

STRENGTHS AND LIMITATIONS OF THE NATALI CODE FOR AEROSOL TYPING FROM MULTIWAVELENGTH RAMAN LIDAR OBSERVATIONS

Doina Nicolae^{1*}, Camelia Talianu^{1,2}, Jeni Vasilescu¹, Victor Nicolae¹, Iwona S. Stachlewska³

¹National Institute of R&D for Optoelectronics, Romania, *nnicol@inoe.ro

²University of Natural Resources and Life Science, Institute of Meteorology, Austria

³University of Warsaw, Faculty of Physics, Institute of Geophysics, Poland

ABSTRACT

A Python code was developed to automatically retrieve the aerosol type (and its predominant component in the mixture) from EARLINET's 3 backscatter and 2 extinction data. The typing relies on Artificial Neural Networks which are trained to identify the most probable aerosol type from a set of mean-layer intensive optical parameters. This paper presents the use and limitations of the code with respect to the quality of the inputted lidar profiles, as well as with the assumptions made in the aerosol model.

1 INTRODUCTION

Global and local properties of atmospheric aerosols have been extensively observed using both space-borne and ground-based instruments, especially during the last decade. Researchers are exploiting satellite remote sensing observations to characterize aerosol layers and to both, assess and propose parameterizations in regional and global models. Although detailed studies of aerosol optical properties have been made, mixed aerosols are still difficult to characterize.

An important advance in the remote sensing of aerosols is the development of continental-scale ground based lidar networks, which provide quality assured optical profiles on a large temporal and spatial scale. EARLINET, the European Aerosol Research Lidar Network data (<http://ww.earlinet.org>) is relevant for climatology, but also for special events, with strong aerosol influence: Saharan dust outbreaks, forest-fire smoke plumes, photochemical smog or volcano eruptions [1].

In the frame of the ESA funded NATALI project, we developed an algorithm based on a combination of Artificial Neural Networks, to estimate the most probable aerosol type from a set of multispectral data. The algorithm has been implemented in a Python code and adjusted for

running on the EARLINET profiles (3 backscatters, 2 extinctions and one linear particle depolarization ratio).

The algorithm relies on the ability of specialized Artificial Neural Networks (ANN) to resolve the overlapping values of the intensive optical parameters calculated for each identified layer in the multiwavelength Raman lidar profiles [2]. 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 picked up 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 values of relative humidity. In addition, we included the uncertainty of the observations as a prerequisite hypothesis in order to match real 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.

Two parallel typing schemes were implemented in order to accommodate data sets containing or not the measured linear particle depolarization ratios (LPDR): a) identification of mixtures from 14 aerosol mixtures (high resolution typing) if the LPDR value is available in the input data files; b) identification of 5 predominant aerosol types (low resolution typing) if the LPDR is not provided. For each scheme, three ANNs are run simultaneously, and a voting procedure selects the most probable answer.

2 METHODOLOGY

The whole algorithm has been integrated in a Python code. NATALI software is build on three

modules: (a) Input module: to convert the input files into the specific format read by the ANNs; (b) Typing module: to run the ANNs and decide on the most probable aerosol type; (c) 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 (β_{1064} , β_{532} , β_{355} , α_{532} , α_{355} , and optionally δ_{532}), identifies the layer geometrical boundaries, calculates the intensive optical parameters within each layer, their mean value and associated uncertainty.

Layer boundaries are calculated by applying the gradient method on the 1064 nm backscatter coefficient profile [3]. The inflexion points of the second derivative of the profile data (smoothed with the Savistky-Golay filter) give the top and bottom of each identified layer. Low or fine structure of the aerosol layers is revealed by a respectively higher or lower value of the smoothing parameter (adjustable) *FINESSE*. Only layers with a thickness larger than 300 m are considered relevant, for the reason of significant signal-to-noise ratio.

The intensive optical parameters and their associated uncertainties are computed for the middle part of each layer for which the signal-to-noise ratio is highest (no less than 200m mid-layer), to exclude the margins which are affected by the smoothing.

The linear particle depolarization ratio is picked-up directly from the EARLINET b532 file, if existing. For each layer, and for all of the above arrays, the module calculates averages and associated uncertainties.

Several filters are applied on 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 between acceptable limits [4]
- the relative error of each intensive optical parameter is lower than 50%

For each layer and for each intensive optical parameter, the module generates a number of values (N , adjustable) between [average –

uncertainty] and [average + uncertainty]. Data are then scrambled, considering that any combination has a similar probability to describe the reality. The cluster of possible combinations of intensive optical parameters is prepared for the ANN input.

The typing module runs in parallel the ANNs for each data set representing a layer, and applies the voting procedure to identify the most probable aerosol type.

In the case that the depolarization is available, the module runs in parallel 6 ANNs: 3 for high resolution (A1H, A2H, A3H) and 3 for low resolution typing (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 (for reasons of insufficient quality of the input data), the user has still access to information in low resolution.

If the depolarization is not available, the module runs in parallel 3 ANNs (B1L, B2L, B3L), and returns the most probable predominant aerosol type. By comparison to the low resolution typing when depolarization is available, in this case the ANNs cannot distinguish the Volcanic type, as it overlaps completely (in all existing parameters) with Dust or Continental polluted types. As such, with 3ANNs only 5 predominant types are retrieved, namely: Continental, Continental polluted, Dust, Marine, and Smoke.

A voting procedure selects the most probable answer out of the three (possibly different) individual returns. Selection is done based on the confidence level of the ANN outputs, and the stability over the uncertainty range (percentage of agreement for values between error limits). The output module prepares and saves the files in 2 formats: csv and human-readable (telegrams), and writes the log.

The software's code structure resembles the three module approach described earlier: (a) the input module: `nt_input.py`; (b) the typing module: `nt_typing.py`; (c) the output module: `nt_output.py`. These three modules are orchestrated by the `natali.py` script, which contains the high-level algorithm and calls the required module methods.

Apart from its main purpose, NATALI algorithm proved to have side-applications. In this paper, we

show the capability of the software to test the quality of the optical data and identify incorrect calibration or incorrect cloud screening.

3 RESULTS

Two data sets with continuous Raman measurements at Bucharest and Warsaw were analyzed with NATALI and presented in this paper. No additional filtering was applied to the data. Examples of the optical profiles used as inputs for the NATALI code are presented in Figs. 1 and 3 respectively. The result of the aerosol typing in low resolution (in the absence of the linear particle depolarization ratio, i.e. retrieving only the predominant component in the mixture) is presented in Figs. 2 and 4 respectively. In each of the time series, the black rectangle marks the data set for which the optical profiles are shown.

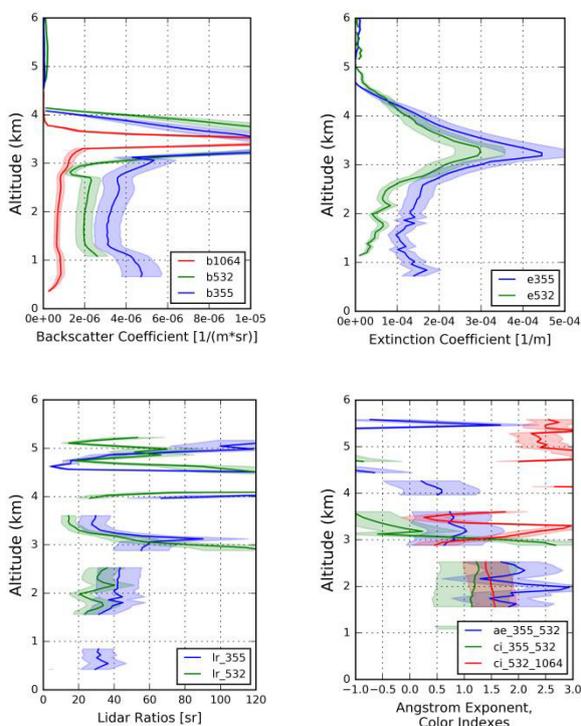


Figure 1 Aerosol optical profiles for Bucharest observations, 09/07/2012: backscatter coefficients (upper left panel), extinction coefficients (upper right panel), lidar ratios (bottom left panel), Angstrom coefficients (bottom right panel)

It can be seen from the examples at Bucharest site that NATALI cannot perform the typing if the uncertainty of the input mean-layer intensive optical parameters is higher than 100%. In this

case, NATALI returns “Unknown” and flags the output. This situation generally appears for high altitude layers, where there are few particles and the signal-to-noise ratio of the lidar, especially of the Raman channels, is low.

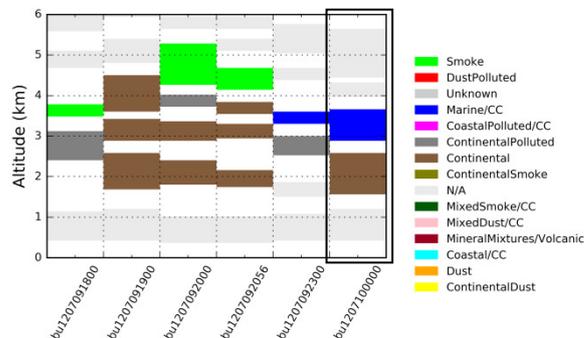


Figure 2 Results of the aerosol typing for Bucharest observations, 09/07/2012

In the incomplete overlap region, the signal-to-noise ratio is high but the values of intensive optical parameters are outside the acceptable ranges, therefore NATALI again returns “Unknown”. This is due either to an incorrect overlap correction of the raw lidar signals, or to missing information (different cuts of the backscatter or extinction profiles. For the layers with significant load and above the full overlap of the lidar NATALI is able to estimate the aerosol type even if the calibration is not perfect and/or some input parameters are not within the typical range. One particular advantage of NATALI is that it can be used to identify regions where clouds were not removed completely. For example, the marked profile in Fig. 2 shows the presence of marine particles at 3-4 km altitude in a continental site. Looking at the corresponding optical profiles (Fig. 1), it is clear that clouds are present in the lidar products. From the optical point of view, cloud particles are close to marine aerosols, large and with a low lidar ratio.

Analysis of the Warsaw observations show the difficulty of NATALI to distinguish between smoke particles and continental polluted. The extinction and backscatter-related Angstrom coefficients are very similar for these aerosol types, with smoke having a slightly higher values of the lidar ratios. However, when considering the uncertainty of the measurements, this differences in the lidar ratio are no longer visible.

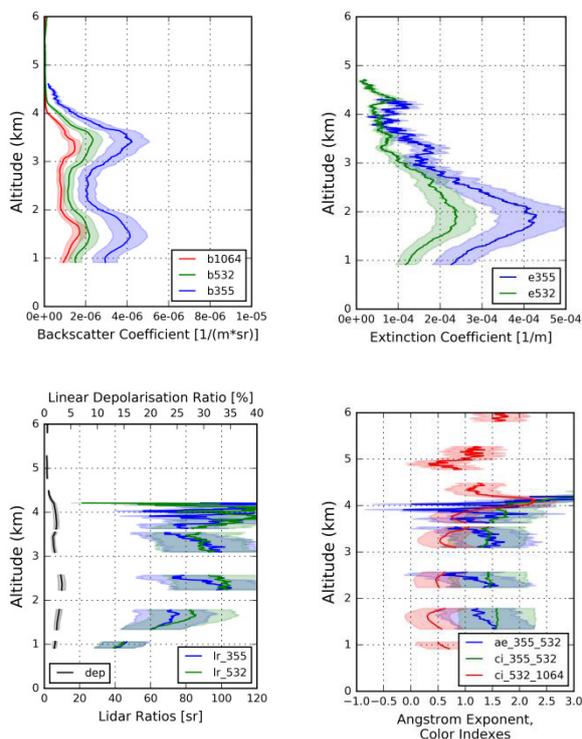


Figure 3 Aerosol optical profiles for Warsaw observations, 10/08/2015: backscatter coefficients (upper left panel), extinction coefficients (upper right panel), lidar ratios (bottom left panel), Angstrom coefficients (bottom right panel)

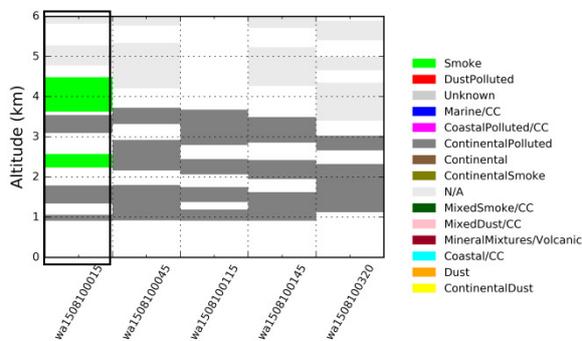


Figure 4 Results of the aerosol typing for Warsaw observations, 10/08/2015

Nevertheless, the output of the code seems fairly stable with time, considering the sensitivity of the retrieval with respect to the calibration of the input optical data.

4 CONCLUSIONS

The aerosol typing algorithm NATALI was embedded into a Python code that allows fast analysis of the multispectral lidar data.

Optical profiles are displayed in parallel with the return of the ANNs regarding the aerosol type (if depolarization information is available) or the predominant component in the mixture (otherwise). The retrieval is stable over a large range of data uncertainty, however it may fail if the signal-to-noise ratio is too low or the optical profiles are not calibrated correctly. The neural network is able to recognize the pattern of noisy data, such pattern has to be corrected, otherwise the retrieval results will be misleading. The sensitivity of the retrieval to the quality of the optical data makes NATALI a useful tool for quick check-up of the lidar products..

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, and grant agreement No. 692014 ECARS. The development and evaluation of Warsaw lidar was done in the frame of the Polish National Science Center project NCN 2012/05/E/ST10/01578

References

- [1] Tesche, M., et al., Optical and microphysical properties of smoke over Cape Verde inferred from multiwavelength lidar measurements, *Tellus B*, 63, 677–694, doi:10.1111/j.1600-0889.2011.00549.x, 2011
- [2] Nicolae, D., et al., Independent retrieval of aerosol type from lidar, *EPJ Web of Conferences* **119**, 18002, DOI: 10.1051/epjconf/201611918002, 2016
- [3] Belegante, L., et al., Retrieval of the boundary layer height from active and passive remote sensors. Comparison with a NWP model, *Acta Geophysica*, 62(2), 276–289, doi:10.2478/s11600-013-0167-4, 2014
- [4] Wandinger U. et al, HETEAC: The Aerosol Classification Model for EarthCARE, *EPJ Web of Conferences* **119**, 01004, DOI:10.1051/epjconf/201611901004, 2016