 and the time

intervals common with lidar measurements are marked with white boxes.

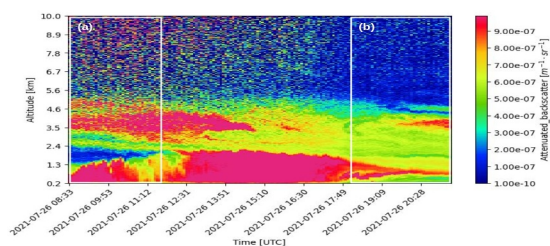


Figure 4. Ceilometer attenuated backscatter profiles

The layers obtained from the W-Net model are shown in Figure 5.

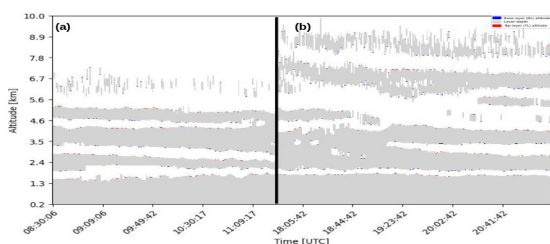


Figure 5. W-Net layers

The sources of aerosol particles arriving at various altitudes and times over MARS on 26 July 2021 obtained from TRACE are shown in Figure 6.

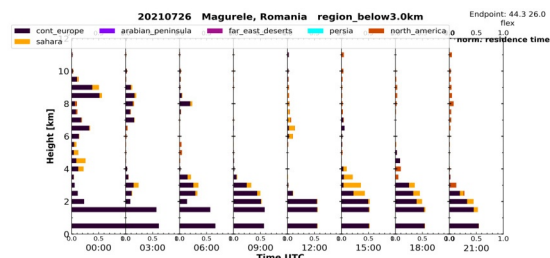


Figure 6. Sources of aerosol particles

5 Results and discussions

The lidar and CAMS extinction coefficient profiles, integrated over the air column, provide the AOD values, which are further cross-checked with the AOD values obtained from photometer. If the results are consistent (relative differences smaller than 0.5), the optical properties retrieved from lidar measurements are taken as reference and further used for aerosol climatology.

The backscatter and extinction profiles are used as input data in NATALI to refine the determination of the aerosol type.

The column AOD values obtained for the selected case study are shown in Table 1.

Table 1. Averaged AOD values

	355 nm	532 nm	1064 nm
Lidar	0.48 ± 0.02	0.33 ± 0.17	0.15 ± 0.01
CAMS	0.50 ± 0.11	0.34 ± 0.08	0.18 ± 0.06
Photometer	0.50 ± 0.12	0.31 ± 0.08	0.15 ± 0.04

The aerosol mixing ratios from CAMS for the the aerosol layers observed at 09:00 UTC are shown in Table 2 and the optical properties of aerosol layers are shown in Table 3.

Table 2. Fraction of aerosol compounds computed from CAMS on 26 July 2021, 09:00 UTC

Alt. [km]	Organic matter	Black carbon	Dust	Sulfate
< 1.75	0.56 ± 0.04	0.03 ± 0.02	0.02 ± 0.02	0.39 ± 0.03
2.0-2.8	0.40 ± 0.08	0.03 ± 0.01	0.40 ± 0.02	0.16 ± 0.04
3.0-4.2	0.26 ± 0.02	0.02 ± 0.02	0.56 ± 0.02	0.15 ± 0.04
4.6-5.4	0.24 ± 0.02	0.02 ± 0.02	0.57 ± 0.01	0.16 ± 0.05

Table 3. Aerosol optical properties computed from NATALI on 26 July 2021, 09:00 UTC

Alt. [km]	LR_532 [sr]	LR_355 [sr]	AE	CR _{1064_532}
< 1.75	49 ± 2	47 ± 3	2.8 ± 0.1	0.5 ± 0.5
2.0-2.8	53 ± 1	50 ± 1	2.0 ± 0.3	0.5 ± 0.2
3.0-4.2	56 ± 1	50 ± 2	0.6 ± 0.4	0.9 ± 0.1
4.6-5.4	55 ± 2	48 ± 2	0.8 ± 0.1	0.8 ± 0.1

Figures 3, 4, 5 and 6 show a complex pattern of contributions to aerosols observed over MARS site on 26 July 2021. The figures 3, 4 and 5 shows consistent aerosol layers above planetary boundary layer (PBL), which reach around 1.5 km. As can see in Figure 4, the layer between 2.0–2.8 km, observed in the first

part of the day from lidar measurements (marked with (a) in Figures 3 and 5), enters the PBL around noon. After noon (no lidar measurements), the upper layers between 2.0–6.0 km disperse and part of them mixes with the PBL. After 18:00 UTC, other layers of aerosols distributed between 2 and 10 km are observed. Also, in Figures 3 and 4, an intrusion of clean air is observed during the evening where the atmosphere is clear up to 1.0 km altitude. The aerosol is pushed up and concentrated above the residual layer.

As can see in Figure 6, for the lower layers (below 1.75 km), Europe is the main source of aerosols. For the mid-altitude layers (between 2.0 and 5.0 km), sources in Europe and Sahara are the main sources of aerosols in the first part of the day (before 15:00 UTC), while the North America makes important contributions, especially after 15:00 UTC. For the high-altitude layers (over 5.0 km), the main contributions come from Sahara in the first part of the day, while the sources from North America have a significant contribution in the second part of the day.

The aerosol type given by NATALI for the PBL is continental-polluted, with a high content of organic matter and poor content of the dust particles. The layers above 2 km are polluted dust, with different dust concentration in the first part of the day and smoke for the second part of the day with different black carbon concentration.

6 Conclusions

For this case, a very good correlation between column AODs measured by lidar and photometer, as well as modeled by CAMS was obtained. As such, lidar measurements can be used to calculate the optical properties of aerosols within each layer and CAMS can be further used to estimate the aerosol mixing ratios specific for each layer.

We have demonstrated that a synergistic methodology for aerosol characterization and classification using aerosol remote sensing observations combined with atmospheric modeling and unsupervised “deep learning”, is able to identify and to classify aerosol layers automatically with a high degree of repeatability and scalability. The method can be applied to observational and modeled data, or their combination, if data are consistent.

7 Acknowledgments

This work is supported by the RI-URBANS project (Research Infrastructures Services Reinforcing Air Quality Monitoring Capacities in European Urban & Industrial Areas, European Union’s Horizon 2020 research and innovation program, Green Deal, European Commission, contract 101036245), by the Core Program within the Romanian National Research Development and Innovation Plan 2022-2027, carried out with the support of MCID, project no. PN 23 05 and through Program 1- Development of the national research-development system, Subprogram 1.2 - Institutional performance - Projects to finance the excellent RDI, Contract no. 18PFE/30.12.2021.

8 References

- [1] C. Talianu, “New algorithm for the retrieval of aerosol’s optical parameters by LIDAR data inversion,” *SCEE*, **11**, 55–61 (2007).
- [2] A. Ansmann, “Combined raman elastic-backscatter LIDAR for vertical profiling of moisture, aerosol extinction, backscatter, and LIDAR ratio,” *Appl. Phys. B*, **55**, 18–28 (1992).
- [3] J.D. Klett, “Stable analytical inversion solution for processing lidar returns,” *Appl. Optics*, **20**, 211–20, (1981)
- [4] F.G. Fernald, “Analysis of atmospheric lidar observations: some comments”, *Appl. Optics*, **23**, 652, (1984)
- [5] C. Manolache, “Unsupervised deep learning models for aerosol layers segmentation”. In *Proceedings of the 2022 14th International Conference on Communications (Institute of Electrical and Electronics Engineers, Bucharest, 2022)*, 1–6
- [6] M.J. McGill, “Machine learning-enabled real-time detection of cloud and aerosol layers using airborne lidar”. *Front. Remote Sens.* **4**, 1116817, (2023)
- [7] D. Nicolae, “A neural network aerosol-typing algorithm based on lidar data”, *ACP*, **18**, 14511–14537, (2018)
- [8] J.E. Yorks, “Aerosol and Cloud Detection Using Machine Learning Algorithms and Space-Based Lidar Data”. *Atmosphere* **12**, 606, (2021)
- [9] I. Pisso, “The Lagrangian particle dispersion model FLEXPART version 10.4”, *GMD*, **12**, 4955–4997, (2019)
- [10] M. Radenz, “Automated time–height-resolved air mass source attribution for profiling remote sensing applications”, *ACP*, **21**, 3015–3033, (2021)