AN AUTOMATIC AEROSOL CLASSIFICATION FOR EARLINET: APPLICATION AND RESULTS

Nikolaos Papagiannopoulos[1]*, Lucia Mona[1], Vassilis Amiridis[2], Ioannis Binietoglou[2,3], Giuseppe D'Amico[1], P. Guma-Claramunt[1], Anja Schwarz[4], Lucas Alados-Arboledas[5,6], Aldo Amodeo[1], Arnoud Apituley[7], Holger Baars[4], Daniele Bortoli[8], Adolfo Comeron[9], Juan Luis Guerrero-Rascado[5,6], Panos Kokkalis[2,10], Doina Nicolae[3], Alex Papayannis[10], Gelsomina Pappalardo[1], Ulla Wandinger[4], and Matthias Wiegner[11]

[1]*Consiglio Nazionale delle Ricerche, Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Tito Scalo (PZ), Italy, [*nikolaos.papagiannopoulos@imaa.cnr.it](mailto:nikolaos.papagiannopoulos@imaa.cnr.it)*
[2] *IAASARS, National Observatory of Athens, Athens, Greece*
[3]*National Institute of R&D for Optoelectronics (INOE), Magurele, Romania*
[4]*Leibniz Institute for Tropospheric Research (TROPOS), Leipzig, Germany*
[5]*Instituto Interuniversitario de Investigación del Sistema Tierra en Andalucía (IISTA-CEAMA), Granada, España*
[6]*Dpto. Física Aplicada, Universidad de Granada, Granada, España*
[7]*Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands*
[8]*Earth Science Institute – ICT, Évora, Portugal*
[9]*Dept. of Signal Theory and Communications, Remote Sensing Lab. (RSLab), Universitat Politècnica de Catalunya, Barcelona, Spain*
[10]*Laser Remote Sensing Unit, Physics Dept., National Technical University of Athens, Greece*
[11]*Ludwig-Maximilians-Universität (LMU), Meteorologisches Institut, Munich, Germany*

ABSTRACT

Aerosol typing is essential for understanding the impact of the different aerosol sources on climate, weather system and air quality. An aerosol classification method for EARLINET (European Aerosol Research Lidar Network) measurements is introduced which makes use the Mahalanobis distance classifier. The performance of the automatic classification is tested against manually classified EARLINET data. Results of the application of the method to an extensive aerosol dataset will be presented.

1   INTRODUCTION

A thorough characterization of atmospheric aerosol particles is a key factor for accurate climate modeling [1]. The vertical profile of atmospheric particles is particularly important for the study of aerosol transport, and of the radiative forcing. Lidar measurements provide vertically resolved information on the aerosol distribution as well as on their optical properties. These parameters obtained can be used to classify into several aerosol types the various layers in the vertical.

Automatic procedures for aerosol typing combine information and based on a set of pre-specified clusters, to determine the aerosol type. Different methods for retrieving aerosol types are applied depending on the lidar abilities. The spaceborne CALIOP (Cloud-Aerosol Lidar with Orthogonal Polarization) lidar utilizes a decision-tree that takes into account lidar-derived information but also external information on geographical location, surface type, and season [2]. In contrast [3] performed an objective, multi-dimensional analysis that exclusively makes use of airborne lidar measurements. Recently, a novel method was implemented using artificial neural networks based on optical data provided by multi-wavelength Raman lidars measurements and advanced aerosol models in order to categorize aerosols [4].

EARLINET provides profiling data on a continental scale. The majority of the measuring stations operate multi-wavelength Raman lidars with the ability of linear particle depolarization ratio measurements. Based on these characteristics, EARLINET constitutes an excellent basis to distinguish different aerosol types.

Here, we apply the Mahalanobis distance

classifier to EARLINET data. Previous to classifying a sensitivity analysis is performed to identify the intensive parameters most adequate for the classification. Examples of aerosol classification are presented, followed by a short discussion of these results.

2 AUTOMATIC AEROSOL TYPE CLASSIFICATION

The Mahalanobis-based classification found great applicability in aerosol studies. For instance, the algorithm developed by [3] is lidar stand-alone classification using four lidar intensive properties with the aim of classifying aerosols into eight types. A slightly different algorithm was introduced by [5] that incorporates uncertainties in the input properties. Recently, [6] used the same classifier to produce an aerosol classification scheme based on long term AERONET data.

2.1 Methodology

The aerosol classification is performed in two parts. First, specific samples of known aerosol types are used to make the model distributions, this is called the training phase. Second, a test dataset is classified by comparison with these models and it is called the testing phase.

The calculated classifier assigns any given N-dimensional point $(x_1, x_2, \ldots, x_N)$ to the cluster that has minimum Mahalanobis distance (D_M) from that point. A cluster is defined by its mean $(\mu_1, \mu_2, \ldots, \mu_N)$ and its covariance matrix, S. In N dimensions,

$$D_M(x) = \sqrt{(x-\mu)^T S^{-1} (x-\mu)} \qquad (1)$$

2.2 Training phase

For the training phase we used EARLINET network-wide typing results already published in literature [7, 8] for a total of 64 samples. The clusters were chosen for the aerosol types to be physically meaningful and mark the sources. The 7 selected aerosol types are: pure dust (D), polluted dust (PD = Dust + Smoke), mixed dust (MD = Dust + Marine), clean continental (CC), polluted continental (PC), smoke (S), and mixed marine (MM). However, aerosol types with similar characteristics were merged to identify whether this merging enhances the correct

prediction. The S and PC were merged into a single category of small, absorbing particles. Further, the pure dust and the dust mixtures (PD and MD) were integrated into one.

Next, we performed a sensitivity analysis to identify which wavelengths and intensive parameters provide the adequate information. In contrast to the work of [3] we used three aerosol intensive properties due to lack of depolarization ratio profiles: the lidar ratio (S_{aer}), the backscatter/extinction related Ångström exponent (B-AE/E-AE), and the ratio of the lidar ratios (RS_{aer}). Two statistical parameters that highlight the strength of the selected classifiers to discriminate between clusters are used: the total and the partial Wilk's lambda (for more info check [7]). The lowest Wilk's total lambda was found to be 0.03 for the set B-AE (IR, UV), S_{aer} (VIS), and RS_{aer}; the B-AE has the most weight in the classification.

Figure 1 shows the characteristics of the training dataset in terms of the above mentioned intensive properties. The crosshairs indicate the standard deviation of the mean cluster properties. The 2-σ ellipses are calculated using the eigenvalues and eigenvectors of the covariance matrix. For the pair S_{aer} and B-AE, the various aerosol clusters tend to populate specific areas of the graph whereas for the pairs S_{aer}, RS_{aer} and B-AE, RS_{aer} no evident grouping is observed. It might be noted that the parameters do not give a good separation of the clusters, however, this is, of course, the benefit of using multivariate analysis.
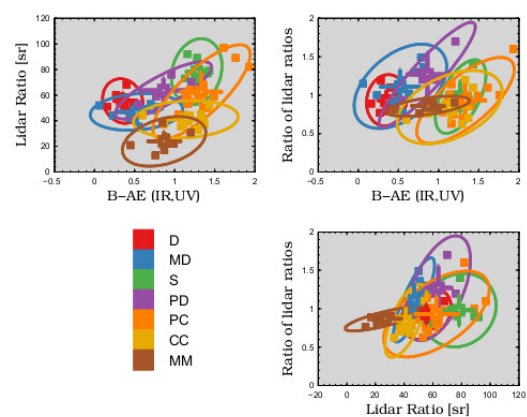

Figure 1 Pre-specified clusters delineated by two-sigma ellipses.

3 RESULTS

EARLINET data collected during the ACTRIS Summer 2012 campaign were chosen to test the automatic typing algorithm. The description of the aerosol type distribution over Europe during the campaign was obtained through a combined use of advanced lidar measurements, backward trajectory analyses and model outputs [9]. The test dataset comprises of 47 samples, 21 of which yield depolarization values.

In Tab. 1, the agreement between the manual and automatic procedures is 64 %, the samples wrongly flagged are 34 %, and the remaining 2 % is not assigned due to the low confidence ($D_M <$ 4.0 for 3 d.f.). The aerosol types that performed worse are the smoke and polluted continental categories owing to the similarities in the intensive properties. Given the noticeable signature of dust particles the agreement is limited to 54 % which can be assigned to the lack of depolarization measurements.

In the light of these discrepancies, first, we introduce a combined smoke and polluted continental category, as already discussed in Sect. 2.1, and the results are shown in the same table. The agreement increased by 13 %. Second, when dust is represented by a single cluster (the pre-specified clusters are limited to four), the agreement reached 87 %.

Table 1 Agreement in percentage of the automatic classification results with manually classified data

# of clusters	Correct Pred.	Wrong Pred.
7	64 %	36 %
6	77 %	23 %
4	87 %	13 %

The current status of EARLINET does not allow incorporating the particle linear depolarization ratio (δ_{aer}) into the automatic typing scheme. Whereas, for [3] the depolarization showed to be a robust means to discriminate the various aerosol types. Therefore, we used literature values for particle linear depolarization ratio (Tab. 2) in order to implement this parameter into the automatic algorithm. Values within the aerosol type range were randomly assigned to each sample and the Wilk's lambda distribution was calculated. Wilk's total lambda is 0.01. For B-AE

is 0.41, for S_{aer} is 0.36, for RS_{aer} is 0.48 and for δ_{aer} is 0.17. δaer for this dataset indicates that this variable has the most weight in the classification.

Table 2 The range of the linear particle depolarization (δ_{aer}) ratio for each of the aerosol types along with the bibliographic reference.

Type	δ_{aer}	References
Dust	0.28-0.33	[10]
Poll. Dust	0.12-0.23	[10]
Mixed Dust	0.13-0.17	[10]
Poll. Continent.	0.03-0.07	[11]
Clean Continent	0.01-0.07	[2]
Smoke	0.04-0.09	[11]
Mixed Marine	0.01-0.05	[10]

The Mahalanobis distance threshold, in this case, for the type inference is 4.3 (4 d.f.). Table 3 shows the agreement between the manual and the automatic procedures is 67 % (64 % without δ_{aer}). The agreement for the combined smoke and polluted continental types, now, becomes 86% (77 % without δ_{aer}); whereas for the combined dust type is 81 % (87 % without δ_{aer}). It becomes evident that the depolarization ratio increases the ability for predicting correctly the aerosol type when considering 7 clusters and 6 clusters, while it performs worse in case of 4 clusters, however the prediction rate is rather high.

Table 3 Agreement in percentage of the automatic classification results with manually classified data.

# of clusters	Correct Pred.	Wrong Pred.
7	67 %	33 %
6	86 %	14 %
4	81 %	19 %

4 CONCLUSIONS

An automatic classification for EARLINET data method based on Mahalanobis distance was presented. A Wilk's lambda analysis was performed and the three best performing classifying parameters were the S_{aer} at 532 nm; the RS_{aer} and the B-AE (IR, UV). The prediction of the automatic classification showed positive results when compared against manually classified EARLINET data. The performance was further enhanced when the aerosol clusters were merged according to aerosol types' similar characteristics. The fewer aerosol clusters make a

coarse aerosol identification and do not represent the wide aerosol spectrum. However, the positive learning success indicates that the user can select the number of clusters depending on the application. Here, we think that six aerosol clusters cover the aerosol situation with a configuration of 3 backscatter and 2 extinction coefficient profiles. Accordingly, the algorithm was trained with depolarization data that were randomly generated based on literature values, and showed the strengthening of correct prediction.

Specifically, for the dust subtype the prediction rate was higher than 50 %, even though the spectral signature is easily identifiable, the paucity of depolarization measurements makes the classification trivial as polluted dust and mixed dust reflect similar behavior with the dust type. For Polluted Continental and Smoke categories the algorithm performed the worse, this was expected as the two subtypes are overlapping and is highly difficult to separate them.

The manageability of the algorithm regarding the training dataset, the number of the clusters (i.e., aerosol types) and the classifying parameters (i.e., available intensive parameters) makes the method easily adaptable and handled by individual users. For instance, the training dataset can be easily enlarged with high quality data coming from a multitude of EARLINET stations and a longer time record. Moreover, new classifying parameters, as particle linear depolarization ratio at more wavelengths and aerosol extinction coefficient in the infrared, can be easily added as the observing capacity increases. Future work includes an extended EARLINET testing dataset for this study.

ACKNOWLEDGEMENTS

References

[1] Laj, P. et al.: Measuring atmospheric composition change, Atmos. Environ., 43, 5351-5414, 10.1016/j.atmosenv.2009.08.020, 2009.

[2] Omar, A. et al.: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, J. Atmos. Ocean. Tech., 26, 1994-2014, 10.1175/2009JTECHA1231.1, 2009.

[3] Burton, S.P. et al.: Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, Atmos. Meas. Tech., 5, 73–98, 10.5194/amt-5-73-2012, 2012.

[4] Nicolae, D. et al.: Using artificial neural networks to retrieve the aerosol type from multi-spectral lidar data, European Geosciences Union, General Assembly, Vol. 17, EGU2015-9793, 2015.

[5] Russell, P.B. et al.: A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry, J. Geophys. Res. Atmos., 119, 9838–9863, 10.1002/2013JD021411, 2014.

[6] Hamill, P. et al.: An AERONET-based aerosol classification using the Mahalanobis distance, Atmos. Env., 140, 213-233, 10.1016/j.atmosenv.2016.06.002, 2016.

[7] Pappalardo, G. et al.: EARLINET correlative measurements for CALIPSO: First intercomparison results, J. Geophys. Res., 115, D00H19, 10.1029/2009JD012147, 2010.

[8] Papagiannopoulos, N. et al.: CALIPSO climatological products: evaluation and suggestions from EARLINET, Atmos. Chem. Phys., 16, 2341-2357, 10.5194/acp-16-2341-2016, 2016a.

[9] Papagiannopoulos, N. et al.: Aerosol classification using EARLINET measurements for an intensive observational period, European Geosciences Union, General Assembly, Vol. 18, EGU2016-16026, 2016b.

[10] Groß, S. et al.: Characterization of Saharan dust, marine aerosols and mixtures of biomass-burning aerosols and dust by means of multi-wavelength depolarization and Raman lidar measurements during SAMUM 2. Tellus B, 63, 706-724, 10.3402/tellusb.v63i4.16369, 2011.

[11] Burton, S.P. et al.: Separating mixtures of aerosol types in airborne High Spectral Resolution Lidar data, Atmos. Meas. Tech., 7, 419-436, 10.5194/amt-7-419-2014, 2014.