M.Sc. Thesis

AEROSOL TYPING

BASED ON MULTIWAVELENGTH LIDAR OBSERVATIONS AND METEOROLOGICAL MODEL DATA

This study aims to compare, the results of the three different aerosol classification methods, “Mahalanobis distance automatic aerosol type classification”, “Neural Network Aerosol Typing Algorithm” (NATALI) and “Source Classification Analysis” algorithm. This comparison aims to investigate reasons of misclassification and to address which aerosol properties are needed to associate an observed aerosol layer to a certain aerosol type.

Mylonaki G. Maria
National Kapodistrian University of Athens, Faculty of Physics, Section of Environmental Physics and Meteorology
6/11/2019

National Technical University of Athens

HELLENIC REPUBLIC
National and Kapodistrian University of Athens

National

Kapodistrian University of Athens

Faculty of Physics

Section of Environmental Physics and Meteorology

AEROSOL TYPING BASED ON MULTIWAVELENGTH LIDAR OBSERVATIONS AND METEOROLOGICAL MODEL DATA

M.Sc. Thesis

Mylonaki G. Maria

M.Sc. Thesis Supervisor:

Lect. Eleni Giannakaki

3-member M.Sc. committee:

Lect. Eleni Giannakaki, Prof. Elena Floca, Prof. Alexandros Papayannis

June 2019

Acknowledgements

We acknowledge support of this work by the project “PANhellenic infrastructure for Atmospheric Composition and climatE change” (MIS 5021516) which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure”, funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund) (2018-2021).

Part of this research has been co-financed by additional funding from the European Union under grant RICA 025991 in the FP6 (2006-2011) and FP7 (2007-2013) under grant agreement no 262254.

The authors gratefully acknowledge the NOAA Air Resources Laboratory for the provision of the HYSPLIT transport and dispersion model and READY website (http://www.ready.noaa. gov) used in this publication. MODIS-FIRMS data were provided by EARTHDATA (EOSDIS-NASA).

I would personally like to thank the three members of the M.Sc. Thesis committee, Lect. Elina Giannakaki, Prof. Alexandros Papayannis and Prof. Elena Floca, for wisely guiding me for almost a year during my M.Sc. Thesis development.

Always thankful to the people *behind the scenes*, my family and friends, who patiently endure the stressful moments and are still here to celebrate the good ones. Dedicated to my colleagues, Rania Soupiona and Christianna Papanikolaou who introduced me to the apothegm “a day without laughter is a day wasted”.

Abstract

The main objective of this M.Sc. Thesis was to compare the aerosol classification results of three different aerosol classification methods, named: (1) “Mahalanobis distance automatic aerosol type classification”, (2) “Neural Network Aerosol Typing Algorithm” (NATALI) and (3) “Source Classification Analysis” (SCAN). The current work has been triggered by both the difficulties on aerosol typing due to aerosol mixtures and the demand on aerosol properties studies to quantify assessments of aerosol radiative impacts on climate. The data selection for this study has been made through the EARLINET database depending on the aerosols optical properties availability, specifically concerning the backscatter (b) extinction (a) and linear particle depolarization ratio (δ) values as 3b+2a+1δ. Ninety seven aerosol layers from four EARLINET stations (Bucharest, Kuopio, Leipzig and Potenza) have been classified.

In this M.Sc. Thesis we firstly introduce a new classification method called SCAN. It uses HYSLPIT (6-days) back-trajectories and takes into account the atmospheric layer height above the aerosol source regions (<1km for marine, <2 km for continental polluted, clean continental and dust, 3km for smoke) and distance from the potential fire spot (<8km) as well as fire spot confidence (>80%). Its classification results are based on the amount of time that the air parcel spends above certain already characterized aerosol source region. It is capable of handling a considerable amount of layers in a negligible time. However, the above criteria may not serve well all the layers under study. Also, SCAN does not take into account the tropospheric humidity which plays a crucial role to the aerosol modification.

We show that Natali is capable to well classify the pure continental polluted aerosol layers but it has a difficulty to classify the pure clean continental and pure smoke aerosol types probably due to the overlapping mean values of aerosol optical properties of these aerosol types, compared to SCAN classification algorithm.

We also show that Mahalanobis is capable to classify the pure continental polluted aerosol layers, having a difficulty to classify the pure clean continental and pure smoke aerosol types probably due to the overlapping mean values of aerosol optical properties of these aerosol types, just like Natali, compared to SCAN classification algorithm. Finally, Mahalanobis is not able to classify aerosol layers which consist of more than two aerosol types with relatively

equal contribution to the layer, compared to SCAN classification algorithm. These layers are classified as “no type” by Mahalanobis algorithm.

Concerning the mean values of aerosol optical properties, we found relatively smaller values compared to those available in the literature, especially to the “lidar ratios” and the “linear particle depolarization ratios“. This behavior might be a result of the dependence of the optical properties on the time the trajectory spent above the source area, concerning the pure types.

Περίληψη

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτέλεσε η σύγκριση των αποτελεσμάτων τριών διαφορετικών μεθόδων, “Mahalanobis distance automatic aerosol type classification”, “Neural Network Aerosol Typing Algorithm” (NATALI) and “Source Classification Analysis” (SCAN), οι οποίες κατηγοριοποιούν τα αιωρούμενα σωματίδια της ατμόσφαιρας σε συγκεκριμένες κατηγορίες. Αφορμή για την εκπόνηση της συγκεκριμένης διπλωματικής εργασίας αποτέλεσε τόσο η δυσκολία για γρήγορη και εύκολη κατηγοριοποίηση των αιωρούμενων σωματιδίων, όσο και η ζήτηση των ιδιοτήτων των αιωρούμενων σωματιδίων από τα κλιματικά και μετεωρολογικά μοντέλα με σκοπό τη μελέτη της επίδρασης των αιωρούμενων σωματιδίων στον καιρό και το κλίμα. Χρησιμοποιήθηκαν δεδομένα από τη βάση δεδομένων του δικτύου “EARLINET”, ανάλογα με τη διαθεσιμότητα των 3b+2a+1δ οπτικών ιδιοτήτων των αιωρούμενων σωματιδίων. Συνολικά, μελετήθηκαν ενενήντα επτά στρωματώσεις αιωρούμενων σωματιδίων από τέσσερις σταθμούς του δικτύου “EARLINET” (Βουκουρέστι, Κουόπιο, Λειψία και Ποτέντσα).

Ο αλγόριθμος Scan χρησιμοποιεί τις οπισθοτροχιές από το μοντέλο HYSPLIT και λαμβάνει υπόψην του το ύψος της οπισθοτροχιάς πάνω από τις πηγές (<1km για τις θαλάσσιες πηγές, <2km για τις ρυπασμένες ηπειρωτικές πηγές, μη ρυπασμένες ηπειρωτικές και πηγές ερημικής σκόνης, <3km για τις πηγές καπνού). Τα αποτελέσματά της κατηγοριοποίησης από αυτόν τον αλγόριθμο στηρίζονται στο χρόνο παραμονής της οπισθοτροχιάς πάνω από τις πηγές αιωρούμενων σωματιδίων. Είναι γρήγορος και αποτελεσματικός αλγόριθμος, ενώ δεν εξαρτάται από τις οπτικές ιδιότητες των αιωρούμενων σωματιδίων. Ωστόσο, τα παραπάνω κριτήρια ενδεχομένως να μην ικανοποιούν όλες τις εμφανιζόμενες στρωματώσεις αιωρούμενων σωματιδίων. Τέλος, το Scan δε λαμβάνει υπόψη του την επίδραση της ατμοσφαιρικής υγρασίας στις οπτικές ιδιότητες των αιωρούμενων σωματίδιων.

Παρατηρήσαμε ακόμη ότι ο αλγόριθμος Natali είναι ικανός να κατηγοριοποιήσει σωστά τα ρυπασμένα ηπειρωτικά αιωρούμενα σωματίδια. Παρατηρείται, όμως, μια δυσκολία στην κατηγοριοποίηση των μη ρυπασμένων ηπειρωτικών αιωρούμενων σωματιδίων καθώς και των αιωρούμενων σωματιδίων από καύση βιομάζας, πιθανώς λόγω των κοινών τιμών των οπτικών ιδιοτήτων που τα χαρακτηρίζουν.

Επιπρόσθετα, παρατηρούμε ότι ο αλγόριθμος Mahalanobis είναι ικανός να κατηγοριοποιήσει σωστά τα ρυπασμένα ηπειρωτικά αιωρούμενα σωματίδια,

όμως, υπάρχει και μια δυσκολία στην κατηγοριοποίηση των μη ρυπασμένων ηπειρωτικών αιωρούμενων σωματιδίων και των αιωρούμενων σωματιδίων από καύση βιομάζας, πιθανώς λόγω των κοινών τιμών των οπτικών ιδιοτήτων που τα χαρακτηρίζουν. Τέλος, παρατηρούμε ότι ο αλγόριθμος Mahalanobis, αντιμετωπίζει μια δυσκολία στο να κατηγοριοποιήσει σωστά τα αιωρούμενα σωματίδια στρωματώσεων οι οποίες περιέχουν μίξη αιωρούμενων σωματιδίων με περίπου ισάξια συνεισφορά στη μελετούμενη στρωμάτωση.

Σε ό,τι αφορά τις μέσες τιμές των οπτικών ιδιοτήτων των αιωρούμενων σωματιδίων που εντοπίζονται σε στρωματώσεις με έναν τύπο αιωρούμενων σωματιδίων, παρατηρούμε ότι είναι μικρότερες συγκριτικά με τις αντίστοιχες μέσες τιμές που αναφέρονται στην σχετική διεθνή βιβλιογραφία. Ενδεχομένως αυτό να οφείλεται στην επίδραση της ατμοσφαιρικής υγρασίας, στις ιδιότητες των αιωρούμενων σωματιδίων, καθώς επίσης, και στην επίδραση του χρόνου παραμονής τους πάνω από την πηγή στις οπτικές ιδιότητες των αιωρούμενων σωματιδίων.

Η παρούσα διπλωματική εργασία αποτελείται από πέντε κεφάλαια. Στο Κεφάλαιο 1, παρουσιάζουμε το αντικείμενο της εργασίας και τη σημασία της τεχνικής των lidar για τη μελέτη των αιωρούμενων σωματιδίων

Στο Κεφάλαιο 2, εισάγουμε τις έννοιες των αιωρούμενων σωματιδίων, καθώς και της τεχνικής “lidar”.

Το Κεφάλαιο 3 στοχεύει στην παρουσίαση της μεθοδολογίας η οποία ακολουθήθηκε κατά τη διάρκεια της παρούσας διπλωματικής εργασίας, συμπεριλαμβανομένης της λεπτομερής παρουσίασης της μεθόδου “SCAN”.

Το αντικείμενο του Κεφαλαίου 4 αποτελεί η συζήτηση των ευρημάτων της παρούσας μεθοδολογίας.

Τέλος, τα συμπεράσματα και οι στόχοι για μελλοντικές μελέτες παρουσιάζονται στο Κεφάλαιο 5.

Contents

1. Introduction

Aerosol particles affect directly the Earth's radiation budget by interacting, mainly, with the solar radiation through the processes of absorption and scattering (aerosol-radiation interaction, "ari") (Hobbs, et al., 1993). Furthermore, aerosols affect clouds formation and behavior both serving as seeds (Cloud Condensation Nuclei, Ice Nuclei) upon which cloud droplets and ice crystals form, and influencing the cloud albedo due to changing concentrations of cloud condensation and ice nuclei, also known as the Twomey effect (aerosol-cloud interaction, "aci") (Hobbs, et al., 1993).

Despite an increasing number of laboratory and field aerosol studies over the last years (IPCC, 2014), the uncertainty associated with the estimated radiative forcing attributed to aerosols and aerosol−cloud interactions has not decreased, owning to the high spatial and temporal variability of aerosol properties (IPCC, 2014). A number of atmospheric studies have targeted to address this issue by providing the mean values of aerosol optical properties of each type as input to the radiation and the numerical weather prediction models (Jacobson, et al., 2005).

One such technique is the active remote sensing of the atmosphere (Weitcamp, et al., 2005). The Lidar technique has received a considerable attention, thanks to the numerous possibilities to retrieve near real time information about the structure and the composition of the atmosphere providing very high vertical (i.e. down to few meters) and temporal (i.e. down to few minutes) resolution. Specifically, multi-wavelength Raman/depolarization lidars can be used for aerosol detection and characterization (i.e. dust, smoke, continental, etc.) as they provide vertically-resolved information of extensive (particle backscatter and extinction coefficient, volume depolarization ratio) and intensive (lidar ratio (LR), Ångström exponent ($AE_{\lambda\alpha/\lambda\beta}$), linear particle depolarization ratio (LPDR)) optical properties (Burton, et al., 2012; Groß, et al., 2013; Nicolae, et al., 2015; Giannakaki, et al., 2016; Soupiona, et al., 2018).

In the light of the above technique, Mahalanobis distance aerosol classification algorithm (Papagiannopoulos, et al., 2018) uses the lidar intensive properties (Lidar Ratio, ratio of Lidar Ratio ($LR_{\lambda1}/LR_{\lambda2}$), Ångström Exponent and Linear Particle Depolarization Ratio (if provided)) in order to classify the measured aerosol layers into a number of aerosol types.

Similarly, Neural Network Aerosol Classification Algorithm (Nicolae, et al., 2018) is based on artificial neural networks (ANNs) trained to estimate the

most probable aerosol type from, solely, a set of multispectral lidar data (Color Index (CI), Color Ratio (CR), LR, $AE_{\lambda\alpha/\lambda\beta}$ and LPDR (if provided)).

One of the classic ways to classify the aerosol layers observed is based on the trajectory analysis in order to find the air mass origin. The method is usually applied in combination with model results (e.g. DREAM) and/or satellite data (MODIS fire system, cf. Giglio, et al., 2013) in a case-by-case level. This methodology certainly involves a rather subjective element as it is not automated.

In this study we use our experience, in the by-eye identification to develop an improved automated layer identification algorithm, based on backward trajectory analysis and satellite data in combination with the application of criteria. The algorithm is called Source Classification Analysis (SCAN) and is firstly introduced in this study. It is based on the amount of time that the air parcel spends above certain already characterized source region.

Since the Mahalanobis algorithm and Natali code are newly published (Papagiannopoulos et al., 2018; Nicolae et al., 2018) they are not widely used. Moreover, Nicolae et al. (2018) published recently a report conserning the strenghts and limitations of the Natali algorithm, focusing mainly on the effect of lidar data quality to the retrievals of this method. Moreover, a first attempt of classification comparison between Natali and Mahalanobis is in progress by Voudouri et al. (2019, in progress), dealing with data from the Thessaloniki lidar station without the use of the linear particle depolarization ratio optical property.

This study aims to compare, the results of the three different aerosol classification methods, “Mahalanobis distance automatic aerosol type classification”, “Neural Network Aerosol Typing Algorithm” (NATALI) and “Source Classification Analysis” algorithm. This comparison aims to investigate reasons of misclassification and to address which aerosol properties are needed to associate an observed aerosol layer to a certain aerosol type.

This Thesis is structured in five chapters.
In Chapter 1, we introduce the lidar technique in the detection of the atmospheric aerosols.
In Chapter 2, we present the theoretical background concerning the aerosols and the lidar technique.
Chapter 3 is aiming to exhibit the methodology of this study, including a detailed presentation of SCAN.
The subject of Chapter 4 is the findings resulting from this study.

Finally, the conclusions and perspectives for future studies are provided in Chapter 5.

2. Theoretical Background

2.1. Aerosols

Whereas an aerosol is technically defined as a suspension of fine solid or liquid particles in a gas (Schmauß, al., 1929), common usage refers to the aerosol as the particulate component only (Poschl, et. al., 2005). Aerosols can be emitted directly as particles (primary aerosols) or formed in the atmosphere by gas-to-particle conversion processes (secondary aerosols) (Poschl, et. al., 2005). Atmospheric aerosols are generally considered to be the particles that range in size from a few nanometers (nucleation and Aitken mode, ultrafine, nanoparticles) to tens of micrometers (accumulation and coarse mode) in diameter (Poschl, et. al., 2005). Once airborne, particles can change their size and composition by condensation of vapor species or by evaporation, by coagulating with other particles, by chemical reaction or by activation in the presence of water supersaturation to become fog and cloud droplets (Poschl, et. al., 2005).

Eventually, particles are removed from the atmosphere by two mechanisms: deposition at Earth's surface (dry deposition) and incorporation into cloud/rain droplets during the formation of precipitation (wet deposition) (Poschl, et. al., 2005).

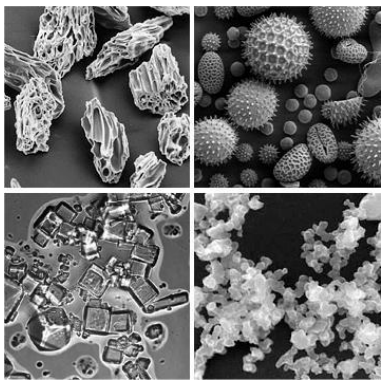

Figure 2.1 ***Pictures of volcanic, pollen, NaCl and biomass burning aerosols captured by Scanning Electron Microscopy*** **(NASA Earth Observatory : Home https://www.earthobservatory.nasa.gov/).**

2.1.1. Tropospheric Aerosol Sources

Particles in the atmosphere arise from natural sources, such as windborne dust, seaspray, and volcanoes, and from anthropogenic activities, such as combustion of fuels. Aerosols contain sulfate, ammonium, nitrate, sodium, chloride, trace metals, carbonaceous material, crustal elements, and water (Seinfeld and Pandis, 2006). The carbonaceous fraction of the aerosols consists of both elemental and organic carbon. Elemental carbon, also called black carbon or graphitic carbon, is emitted directly into the atmosphere, predominantly from

combustion processes. Particulate organic carbon is emitted directly by sources or can result from atmospheric condensation of low-volatility organic gases (Seinfeld, et al., 2006).

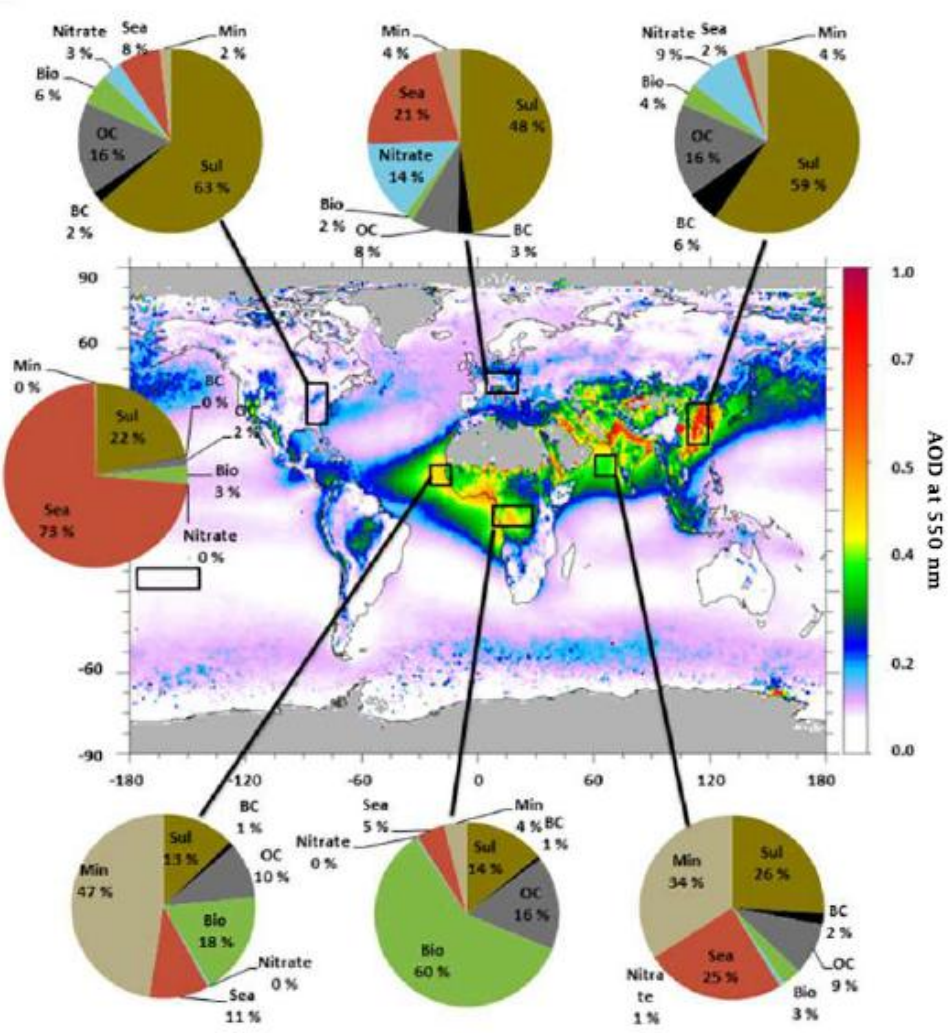

Figure 2.2 ***Global aerosol distribution as seen by Moderate Resolution Imaging Spectroradiometer (MODIS). AOD at 550 nm averaged over the 10-year period 2001-2010*** **(Remer, et al., 2008).** ***Pie charts show how various aerosol types contribute to the total AOD for different regions, as estimated by a global aerosol model*** **(Myhre, et al., 2009).** ***Aerosol types are Sul (sulfate), BC and OC from fossil fuel usage, Bio (OC and BC from BB), Nitrate, Sea (sea salt), and Min (mineral dust). Gray areas indicate lack of MODIS data. Some aerosol types, e.g. sulfate, have enhanced contributions to AOD due to hygroscopic growth. Picture was adapted from*** **Myhre et al. (2013).**

2.1.2. Aerosol Types

Continental polluted

Continental polluted aerosols are mixtures of primary particulate emissions from industries, transportation, power generation, and natural sources and secondary material formed by gas-to-particle conversion mechanisms (Seinfeld, et al., 2006). The aerosol size distribution is quite variable in urban area. Extremely high concentrations of fine particles (less than 0.1 μm in

diameter) are found close to sources (e.g., highways), but their concentration decreases rapidly with distance from the source (Seinfeld, et al., 2006).

Anthropogenic particles show a strong wavelength dependence of their optical properties, i.e., high Ångström exponent values. Moreover, they are typically small and do not significantly depolarize the backscattered light ($\delta_{aer}^{532} = 0.04 \pm 0.04$; (Heese, et al., 2016), and due to the high carbon content, these particles reveal high lidar ratios (Giannakaki, et al., 2010).

Table 2.1 ***Pure aerosol types and components (Nicolae, et al., 2018).***

Aerosol types	Basic component types	Range variation of the number density mixing ratios for aerosol components (limits are consistent with OPAC and literature)	Aspect ratio
Continental	Water soluble Insoluble Soot	0.4914-0.5914 0.0086-0.0086 0.4000-0.5000	1.100
Continental polluted	Water soluble Insoluble Soot Sulfate	0.1998-0.2998 1.8E-4-1.8E-4 0.6000-0.7000 0.1000-0.1000	1.040
Smoke	Water soluble Soot Sulfate	0.3900-0.4900 0.5000-0.6000 0.0100-0.0100	1.150
Dust	Water soluble Mineral Nucleation mode Accumulation mode Coarse mode Soot	0.1949-0.2949 0.1170-0.1170 0.0880-0.0880 0.6E-4-0.6E-4 0.5000-0.6000	0.870
Marine	Water soluble Sea Salt Accumulation mode Coarse mode Insoluble	0.1652-0.1662 0.8320-0.8320 0.0e+00-0.1e-06 0.5000-0.6000	1.007
Volcanic	Mineral Nucleation mode Accumulation mode Coarse mode Soot Sulfate	 0.0915-0.1070 0.1470-0.1719 0.4e-4-0.5e-4 0.0391-0.0457 0.6753-0.7224	0.850

Clean Continental

The clean continental type differentiates from the polluted continental type due to its less light absorbing properties. The clean continental type shows low depolarizing ability with values lower than 0.07 (Omar, et al., 2009), low lidar ratio values, i.e., 20–40 sr and relatively high Ångström exponents, i.e., 1.0–2.5 (Ansmann, et al., 2001; Giannakaki, et al., 2010).

Marine (sea spray)

In the absence of significant transport of continental aerosols, particles over the remote oceans are largely of marine origin (Prospero, et. al., 1989). Marine atmospheric particle concentrations are normally in the range of 100-300 cm^{-3}. Their size distribution is usually characterized by three modes: the Aitken *(*D_p < 0.1 μm) the accumulation (0.1 < D_p < 0.6 μm), and the coarse one (D_p > 0.6 μm) (Fitzerald, 1991).

When sea water is emitted as sea spray or spume drops, the chlorine-to-sodium mass ratio, originally 1.8:1, sometimes decreases because the chlorine is removed by sea-spray acidification (Hitchcock, 1980). Sea-spray acidification occurs when sulfuric or nitric acid enters a sea-spray drop and forces chloride to evaporate as hydrochloric acid. Some sea-spray drops lose all of their chloride in the presence of sulfuric or nitric acid.

The size of a sea-spray drop is also affected by dehydration (loss of water) which occurs when water from a drop evaporates due to a decrease in the relative humidity between the air just above the ocean surface and that a few meters higher. Dehydration increases the concentration of solute in a drop (Hitchcock, 1980).

The sea-salt particles feature a predominant coarse mode, however, they are spherical in humid conditions and weakly absorbing, in contrast to the dust particles. Therefore, they yield low particle lidar ratio values, are almost non depolarizing and exhibit low Ångström exponent values (Burton, et al., 2013; Dawson, et al., 2015). This aerosol type is mainly identifiable by the low value of the particle lidar ratio, i.e., 15–25 sr at 532nm (Burton, et al., 2012).

Dust aerosols

Desert areas around the world emit huge quantities of dust aerosols which also actually extend considerably over adjacent regions, such as oceans (Jaenicke, et al., 1978) and can be transported over very long distances (Prospero, et al., 1989; Papayannis, et al., 2008; Mona, et al., 2012). The shape of the dust size distribution is similar to that of remote continental aerosol but depends strongly on the wind velocity. Its number distribution tends to exhibit three overlapping modes at diameters of 0.01 μm or less, 0.05 μm, and 10 μm, respectively (Jaenicke, et al., 1978).

The optical properties are considerably different from the other types, thus making them easy to identify. The irregular shape and the large size (< 50μm; Mahowald, et al., 2014) lead to a significant high depolarization of the

backscattered radiation (e.g., δ_{aer}^{532} = 0.34 ± 0.02 for Saharan dust over Germany, cf. Wiegner, et al., 2011), and to medium lidar ratio values (e.g., S_{aer}^{532} = 55±10 sr (cf. Tesche, et al., 2013; Mona, et al., 2014). Desert aerosols are spectrally neutral to backscatter and extinction, and thus produce low Ångström exponent values (Wiegner, et al., 2011). Therefore, the aerosol lidar ratio, particle linear depolarization ratio, and the Ångström exponent are excellent physical parameters to characterize mineral dust and to distinguish it from other aerosol types.

Moreover, it needs to be taken into account that the dust optical properties depend on their source region and the transport pattern (Valenzuela, et al., 2014), with an important variability mainly shown in the respective lidar ratio values (Nisantzi, et al., 2015). Recently, Mamouri et al. (2013) showed that dust originating from the Arabian desert produced significantly lower lidar ratio values (34–39 sr at 532nm) than respective values (50–60 sr at 532nm) for dust particles origination from the western Saharan region.

Table 2.2 ***Conventional names of the aerosol types (Nicolae, et al., 2018).***

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

Mixed dust aerosols

Dust can be transported over continental scales. In particular, Saharan dust outbreaks to Europe and across the Atlantic Ocean have been deeply investigated, as shown by Ansmann et al. (2003), Papayannis et al. (2014) and Binietoglou et al. (2015). The study of Papayannis et al. (2008) indicated a large variability of the measured lidar ratio and Ångström exponent values among the different sites, suggesting mixing at different levels. Additionally, the mixture processes also produce large variability of intensive properties as measured at the same site (Mona, et al., 2014).

Taking into account the complex structure and composition of the dust aerosols over Europe, as well as the effects of transport and mixing on the optical and microphysical properties of these particles we consider the use of three dust groups: pure dust, mixed dust and polluted dust. The pure dust group refers to

particles for which the mixing with other aerosol types is negligible. Mixed dust refers to dust particles mixed with marine particles. This leads to less depolarizing, and less absorbing particles with respect to pure dust particles. Papagiannopoulos et al. (2016a) found this mixture to be important in the Mediterranean region and suggested its inclusion in the CALIPSO retrieval scheme for improving the accuracy of aerosol backscatter and extinction coefficient profiles.

Dust polluted aerosols

The dust polluted category consists of dust dominated mixtures with smoke and/or continental particles, which produce lower depolarization, higher lidar ratios and enhanced Ångström exponent values owing to the presence of small, spherical particles (Groß, et al., 2011; Burton, et al., 2012; Tesche, et al., 2013).

Biomass burning aerosols

Biomass burning is a major global source of atmospheric aerosols. Generally, smoke particles are relatively small, spherical, and highly absorbing that produce low depolarization, high Ångström exponents, and large lidar ratios (Amiridis, et al., 2009; Baars, et al., 2012; Nicolae, et al., 2013; Giannakaki, et al., 2016). The optical properties of smoke particles may vary due to the vegetation type of the emitting source, the combustion type (smouldering or flaming fires), and atmospheric conditions (Balis, et al., 2003). Furthermore, the particles are susceptible to changes of their optical properties during their lifetime in the atmosphere (Nicolae, et al., 2013).

Several EARLINET-based studies have focused on observations and characterization of smoke plumes (Amoiridis, et al., 2009; Ansmann, et al., 2009; Tesche, et al., 2011), demonstrating that it is a frequently encountered aerosol type over Europe. In particular, biomass burning aerosol originating from forest fires in Canada and Siberia is regularly observed between May and October (Amiridis, et al., 2009; Ortiz-Amezcua, et al., 2017). However, the similarities of the physical characteristics of smoke particles and continental particles result in similar optical properties, making these types difficult to distinguish.

Volcanic aerosols

Volcanoes are another important source of atmospheric aerosols. Volcanic eruptions eject great amounts of material in the atmosphere (tephra), while the fraction smaller than 2mm is labeled as volcanic ash. Most of these aerosols

will settle only a few tens of kilometres away from the volcano but smaller particles can travel thousands of kilometres and affect wider areas (Mattis, et al., 2010; Sicard, et al., 2012; Papayannis, et al., 2012; Kokkalis, et al., 2013; Pappalardo, et al., 2013). The optical properties of volcanic ash aerosols is generally similar to the one of desert dust, as was shown by Ansmann et al. (2010) and Wiegner et al. (2012) for fresh ash with particle linear depolarization ratios reaching 0.37 and lidar ratios of 50–65 sr. Aged volcanic particles as observed by Papayannis et al. (2012) indicate higher sphericity less non-sphericity with depolarization ratio values of about 0.1–0.25 and lidar ratios for 355nm within the range 55–67 sr and for 532nm 76–89 sr.

2.2. Atmospheric lidar

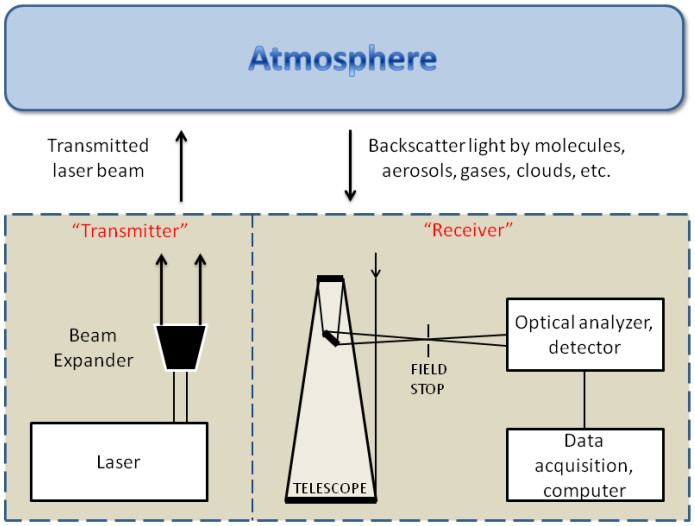

Figure 2.3 ***Lidar schematic configuration (Weitkamp, 2005).***

A simplified representation of a typical lidar system set-up is demonstrated in Figure 2.3. The transmission unit consists of a pulsed laser source, followed by a series of high reflection mirrors, and a beam expander which sends the collimated laser beam, vertically, up to the atmosphere. Part of the transmitted radiation is scattered by the atmospheric components (i.e., gases, molecules, aerosols, clouds) backward to the lidar system, where it is collected by an optical telescope. The backscattered laser light is driven to an optical analyzer (i.e. spectrometer) where the optical signal is, first, spectrally separated, then, amplified and, finally, transformed to an electrical signal. Finally, the signal is digitized and stored in a computer unit for further signal processing and analysis.

2.2.1. Aerosol optical properties from a multi-wavelength elastic/Raman lidar system

The backscatter coefficient $\beta(R,\lambda)$ defines the scattering coefficient at the backward direction (i.e., at a scattering angle of θ = 180°). This parameter determines the strength of the lidar return at the operational wavelength λ at a distance R far from the lidar system. The mathematical expression of the backscatter signal is:

$$\boldsymbol{\beta(R,\lambda) = \sum N_j(R) \frac{d\sigma_{j,sca}}{d\Omega}(\pi,\lambda)[m^{-1}sr^{-1}]} \quad \mathbf{1}$$

where N_j denotes the concentration of scattering "particles" of kind j in the illuminated volume and $d\sigma_{j,sca}/d\Omega$ is the "particles" differential scattering cross section for the backward direction. Nevertheless, since this backscatter coefficient denotes the backscattered light from both aerosols and molecules in the atmosphere, the contribution of molecules to the total backscatter could be defined, so as the aerosol backscatter coefficient can be estimated.

The most commonly used method for calculating the aerosol backscatter coefficient is the Fernald-Klett method exclusively described in literature Fernald (1984) and Klett (1981). The main uncertainty for the accurate calculation of aerosol backscatter coefficient in this method is introduced by the assumption of the lidar ratio value, which is characteristic for the type of the detected aerosols. Another uncertainty can be introduced by the calibration height where aerosol contribution in the backscatter profile is considered negligible (Weitcamp, 2005).

The most accurate technique to derive the aerosol extinction and backscatter coefficients is the so-called Raman lidar one. The main advantage of this technique compared to the elastic one is that it allows the independent retrieval of the aerosol extinction and backscatter coefficients, without any assumption of the aerosol S_λ.

The Raman lidar equation includes the backscatter coefficients $\beta_{mol}(\lambda,z)$, $\beta_{aer}(\lambda,z)$ and extinction coefficients $\alpha_{mol}(\lambda,z)$, $\alpha_{aer}(\lambda,z)$ of molecules and aerosols respectively. The solution of this differential equation with making use of the Raman channels has been proposed (Ansmann, et al., 1990, Papayannis, et al., 1990) for the simultaneous retrieval of the aerosols' backscatter and extinction coefficients at λ = 355 nm and λ = 532 nm. The inelastic Raman signals are only due to the scattering from Nitrogen N_2 molecules and therefore, the backscattering cross-section determines the intensity of the signals.

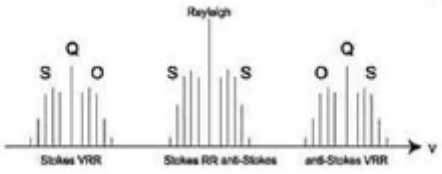

Figure 2.4 ***Raman spectrum of a diatomic molecule. The Rayleigh line is at the same frequency of the incident photon. Just around the Rayleigh line, the rotational Stokes and anti-Stokes lines lie to the left towards lower energies and to the right towards higher energies respectively*** **(Wandinger, 2005).**

In 1928, Sir Raman observed the frequency-shifted lines in the spectrum of scattered light (Raman, et al., 1990). The frequency shift relative to the incident light frequency corresponds to the vibrational and rotational frequencies of the scattering molecules. The frequency shift can be towards lower energies $\lambda_{Ra} > \lambda_0$ (Stokes lines) or higher energies $\lambda_{Ra} < \lambda_0$ (anti-Stokes lines). The total Raman spectrum of a diatomic molecule can be seen in Figure 2.4. When the scattered light is coherent to the incident light, Rayleigh scattering occurs which is much stronger than the Raman scattering. When only molecular rotations occur, the Rotational Raman (RR) lines are just around the Rayleigh line. However, the Vibrational Rotational Raman (VRR) lines are better separated in frequency from the incident light. VRR lines consist of Q, S and O branches. The anti-Stokes lines are weaker than the Stokes lines and usually cannot be observed.

A molecule is always at a specific vibrational-rotational energy. The frequency shift which is caused by the change in vibrational, rotational or vibrational-rotational energy level of the scattering molecule is estimated by

$$\boldsymbol{\Delta\bar{\upsilon} = \bar{\upsilon}_i - \bar{\upsilon}_s = \frac{\Delta E}{hc_0}\mathbf{1}}$$

where $\bar{\upsilon}_i$ is the frequency of the incident photon, $\bar{\upsilon}_s$ is the frequency of the scattered photon, ΔE is the difference between the molecular energy levels, h is Planck's constant and c_0 the speed of light in vacuum (Long, 1977).

Figure 2.5 illustrates the rotational-vibrational Raman spectra for several atmospheric molecules such as silicon dioxide (SiO_2), oxygen (O_2), nitrogen (N_2) and water vapor (H_2O) in gas, liquid and solid phase. The incident light at 355 nm excites the atmospheric nitrogen molecules which emits light at the Stokes line of 387 nm. When the incident beam is at 532 nm, the Raman scattered photon is detected at 607 nm. The Raman lidar can also detect Raman lines of water vapor at 407 nm when incident light beam is at 355 nm.

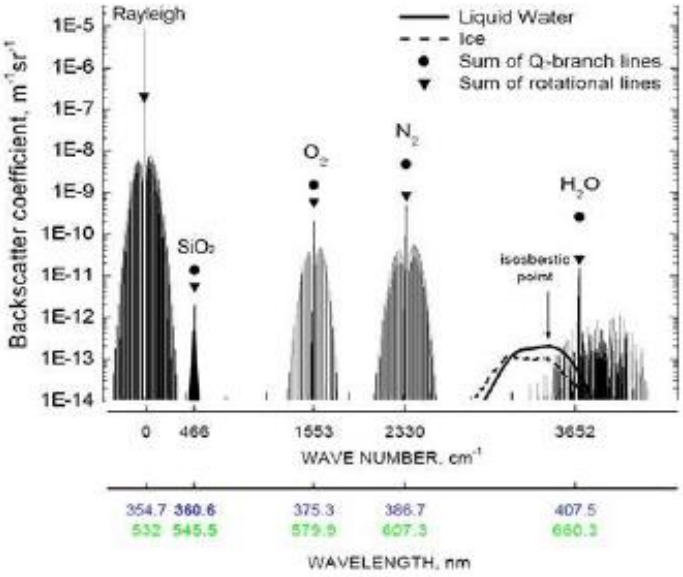

Figure 2.5 ***Rotational vibrational Raman lines. The Raman backscatter spectrum of the atmosphere for incident laser wavelength of 355 and 532 nm. Picture was adapted from*** **Avila et al. (1999).**

2.2.2. EARLINET Lidar Network

EARLINET (www.earlinet.org) was established in 2000, providing aerosol profiling data on a continental scale, and now is part of the Aerosols, Clouds, and Trace gases Research InfraStructure (ACTRIS; www.actris.eu/). In these 19 years of continuous existence, EARLINET has evolved both in the number of contributing stations, as well as in its observing capacity (Pappalardo, et al., 2014). Currently, 30 stations are submitting aerosol extinction and/or backscatter coefficient profiles to the EARLINET database, according to EARLINET's measurement schedule (one daytime and two nighttime measurements per week). Therefore, these systematic observations consolidate a 4D European quantitative and statistically significant aerosol survey. Further measurements are devoted to special events, such as volcanic eruptions, forest fires, and desert dust outbreaks. Moreover, EARLINET provides correlative measurements during CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) overpasses on each EARLINET station in order to validate satellite products (Mamouri, et al., 2009; Mona, et al., 2009).

The majority of the EARLINET stations (67% of the stations) as described by Pappalardo et al. (2014) operate multi-wavelength Raman lidars that combine a set of elastic and nitrogen inelastic channels, typically consisting of three elastic and two inelastic Raman channels (the so-called $3\beta + 2\alpha$ configuration). In particular, they provide the aerosol extinction (at 355nm and 532nm), and backscatter coefficients (at 355 nm, 532 nm, and 1064 nm). This configuration allows the retrieval of the range-resolved particle lidar ratio at 355nm and

532nm (S_{aer}). This intensive parameter depends on the shape, size, and chemical composition of the aerosol (Müller, et al., 2007). When the lidar ratio is available for more than one wavelength, the corresponding color ratio can be also retrieved ($S_{aer}^{\lambda 1}/S_{aer}^{\lambda 2}$).

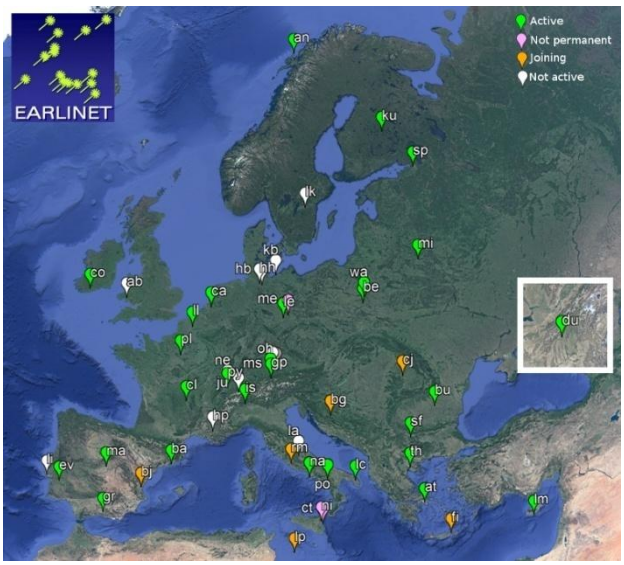

Figure 2.6 ***EARLINET Lidar stations (last updated: 21/03/2019)***

This quantity is a robust means to characterize the ageing status of smoke particles, as well as the spectral dependence of aerosol (Müller, et al., 2007a; Nicolae, et al., 2013). The combination of the optical data allows the retrieval of the size sensitive backscatter and/or extinction related Ångström exponent and can be calculated as

$$\kappa_X = \frac{\ln [X(\lambda_1)/X(\lambda_2)]}{\ln (\lambda_1/\lambda_2)} \quad \mathbf{2}$$

with X denoting the backscatter β or extinction coefficient α for a set of wavelengths, λ_1 and λ_2. Moreover, 52% of EARLINET stations (Pappalardo, et al., 2014) are equipped with depolarization channels, thus providing profiles of the particle linear depolarization ratio. It can be calculated according to (Biele, et al., 2000; Freudenthaler, et al., 2009:

$$\delta_{aer}^{\lambda} = \frac{(1+\delta_m)\delta_v R-(1+\delta_v)\delta_m}{(1+\delta_m)R-(1+\delta_v)} \quad \mathbf{4}$$

with R the backscatter ratio, δ_m the molecular depolarization, and δ_v the volume depolarization ratio (Bravo-Aranda, et al., 2016). This parameter provides information on the particle shape, thus enhancing the aerosol typing strength of the network. Under favourable conditions, the aerosol microphysical properties, such as the effective radius, the volume concentration and the

refractive index can also be retrieved through complex numerical algorithms (Müller, et al., 2004; Veselovskii, et al., 2010; Samaras, et al., 2015; Chaikovsky, et al., 2016).

2.3. Hybrid Single-Particle Lagrangian Integrated Trajectory Model, *HYSPLIT*

The HYSPLIT model (Draxler, et al., 1998) is a complete system for computing simple air parcel trajectories, as well as complex transport, dispersion, chemical transformation, and deposition simulations. HYSPLIT continues to be one of the most extensively used atmospheric transport and dispersion models in the atmospheric sciences community. A common application is a back trajectory analysis to determine the origin of air masses and establish source-receptor relationships.

The model calculation method is a hybrid one between the Lagrangian approach (using a moving frame of reference for the advection and diffusion calculations as the trajectories or air parcels move from their initial location) and the Eulerian methodology, which uses a fixed three-dimensional grid as a frame of reference to compute pollutant air concentrations (The model name, no longer meant as an acronym, originally reflected this hybrid computational approach). HYSPLIT has evolved over more than 30 years, from estimating simplified single trajectories based on radiosonde observations to a system accounting for multiple interacting pollutants transported, dispersed, and deposited over local to global scales.

The dispersion of a pollutant is calculated by assuming either puff or particle dispersion. In the puff model, puffs expand until they exceed the size of the meteorological grid cell (either horizontally or vertically) and then split into several new puffs, each with its share of the pollutant mass. In the particle model, a fixed number of particles are advected about the model domain by the mean wind field and spread by a turbulent component. The model's default configuration assumes a 3-dimensional particle distribution (horizontal and vertical).

2.4. Moderate Resolution Imaging Spectroradiometer, *MODIS*

As part of NASA's Earth Observing System (EOS), MODIS is carried on both the Terra and Aqua satellites. The MODIS active fire products fall within the suite of terrestrial products and provide information about actively burning fires, including their location and timing, instantaneous radiative power, and

smoldering ratio, presented at a selection of spatial and temporal scales (Giglio et al., 2002; Kaufman et al., 1998).

An algorithm examines each pixel of the MODIS swath and ultimately assigns it to one of the following classes: missing data, cloud, water, non-fire, fire, or unknown. Pixels lacking valid data are immediately classified as missing data and excluded from further consideration. Cloud and water pixels are identified using cloud and water masks and are assigned to the classes cloud and water, respectively. The fire detection algorithm considers only those land pixels that remain.

Figure 2.9 ***World satellite image. Red dots represent possible fire sources for the time period 2019/03/27-2019/03/28 as seen by MODIS Aqua and Terra satellites (https://firms.modaps.eosdis.nasa.gov/map/).***

There are two logical paths through which fire pixels can be identified. The first consists of a simple absolute thresh-old test. This threshold must be set sufficiently high so that it is triggered only by very unambiguous fire pixels, i.e. those with very little chance of being a false alarm. The second path consists of a series of contextual tests designed to identify the majority of active fire pixels that are less obvious.

The variable ‘Confidence’ is based on a collection of intermediate algorithm quantities used in the detection process. It is intended to help users gauge the quality of individual hotspot/fire pixels. Confidence values are set to low, nominal and high. Low confidence daytime fire pixels are typically associated with areas of sun glint and lower relative temperature anomaly (<15K) in the mid-infrared channel. Nominal confidence pixels are those free of potential sun glint contamination during the day and marked by strong (>15K) temperature anomaly in either day or nighttime data. High confidence fire pixels are associated with day or nighttime saturated pixels.

3. Methodology

3.1. Experimental sites

Lidar station selection depended on the availability of the vertical profiles of the full set (3β + 2α + δ) of aerosol optical properties: backscatter coefficient (b_{355}, b_{532}, b_{1064}), extinction coefficient (e_{355}, e_{532}), Lidar Ratio (LR_{355}, LR_{532}), Ångström exponent ($AE_{355/532}$, $CI_{355/532}$, $CI_{532/1064}$) and Linear Particle Depolarization Ratio ($LPDR_{532}$) at the EARLINET database (https://www.earlinet.org) within the years 2014-2018. The 4 lidar stations performed the used measurements are: Kuopio, Finland; Bucharest, Romania; Leipzig, Germany; and Potenza, Italy (Table 3.1).

Table 3.1 ***EARLINET lidar station information.***

Location	ACTRIS Code	Institute	Coordinates (lat, long, altitude amsl)	Reference	No of layers	Year
Bucharest	INO	National Institute of R&D for Optoelectronics (INOE) Link	44.35 N, 26.03 E, 93 m	Nemuc, et al. 2013	7	2017
Kuopio	KUO	Finnish Meteorological Institute (FMI), Atmospheric Research Centre of Eastern Finland, Kuopio Link	62.74 N, 27.54 E, 190 m	Althausen, et al., 2009, Engelmann, et al., 2016	9	2015, 2016
Leipzig	LEI	Leibniz Institute for Tropospheric Research, Leipzig Link	51.35 N, 12.43 E, 90 m	Althausen, et al., 2009, Engelmann, et al., 2016	17	2018
Potenza	POT	Consiglio Nazionale delle Ricerche - Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Potenza Link	40.60 N, 15.72 E, 760 m	Madonna, et al., 2011	64	2015-2016

Bucharest

Romania is crossroad of different air masses (Fig. 3.1) that arrive from regions (Greece, Ukraine, Russia) that are affected episodically by strong biomass burning events (Amiridis, et al., 2010). Nicolae, et al., (2013) studied optical and microphysical properties of long-range transported biomass burning (BB) aerosols and their variation with atmospheric evolution (ageing) observed over the lidar station of Magurele, Bucharest.

The city and region of Bucharest is urban with an intense traffic and surrounded by industrial platforms, which gives rise to a variety of aerosol and gaseous pollutants (Nicolae, et al., 2006). However the most relevant in the characterization of aerosol properties in this area is the desert dust events which distort the atmospheric composition (Nemuc, et al., 2008).

Kuopio

Kuopio (population 91.000) is the principal town of the province of Northern Savo, in the eastern part of central Finland, 330 km northeast from Helsinki, the capital of Finland (Fig. 3.1). The main part of the district of Northern Savo, and especially the neighborhood of Kuopio, belongs to the southern boreal climatic zone and is characterized by forests with conifer (mostly pine and spruce) and deciduous (mostly birch) trees, an undulating terrain with rocky soil and moderate height hills, and lots of long lakes in the northwest–southeast direction. The vast lake district acts as a heat storage and increases the nightly temperatures in summers, thus lengthening the growing period (Leskinen, et al., 2009).

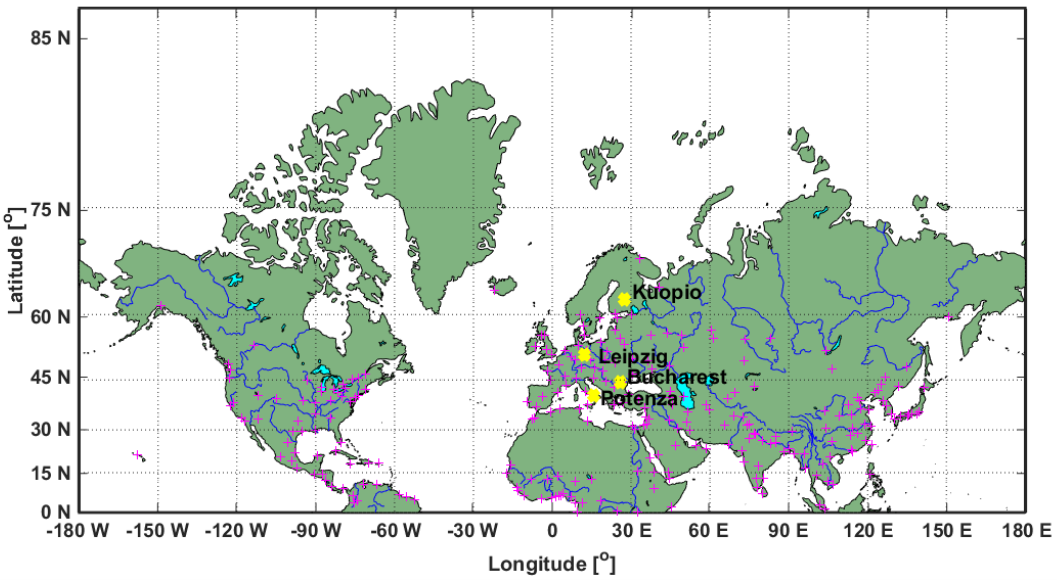

Figure 3.1 ***Geographical location of the 4 EARLINET stations considered in this study (yellow diamonds). Major European cities are also presented with pink dots. Major European lakes are indicated with cyan and rivers with blue.***

The Puijo measurement station is on the top of an observation tower, 306 m a.s.l. and 224 m above the surrounding lake level. The tower is a 75 m high building on the Puijo hill, approximately 2 km northwest of the center of Kuopio (Leskinen, et al., 2009).

The most significant local sources are traffic on highways (national/European highway 5/E63 and national highway 17), especially between Kuopio and Siilinjärvi with approximately 30.000 vehicles/day, the local traffic in Kuopio, and point sources, such as a district heating plant 3 km south of Puijo and a pulp mill 5 km north-east of Puijo (Leskinen, et al., 2009).

Leipzig

Mattis et al. (2008) presented geometrical properties and seasonal variations of appearance of aerosol particle pollution in the free troposphere over the central European lidar site at Leipzig, Germany (Fig 3.1). The lofted particle layers largely resulted from intercontinental long-range transport. In 19% of all regular measurements, free-tropospheric pollution was advected from North America. Forest-fire smoke from Canada and anthropogenic pollution from urban areas of the United States of America and Canada were the sources of the particle layers. They find a strong seasonal dependence of occurrence of these layers with a peak in June–August of each year. In a few cases they observed forest-fire smoke advected from Siberia and east Asia with winds from westerly directions. Pollution advected from areas north of 70°N presented another transport channel. That pollution consisted of Arctic haze or mixtures of haze with anthropogenic pollution. The main occurrence of such particle layers was around springtime of each year. Import of mineral dust from the Sahara represents another transport path. Most of such cases were observed during late spring time and summer time. Free-tropospheric pollution advected from east and southeast Europe and Russia presented one transport channel from within the Euro-Asian continent.

Potenza

CIAO, the most advanced infrastructure for ground based remote sensing in Italy, is located in Tito Scalo, Potenza, Southern Italy, on the Apennine mountains (40.60◦N, 15.72◦E, 760 m a.s.l.), less than 150 km from the West, South and East coasts (Figure 3.1). The site is in a plain surrounded by low mountains (<1100 m a.s.l.). The observatory operates in a typical mountain weather strongly influenced by Mediterranean atmospheric circulation, resulting in generally dry, hot summers and cold winters. In this location, phenomena like orographically-induced effects on cloud formation can be studied (Madonna, et al. 2011). The site is particularly interesting for studying aerosol properties because it is affected by a quite large number of Saharan dust intrusions per year (Mona, et al., 2014) and it is located 300 km far from the Etna Volcano (Pappalardo, et al., 2004).

3.2. Experimental data

48 dates of lidar observations by multiwavelength Raman and depolarization lidars from the aforementioned lidar stations have been studied for this Thesis.

Table 3.2 ***Dates of lidar observations used for this study (2015-2018).***

Year (yyyy)	Date (dd/mm)	INO	KUO	LEI	POT	No of layers	SUM (dates)	SUM (layers)
2015	22/01		*			1	10	**21**
	13/04				*	2		
	07/05				*	2		
	04/06				*	3		
	11/06				*	4		
	16/07				*	1		
	20/07				*	1		
	30/07		*		*	1+4		
	24/08		*			1		
	03/09				*	1		
2016	04/01		*			1	14	**21**
	07/01		*			2		
	21/03		*			1		
	28/03		*			1		
	04/04				*	1		
	11/04				*	1		
	14/04		*			1		
	23/06				*	2		
	27/06				*	2		
	04/07				*	3		
	07/07				*	1		
	11/07				*	2		
	25/08				*	2		
	29/08				*	1		
2017	23/03				*	1	17	**38**
	30/03				*	2		
	27/04				*	1		
	22/05				*	1		
	08/06				*	2		
	12/06				*	1		
	15/06				*	4		
	22/06				*	4		
	13/07				*	1		
	19/07				*	2		
	20/07				*	4		
	24/08	*				3		
	28/08				*	5		
	31/08	*			*	2+2		
	07/09	*				1		
	21/09	*				1		
	19/10				*	1		
2018	21/06			*		1	7	**17**
	19/07			*		3		
	26/07			*		2		
	30/07			*		3		
	02/08			*		2		
	16/08			*		1		
	27/08			*		5		
SUM (dates)		4	8	7	31		**Sum of layers throughout the years: 97**	
SUM (layers)		**7**	**9**	**17**	**64**			

4 dates by Bucharest site for the year 2017, 8 of Kuopio site for the years 2015-2016, 7 of Leipzig site for the year 2018 and 31 of Potenza site for the years 2015-2017 (Table 3.2). For some dates, more than one layer has been observed increasing the number of layers to 97.

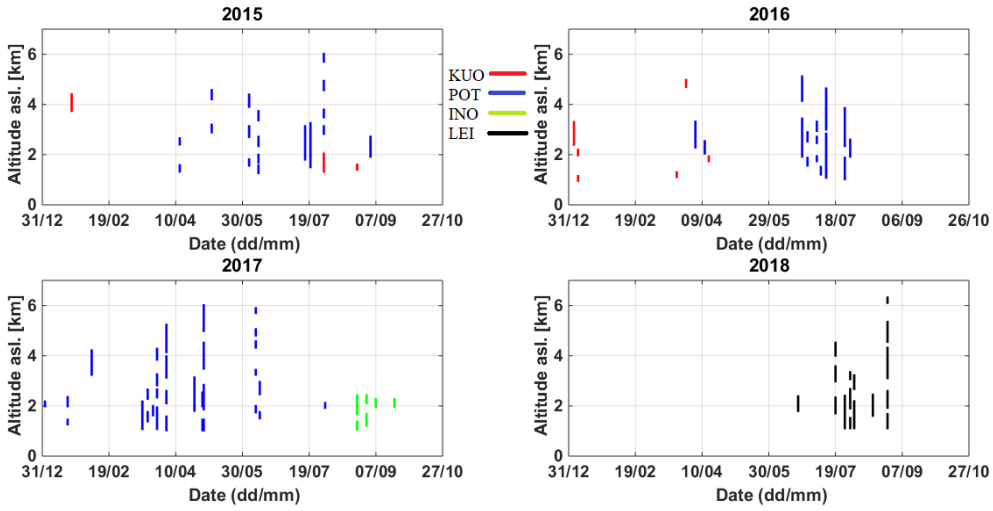

Figure 3.2 ***Geometrical properties of the aerosol layers for this study (2015-2018).***

The aerosol layer boundaries were calculated by applying the gradient method on the 1064 nm backscatter coefficient profile (Belegante, et al., 2014).

In Figure 3.2 the bottoms and tops of the aerosol layers are presented for each year (2015, 2016, 2017, 2018). The horizontal axis stands for the date (dd/mm) of the observed layers beginning from 01/01 to 27/10 of each year and the vertical axis stands for the altitude [km]. The red lines present the aerosol layers arriving over Kuopio, the blue lines present the aerosol layers arriving over Potenza, the green lines present the aerosol layers arriving over Bucharest and finally, the black lines present the aerosol layers arriving over Leipzig.

3.3. Classification

3.3.1. Source Classification Analysis, *SCAN*

SCAN aerosol classification uses the backward trajectories created with HYSPLIT in combination with FIRMS satellite fire data to provide the possible source of air mass observed above the stations, counting the time that the air parcel spends above certain source regions (eg. dust, marine, continental polluted and clean continental).

To identify the source region of the aerosol layers, SCAN makes use of a number of criteria. Firstly, it associates each observed layer to the back-trajectories calculated by HYSPLIT depending on the mean height of the observed layer and its arriving date and time. For this study, 6 days back-

trajectories (144 hours) have been calculated for 6 different arriving heights (1500m, 2000m, 2500m, 3000m, 4000m, 5000m) for each layer's date.

Using the mean of each aerosol layer's geometrical properties, each aerosol layer is associated to the closest (in terms of the minimum difference with HYSPLIT's arriving height) HYSPLIT trajectory (Table 3.3)

Table 3.3 ***Geometrical properties of the aerosol layers for this study (2015-2018) and HYSPLIT arriving height and time.***

Year (yyyy)	Date (dd/mm-hh:mm UTC)	Bottom (m)	Top (m)	HYSPLIT arriving height (m)
2015	22/01-18:05	3744	4402	4000
	13/04-19:45	1330	1570	1500
		2410	2650	2500
	07/05-20:45	2890	3190	3000
		3910	4570	4000
	04/06-19:17	1570	1810	1500
		2710	3130	3000
		3910	4590	4000
	11/06-23:35	1270	1570	1500
		1690	1990	2000
		2350	2710	2500
		3370	3730	4000
	16/07-21:00	1810	3130	2500
	20/07-23:59	1510	3250	3000
	30/07-19:25	1324	2041	1500
	21:26	2830	3130	3000
		3490	3790	4000
		4570	4930	5000
		5710	6010	5000
	24/08-18:00	1204	1802	1500
	03/09-20:16	1930	2710	2500
	04/01-15:18	2399	3296	3000
	07/01-15:19	965	1145	1500
		1982	2191	2000
	21/03-18:21	1115	1294	1500
	28/03-23:11	4701	4970	5000
	04/04-18:47	2290	3310	3000
	11/04-20:27	2050	2530	2500
	14/04-21:00	1742	1922	2000
2016	23/06-19:30	1930	3430	3000
		4150	5110	5000
	27/06-21:07	1570	1870	1500
		2530	2890	2500
	04/07-19:56	1750	1930	2000
		2470	2710	2500
		2950	3310	3000
	07/07-19:01	1210	1510	1500
	11/07-19:22	1090	2830	2000
		3010	4630	4000
	25/08-18:29	1030	1870	1500
		2350	3850	3000
	29/08-19:16	1930	2590	2500
	23/03-18:10	1990	2170	2000
2017	30/03-18:08	1270	1450	1500
		1990	2350	2000
	27/04-19:48	3250	4210	4000
	22/05-19:01	1090	2170	1500
	08/06-19:14	1390	1750	1500
		2290	2650	2500
	12/06-19:29	1630	1990	2000
	15/06-20:00	1090	1930	1500

		2350	2650	2500
		2830	3250	3000
		3850	4270	4000
	22/06-19:18	1030	1570	1500
		2110	2590	2500
		3130	3970	4000
		4150	5230	5000
	13/07-20:15	1810	3130	2500
	19/07-18:57	1030	1450	1500
		1990	2530	2500
	20/07-20:15	1030	1450	1500
		1870	2830	2500
		3490	4510	4000
		4990	6010	5000
	24/08-16:56	1068	1368	1500
		1968	2418	2000
	17:59	1668	2268	2000
	28/08-18:45	1750	1990	2000
		3250	3430	3000
		4330	4570	4000
		4810	5050	5000
		5710	5890	5000
	31/08-17:59	1218	1668	1500
		2118	2418	2500
	18:26	1510	1750	1500
		2470	2950	2500
	07/09-16:33	1968	2268	2000
	21/09-16:16	1968	2268	2000
	19/10-17:14	1930	2110	2000
	21/06-20:30	1807	2375	2000
	19/07-21:30	1710	2330	2000
		2988	3571	3000
		4011	4512	4000
	26/07-20:00	1112	1352	1500
		1434	2398	2000
	30/07-21:45	1112	1516	1500
		1904	2667	2500
2018		3070	3339	3000
	02/08-23:00	1120	2181	1500
		2681	3212	3000
	16/08-21:20	1613	2442	2000
	27/08-19:00	1112	1673	1500
		1934	2592	2500
		3107	4310	4000
		4557	5341	5000
		6111	6313	5000

The back-trajectories calculated by HYSPLIT are shown in Figure 3.3. The blue color illustrates the back-trajectories arriving over Potenza, the black color illustrates the back-trajectories arriving over Leipzig, the red color illustrates the back-trajectories arriving over Kuopio and finally, the green color illustrates the back-trajectories arriving over Bucharest.

Afterwards, SCAN assumes predifened, in terms of aerosol sources, regions (Fig. 3.4 colored squares) (Penning de Vries, et al., 2015): marine (Fig. 3.4 blue squares), clean continental (Fig. 3.4 brown squares), polluted continental (Fig. 3.4 black squares) and dust (Fig. 3.4 orange squares).

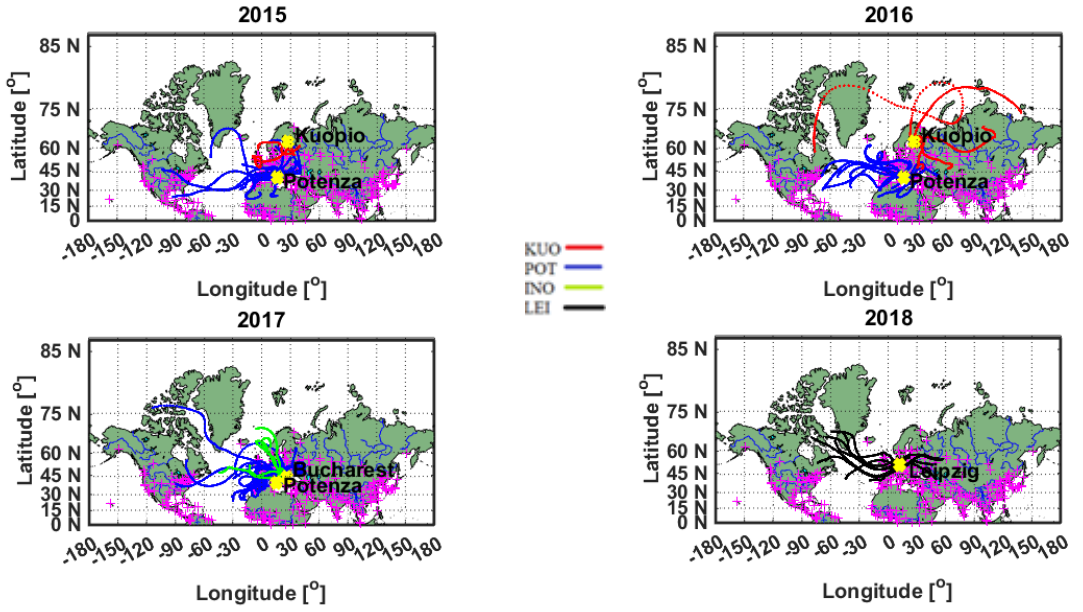

Figure 3.3 ***HYSPLIT 6-days back-trajectories arriving at Potenza (blue lines), Bucharest (green lines), Kuopio (red lines), Leipzig (black lines), used for this study (2015-2018).***

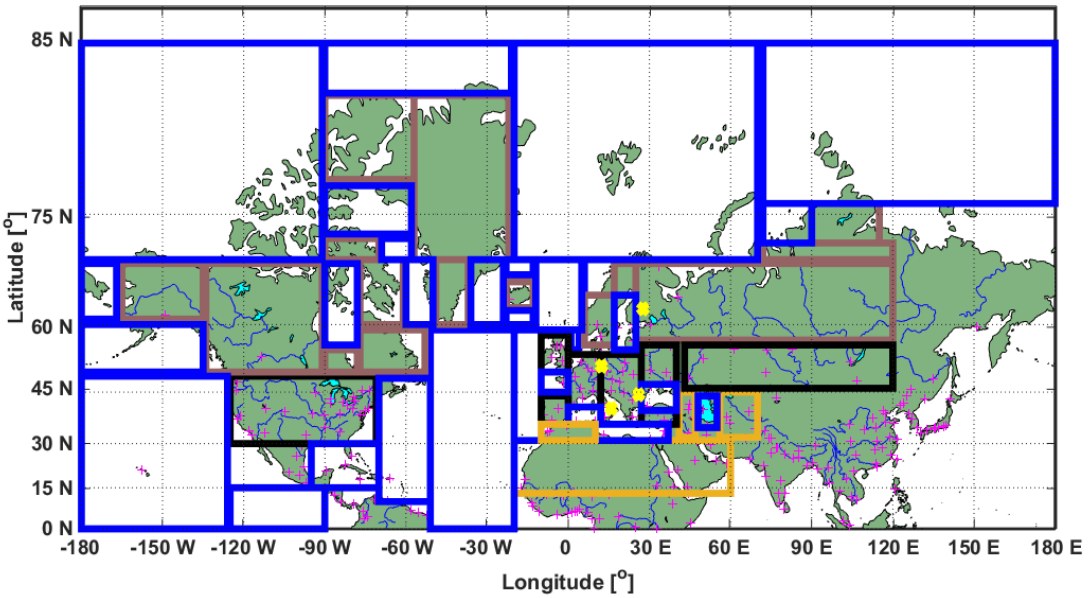

Figure 3.4 ***SCAN's colored squares according to the aerosol source type. Orange squares stand for dust aerosol sources, blue squares for marine aerosol sources, brown squares for clean continental aerosol sources and black squares for continental polluted aerosol sources.***

Finally, fire spots are defined using FIRMS fire/hotspot information according to the duration of HYSPLIT back-trajectories. A hotspot is assumed significant if the value of variable 'confidence' is higher than 80% (Amiridis, et al., 2010). SCAN counts the time that the air parcel spends above the fire hotspots, or within 8km distance from them, with height below 3km taking into account that the mean injection height for fires is 3km according to Amiridis et al., 2010.

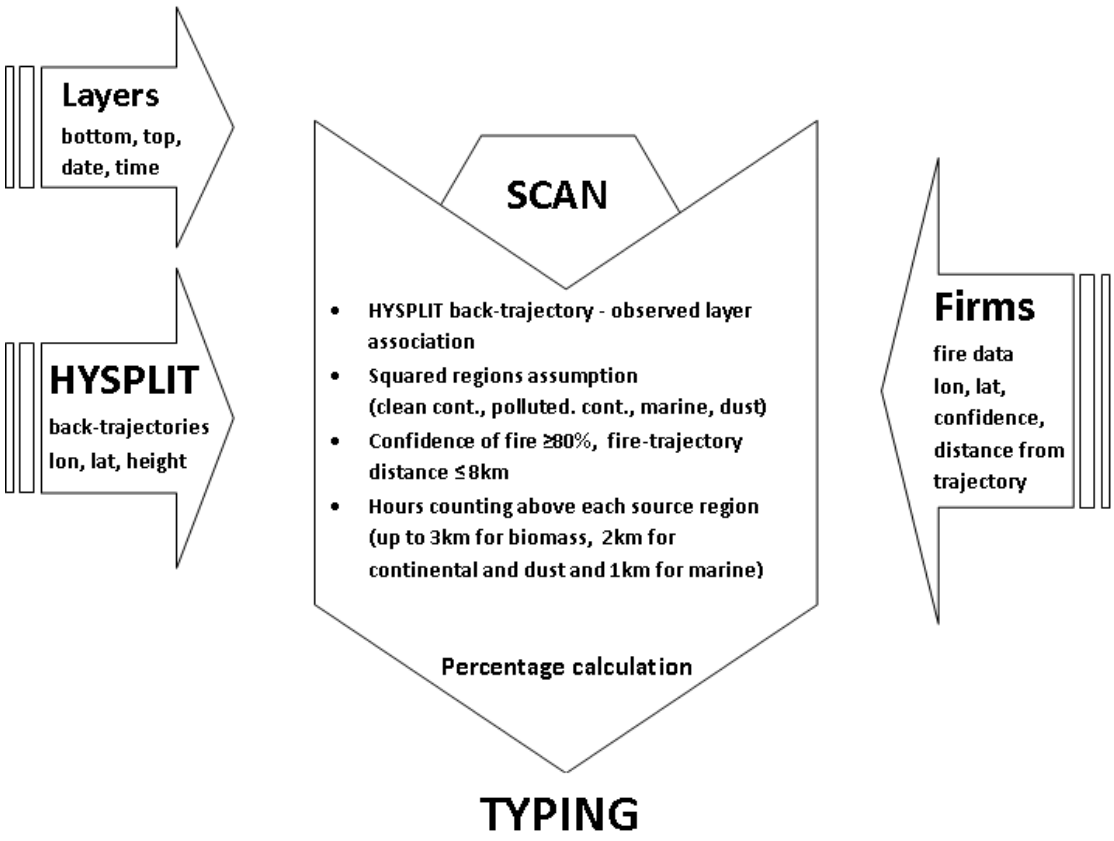

Figure 3.5 *Schematics of the SCAN algorithm for aerosol typing.*

If these criteria are indulged into more than one region, a mixture of more than one aerosol type is assumed. Else, if they are not indulged to none of the regions, then the source of the observed layer is considered as unknown.

3.3.2. Neural network aerosol-typing algorithm, *NATALI*

NATALI considers six classes of pure aerosol: continental, continental polluted, dust, marine, smoke, and volcanic (cf. section 2.1) (Nicolae, et al., 2018). It uses an aerosol model to calculate the optical properties of the aforementioned pure aerosols which are generated by a single source (e.g. dust produced by the deserts, marine particles produces by the oceans). This 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 iteratively compute the intensive optical properties of each aerosol type. The chemical composition of each pure aerosol type was picked up from the OPAC (Optical Properties of Aerosols and Clouds) software package (Hess, et al., 1998). 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 rescaled to the usual lidar wavelengths (i.e. 355,

532, and 1064 nm) using an Ångström exponent equal to 1. The mixtures were obtained by linear combination of pure aerosol properties.

Table 3.4 ***Correspondence between the aerosol types defined in the algorithm, as they can be retrieved by NATALI in high resolution and low resolution (Nicolae, et al., 2018).***

Aerosol types	High-resolution type	Low-resolution typing with LPDR	Low-resolution typing without LPDR
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 and dust	Continental dust	Continental or dust	Continental or dust
Dust and marine	Marine mineral	Dust or marine	Dust or marine
Volcanic and marine	Marine mineral	Durst or marine	Dust or marine
Continental and smoke	Continental smoke	Continental polluted or smoke	Continental polluted or smoke
Dust and smoke	Dust polluted	Dust or smoke	Dust or smoke
Continental and marine	Coastal	Continental or marine	Continental or marine
Continental polluted and marine	Coastal polluted	Continental polluted or marine	Continental polluted or marine
Continental and dust and marine	Mixed dust	Continental or dust	Continental or dust
Continental and smoke and marine	Mixed smoke	Continental polluted or smoke	Continental polluted or smoke

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 (β_{1064}, β_{532}, β_{355}, α_{532}, α_{355}, and optionally δ_{532} nm), identifies the layer geometrical boundaries and calculates within each layer the mean intensive optical parameters (i.e. Ångström exponent, colour indexes colour ratios, lidar ratios, particle linear depolarization ratio) and their associated uncertainty).

The input 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.

Three classification schemes are used with different aerosol type (classification) resolutions. First, when particle depolarization is available and all optical parameters are provided with a high-quality (uncertainty of the aerosol extinction coefficient $\leq$ 50 %, uncertainty of the aerosol backscatter coefficient $\leq$ 20 %, uncertainty of the particle linear depolarization ration $\delta \leq$ 30 %), the typing is performed in high resolution (AH) mode. This means that

the mixtures can be resolved and the number of outputs types is 14 (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) mode. In this case, the number of outputs types is six (i.e. pure with maximum 30% traces of other types).

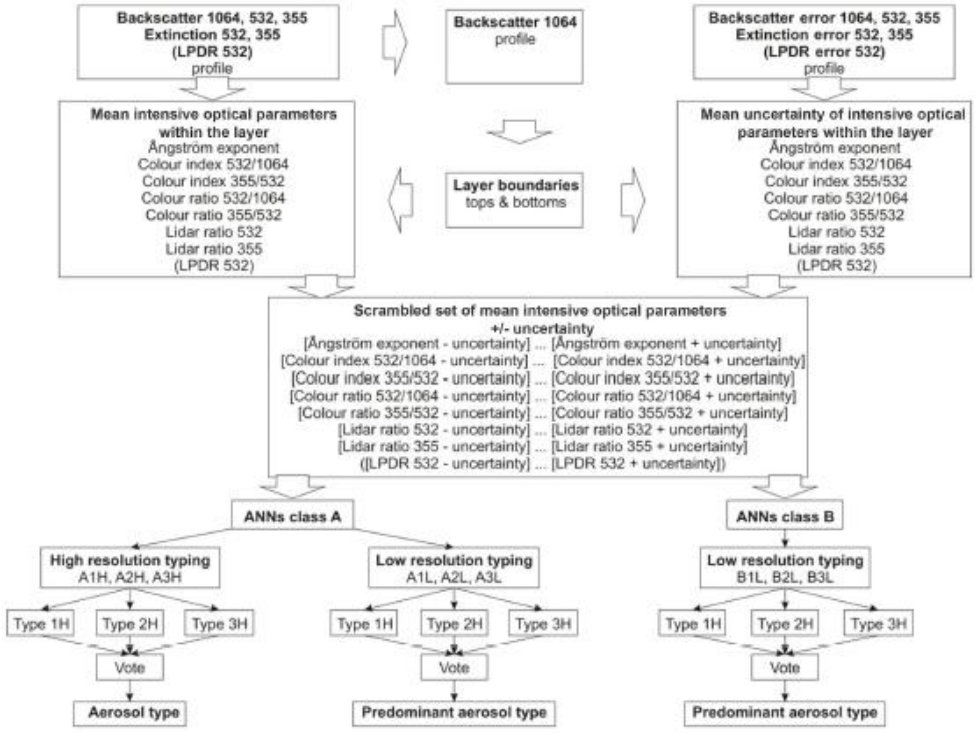

Figure 3.6 ***Schematics of the NATALI algorithm for aerosol typing (Nicolae, et al., 2018).***

Third, when the particle depolarization is not available, the typing is performed in low resolution mode, again meaning that the aerosol mixtures cannot be resolved. In this case, the predominant aerosol type is retrieved for four 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 pollution and are therefore excluded as output type.

A voting procedure selects the most probable answer out of the three (possibly different) individual returns. 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 (i.e. the percentage of agreement for values between error limits).

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 layer boundaries are moved towards the median height until the SNR criteria (<5) is met; if the criteria cannot be satisfied with a layer height greater than 300 m, the layer is discarded.

The netcdf files from EARLINET database containing the optical properties (b_{355}, b_{532}, b_{1064}, e_{355}, e_{532}, d_{532}) useful for this study were used as inputs to the Natali classification algorithm. For each case, the 'finesse' variable was adapted accordingly in order to be achieved the best match of the geometrical properties of each layer compared to the ones calculated manually.

3.3.3. Automatic observation-based aerosol typing method, *Mahalanobis* distance aerosol classification algorithm

Distance-based classification methods aim to assign an observation to a particular class based on the distance of the observation from each class center. In general, the Mahalanobis distance between an observation $\boldsymbol{x}=(x_1,\ldots,x_p)^t$ and the mean class $\overline{\boldsymbol{x}}=(\overline{x_1},\ \ldots,\ \overline{x_p})^t$ in the p-dimensional space $\mathbb{R}^p$ is defined as

$$\boldsymbol{D_M(x,\overline{x})} = \sqrt{(\boldsymbol{x}-\overline{\boldsymbol{x}})^T\boldsymbol{S}^{-1}(\boldsymbol{x}-\overline{\boldsymbol{x}})} \quad \mathbf{5}$$

where $\boldsymbol{S}$ is the class covariance matrix. The surfaces identified by the equation D_M = const. are ellipsoids that are centered around the mean $\overline{\boldsymbol{x}}$ **(**McLachlan, 1999)**.**

The Mahalanobis distance of an observation from an aerosol class can be estimated, and assigned to the aerosol class for which the distance is minimum. Two screening criteria are applied to the minimum distance following the procedure of Burton et al. (2012). The methodology uses 3 and 4 classifying parameters and the minimum accepted distance for a measurement to be labeled is 4 and 4.3, respectively. Moreover, the normalized probability of the aerosol class needs to be higher than 50 %. Otherwise, the type assignment is difficult as the measurement can be equidistant from 2 or more aerosol type classes, and possibly indicate the mixing of these aerosol types (Papagiannopoulos, et al., 2018).

EARLINET aerosol classified layers from Pappalardo et al. (2013), Papagiannopoulos et al. (2016), and Schwarz et al. (2016) were used as dataset for the training phase of the Mahalanobis algorithm. EARLINET observations from 2008 to 2010 present the backbone of the reference dataset. 8 aerosol

classes are used: clean continental (CC), continental polluted (CP), pure dust (D), mixed dust (MD = dust + marine), dust polluted (DP = dust + smoke and/or dust + polluted continental), mixed marine (MM), smoke (S), and volcanic (V).

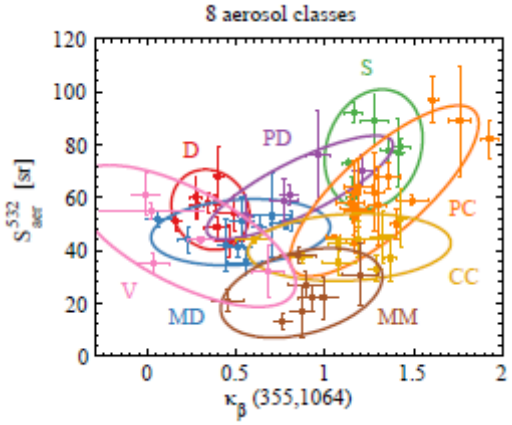

Figure 3.7 ***Colored pre-specified classes and 90 % confidence ellipses for 8 and 4 aerosol classes. The error bars correspond to the standard deviation of the selected mean intensive properties. CC stands for clean continental, D stands for dust, MD stands for mixed dust, MM stands for mixed marine, PD stands for polluted dust, PC stands for polluted continental, S stands for smoke, and V stands for volcanic particles (Papagiannopoulos, et al., 2018).***

Table 3.5 shows the characteristics of the reference dataset in terms of the S^{532}_{aer} and κ_β(355,1064) for the 8 classes, already mentioned. The coloring corresponds to the various classes and the crosshairs indicate the standard deviation of each of the aerosol layers. The 90 % confidence ellipses are calculated using the eigenvalues and eigenvectors of the covariance matrix and define the region that contains 90 % of all the points that can be drawn from the underlying normal class distribution.

Calibrated particle linear depolarization ratio profiles were not available in the selected dataset. So, Papagiannopoulos et al. (2018) used general literature values for particle linear depolarization ratio at 532 nm (Table 3.6) in order to train the algorithm.

Table 3.5 ***Reference dataset: mean type-dependent intensive properties along with the standard deviation (Papagiannopoulos, et al., 2018).***

Type	κ_β(355,1064)	κ_β(532,1064)	κ_β(355,532)	κ_a(355,532)	S^{355}_{aer}	S^{532}_{aer}
CC	1.0 ± 0.2	1.0 ± 0.3	1.3 ± 0.3	1.7 ± 0.6	50 ± 8	41 ± 6
CP	1.3 ± 0.3	1.3 ± 0.2	1.4 ± 0.6	1.7 ± 0.5	69± 12	63 ± 13
D	0.4 ± 0.1	0.4 ± 0.1	0.3 ± 0.2	0.3 ± 0.4	58 ± 12	55 ± 7
MD	0.5 ± 0.2	0.4 ± 0.3	0.7 ± 0.3	0.5 ± 0.3	42 ± 4	47 ± 6
DP	0.9 ± 0.3	0.8 ± 0.1	1.0 ± 0.5	0.6 ± 0.2	54 ± 8	64 ± 9

MM	0.8 ± 0.1	0.8 ± 0.2	1.0 ± 0.3	0.9 ± 0.3	25 ± 7	24 ± 8
S	1.3 ± 0.1	1.3 ± 0.1	1.2 ± 0.3	1.3 ± 0.3	81 ± 16	78 ± 11
V	0.1 ± 0.1	0.4 ± 0.3	0.2 ± 0.3	0.2 ± 0.3	50 ± 11	48 ± 13

Table 3.6 ***The mean and standard deviation of the particle depolarization ratio used for the pre-specified classes and the corresponding bibliographic references (Papagiannopoulos, et al., 2018).***

Type	δ_{aer}^{532}	References
Clean continental	0.04 ± 0.02	Burton et al. (2013)
Continental polluted	0.05 ± 0.03	Burton et al. (2013)
Dust	0.30 ± 0.01	Gross et al. (2011)
Mixed dust	0.15 ± 0.02	Gross et al. (2016)
Dust polluted	0.20 ± 0.05	Burton et al. (2013)
Marine	0.03 ± 0.01	Gross et al. (2013)
Smoke	0.10 ± 0.04	Burton et al. (2013)
Volcanic	0.33 ± 0.03	Pappalardo et al. (2013)

As indicated by Papagiannopoulos et al. (2018), of the most importance in the aerosol typing classification has the set of parameters: κ_β(355,1064), S_{aer}^{532}, and $S_{aer}^{532}/S_{aer}^{355}$. The decision for the selected parameters stems solely from the lowest arithmetic value of the total Wilks' lambda, Λ (Wilks, 1963). To this set, the particle linear depolarization ratio at 532 nm was added.

The mean values of the optical properties (b_{355}, b_{532}, b_{1064}, e_{355}, e_{532}, d_{532}) for each layer calculated both manually and by Natali were used as inputs for the Mahalanobis distance automatic aerosol classification algorithm.

3.4. Aerosol categories

Four different categories were created depending on the results from SCAN: "pure", "mixture of 2", "mixture of 3" and "no type". The "pure" category consists of the cases that SCAN counted only 1 aerosol type during their back-trajectory. The "mixture of 2" consists of the cases that SCAN counted 2 aerosol types during their back-trajectory. The "mixture of 3" consists of the cases that SCAN counted 3 aerosol types during their back-trajectory. Finally, "no type" consists of the cases that SCAN was unable to identify the potential source of the observed layers. Each of the categories mentioned above, was, then, divided into subcategories depending on the typing results of SCAN. The "pure" category was divided into 5 subcategories "Continental Polluted" (cp), "Clean Continental" (cc), "Smoke" (bb), "Dust" (d) and "Marine" (m). The "mixture of 2" category was divided into 5 subcategories, "Continental

Polluted and Smoke", "Continental Polluted and Clean Continental" (cp+cc), "Continental Polluted and Marine" (cp+m), "Clean Continental and Marine" (cc+m) and "Continental Polluted and Dust" (cp+d). Finally, the "mixture of 3" category was divided into 3 subcategories "Continental Polluted, Clean Continental and Marine" (cp+cc+m), "Continental Polluted, Smoke and Marine" (cp+bb+m) and "Continental Polluted, Smoke and Dust" (cp+bb+d).

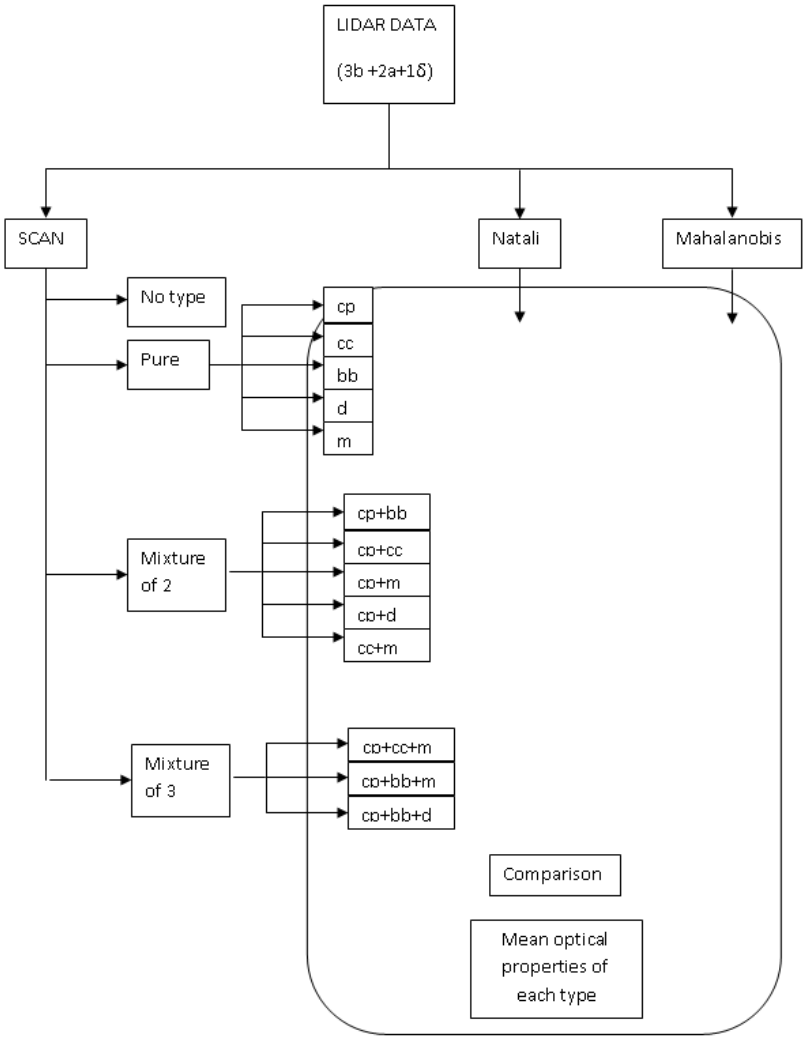

Figure 3.8 ***Schematics of the methodology of this study.***

The shorthands of Natali's and Mahalanobis's results were used so that the comparison between the 3 different methods be feasible (Table 3.7).

Table 3.7 ***Correspondence between the aerosol types and shorthand, as they can be retrieved by Natali, Mahalanobis and SCAN.***

Aerosol types	Natali	Mahalanobis	SCAN
Continental (cc)	Continental	Continental	Clean Continental
Continental Polluted (cp)	Continental polluted	Continental polluted	Continental polluted
Smoke (bb)	Smoke	Smoke	Smoke

Dust (d)	Dust	Dust	Dust
Marine (m)	Marine	Marine	Marine
Continental and dust (cp+d)	Continental dust	Dust polluted	Continental and dust
Dust and marine (d+m)	Marine mineral	Mixed dust	Dust and marine
Continental and smoke (cp+bb)	Continental smoke	-	Continental polluted and smoke
Dust and smoke (d+bb)	Dust polluted	Dust polluted	Dust and smoke
Continental and marine (cc+m)	Coastal	-	Clean continental and marine
Continental polluted and marine (cp+m)	Coastal polluted	-	Continental polluted and marine
Continental polluted and clean continental (cp+cc)	-	-	Continental polluted and clean continental
Continental and dust and marine (cc+d+m)	Mixed dust	-	Clean continental and dust and marine
Continental and smoke and marine (cc+bb+m)	Mixed smoke	-	Clean continental and smoke and marine
Continental polluted and smoke and marine (cp+bb+m)	-	-	Continental polluted and smoke and marine
Continental and smoke and dust (cp+bb+d)	-	-	Continental and smoke and dust
Continental and clean continental and marine (cp+cc+m)	-	-	Continental and clean continental and marine

The performance of Mahalanobis and Natali algorithms was studied for each of the above categories.

4. Results

The results of the 3 different classification algorithms can be found at the Table 4.1. The date (dd/mm-hh:mm), bottom (m) and top (m) of each layer is also given.

Table 4.1 ***Classification results from SCAN, Natali and Mahalanobis, for each date and layer.***

Year (yyyy)	Date (dd/mm-hh:mm UTC)	Bottom (m)	Top (m)	SCAN classification	Natali classification	Mahalanobis classification
2015	22/01-18:05	3744	4402	no type	cc	cc
	13/04-19:45	1330	1570	cp	cp	dp
		2410	2650	cp	cc	cc
	07/05-20:45	2890	3190	no type	m/cc	cp
		3910	4570	no type	bb	no type
	04/06-19:17	1570	1810	cp	cp	cp
		2710	3130	cp	cc	cc
		3910	4590	cp	cp	cp
	11/06-23:35	1270	1570	cp+m	bb	no type
		1690	1990	cp	cp	cp
		2350	2710	no type	cp	cp
		3370	3730	no type	cc	cc
	16/07-21:00	1810	3130	bb	cc	cp+m/cc
	20/07-23:59	1510	3250	cp+bb+m	cc	cc
	30/07-19:25	1324	2041	cp+cc+m	cp	cp
	21:26	2830	3130	cp+d	m/cc	cp
		3490	3790	cp	m/cc	m
		4570	4930	cp+m	d+m	no type
		5710	6010	cp+m	cc+d+m/cc	m
	24/08-18:00	1204	1802	cc	cp	cp
	03/09-20:16	1930	2710	no type	cc+d+m/cc	no type
2016	04/01-15:18	2399	3296	no type	cp+m/cc	no type
	07/01-15:19	965	1145	cp+cc	cp+m/cc	no type
		1982	2191	cp+cc	cp	cp
	21/03-18:21	1115	1294	cc+m	cp	cc
	28/03-23:11	4701	4970	no type	d	no type
	04/04-18:47	2290	3310	d	cc+d+m/cc	d+m
	11/04-20:27	2050	2530	no type	cc+d+m/cc	d+m
	14/04-21:00	1742	1922	cc	bb	no type
	23/06-19:30	1930	3430	cp	m/cc	cp
		4150	5110	d	no type	d+bb/cp
	27/06-21:07	1570	1870	cp+cc+m	cc+bb+m/cc	no type
		2530	2890	no type	m/cc	no type
	04/07-19:56	1750	1930	cp+bb	cp+m/cc	cc
		2470	2710	no type	cc	cc
		2950	3310	no type	m/cc	cc
	07/07-19:01	1210	1510	cp	cp	cp
	11/07-19:22	1090	2830	cp+d	cc	cc
		3010	4630	no type	cp+m/cc	no type
	25/08-18:29	1030	1870	cp+bb	cc	cp
		2350	3850	cp	cp+m/cc	cc
	29/08-19:16	1930	2590	bb	cc	cc
2017	23/03-18:10	1990	2170	cp+d	cp	no type
	30/03-18:08	1270	1450	cp+bb	no type	no type
		1990	2350	cp+bb	cc	cc
	27/04-19:48	3250	4210	no type	cc+d+m/cc	d+m
	22/05-19:01	1090	2170	cp+bb	cp	no type
	08/06-19:14	1390	1750	cp	cp+m/cc	cc
		2290	2650	m	cp+m/cc	no type
	12/06-19:29	1630	1990	cp	cp+m/cc	cc

	15/06-20:00	1090	1930	cp+bb	cc	no type
		2350	2650	cp+bb	m/cc	no type
		2830	3250	d	m/cc	cc
		3850	4270	d	m/cc	cc
	22/06-19:18	1030	1570	cp+bb	cp+bb/cc	no type
		2110	2590	no type	cp+bb/cc	no type
		3130	3970	no type	m/cc	no type
		4150	5230	no type	cc	cc
	13/07-20:15	1810	3130	no type	cc	cc
	19/07-18:57	1030	1450	cp+bb	cp	cp
		1990	2530	bb	cp	cc
	20/07-20:15	1030	1450	cp	cp+bb	no type
		1870	2830	bb	cp+m/cc	cc
		3490	4510	no type	cc+d+m	no type
		4990	6010	no type	cp+m	no type
	24/08-16:56	1068	1368	cp	d+m	no type
		1968	2418	cp+bb	cc	no type
	17:59	1668	2268	cp+bb	cc	no type
	28/08-18:45	1750	1990	cp+bb	cc	cc
		3250	3430	no type	cc+d+m/cc	d+m
		4330	4570	no type	no type	no type
		4810	5050	no type	cc+bb+m	no type
		5710	5890	no type	cp+m/cc	cc
	31/08-17:59	1218	1668	cp+bb	cc	no type
		2118	2418	bb	cc	cc
	18:26	1510	1750	cp+bb	cc	no type
		2470	2950	cp+bb+d	cc	cc
	07/09-16:33	1968	2268	cp	cc	cc
	21/09-16:16	1968	2268	cp+bb	cp	no type
	19/10-17:14	1930	2110	cp	cc+bb+m/cc	no type
	21/06-20:30	1807	2375	cp+cc+m	m/cc	m
	19/07-21:30	1710	2330	cp	cc	cc
		2988	3571	no type	no type	no type
		4011	4512	no type	cc	no type
	26/07-20:00	1112	1352	cp	cp	cp
		1434	2398	cp	cp	cp
	30/07-21:45	1112	1516	cp+cc	cp	cp
		1904	2667	no type	cp	cc
2018		3070	3339	no type	cp+m/cc	cc
	02/08-23:00	1120	2181	cp	cp	cp
		2681	3212	no type	cp	cp
	16/08-21:20	1613	2442	cp	cc	cc
	27/08-19:00	1112	1673	cp	cc	cp
		1934	2592	no type	cp	no type
		3107	4310	cc	cp+bb	no type
		4557	5341	no type	cp+bb	no type
		6111	6313	no type	cp+bb	no type

The comparison between these three classification algorithms are shown in the figures below (Fig. 4.1, Fig. 4.2, Fig. 4.3). The axes demonstrate the aerosol types introduced in Table 3.8. The numbers in the cells indicate the number of the aerosol layers classified by each aerosol classification algorithm.

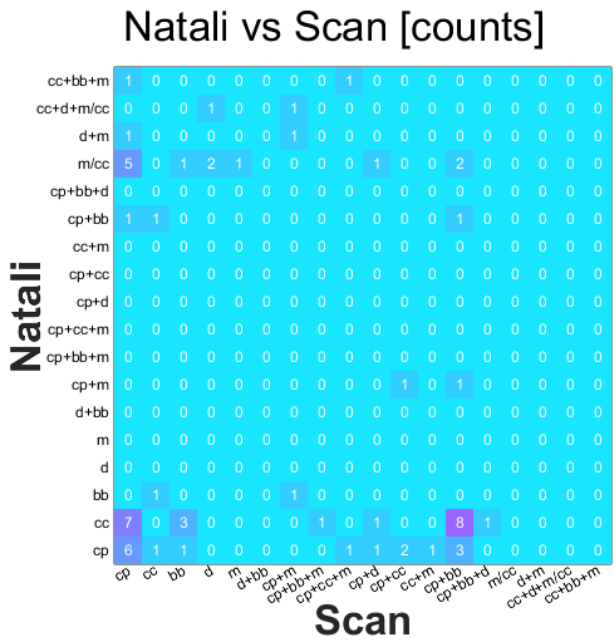

Figure 4.1 ***Heatmap of Scan's and Natali's results comparison. The numbers in the cells indicate the number of layers classified by NATALI (row label) and SCAN (column label).***

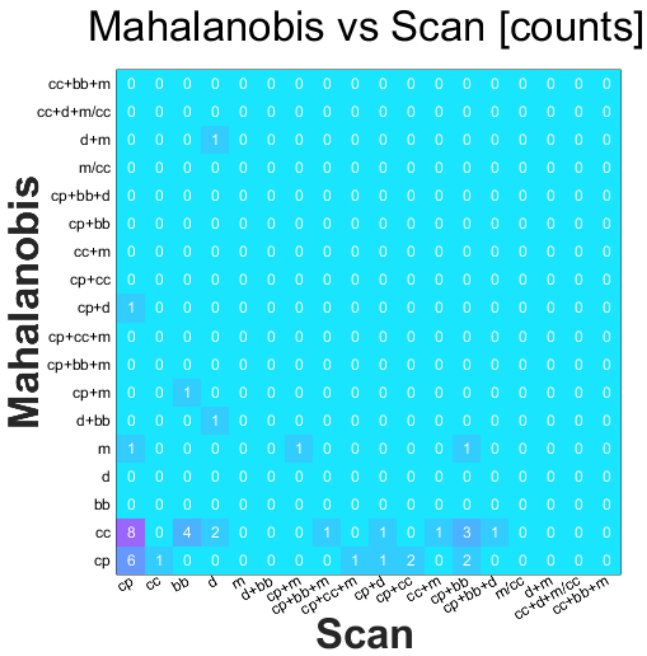

Figure 4.2 ***Heatmap of Scan's and Mahalanobis's results comparison. The numbers in the cells indicate the number of layers classified by MAHALANOBIS (row label) and SCAN (column label).***

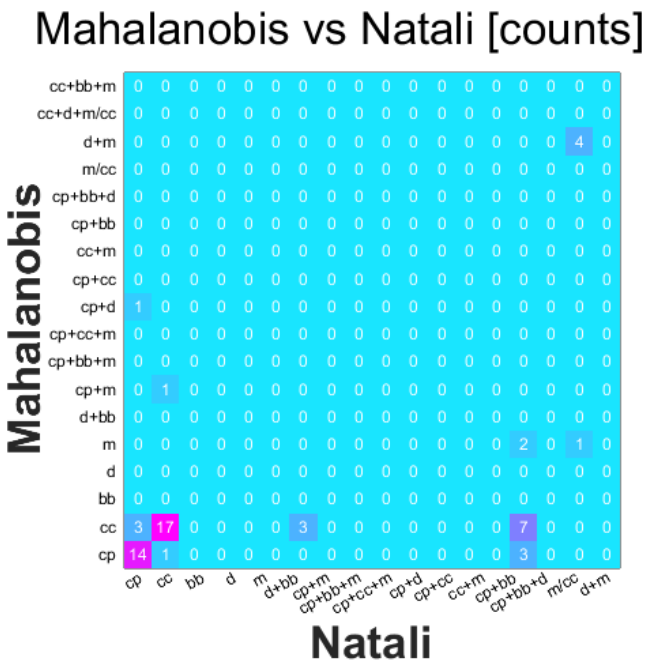

Figure 4.3 ***Heatmap of Natali's and Mahalanobis's results comparison. The numbers in the cells indicate the number of layers classified by MAHALANOBIS (row label) and NATALI (column label).***

4.1. Pure types

To further investigate the discrepancies between the three aerosol classification algorithms, SCAN's results are classified into 4 major categories: "pure" (36%, 35 cases, yellow slice of the pie, Fig. 4.4), "mixture of 2" (26%, 25 cases, cyan slice of the pie, Fig. 4.4), "mixture of 3" (5%, 5 cases, magenta slice of the pie, Fig. 4.4) and "no type" (33%, 32 cases, grey slice of the pie, Fig. 4.4).

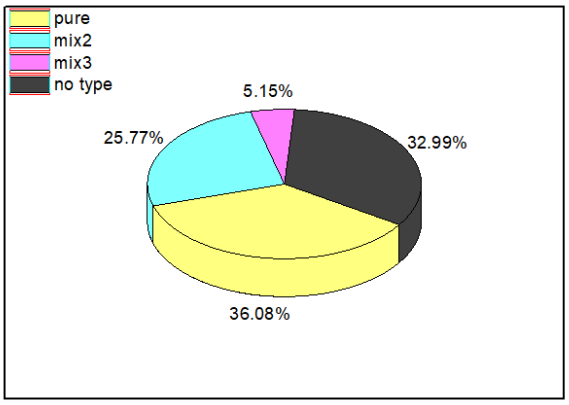

Figure 4.4 ***Major categories of SCAN's results.***

Each of the above categories is divided into minor subcategories depending on the typing results of SCAN.

The "pure" major category is divided into 5 minor subcategories: "Continental Polluted", "Clean Continental", "Smoke", "Marine" and "Dust".

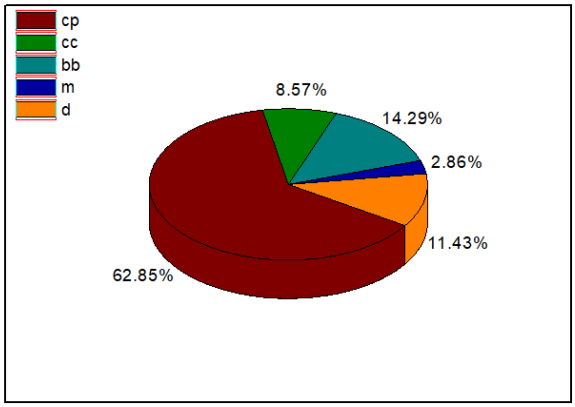

Figure 4.5 ***Minor categories of SCAN's results for major category "pure".***

The "Continental Polluted" category (dark red slice of the pie, Fig. 4.5) represents the 63% (22 cases) of the pure aerosol types detected by SCAN. The "Smoke" category (dark grey slice of the pie, Fig. 4.5) represents the 14% (5 cases) of pure aerosol types detected by SCAN. The "Dust" category (orange slice of the pie, Fig. 4.5) represents the 11% (4 cases) of the pure aerosol types detected by SCAN. The "Clean Continental" category (green slice of the pie, Fig. 4.5) represents the 8% of the pure aerosol types detected by SCAN. Finally, the "Marine" category (blue slice of the pie, Fig.4.5) represents the 3% (1 case) of the pure aerosol types detected by SCAN.

4.1.1. Continental Polluted

The Mahalanobis algorithm classified as "continental polluted" the 36% of the SCAN "continental polluted" cases (Fig. 4.6, left part of the figure, black column). At the same time, Mahalanobis classified as "clean continental" the 40% of the same cases (Fig. 4.6, left part of the figure, blue column), while it couldn't classify the 14% of the SCAN's cp cases (Fig. 4.6, left part of the figure, orange column). The 5% of the SCAN's cp cases were classified by Mahalanobis as mixtures, specifically as "dust and smoke" (Fig. 4.6, left part of the figure, olive column). Finally, the 5% of the SCAN's cp cases were classified by Mahalanobis as "marine" (Fig. 4.6, left part of the figure, magenta column).

Natali algorithm classified as "continental polluted" the 36% (Fig. 4.6, right part of the figure, black column) of the same cases mentioned above ("continental polluted"), and as "clean continental" the 27% of them (Fig. 4.6, right part of the figure, blue column). The 27% of the SCAN's cp cases were classified by Natali as mixtures, specifically as "dust and marine" (Fig. 4.6, right part of the figure, medium blue column), as "continental polluted and marine" (Fig. 4.6, right part of the figure, brown column), as "continental polluted and smoke" (Fig. 4.6, right part of the figure, pink column) and "clean

continental, smoke and marine" (Fig. 4.6, right part of the figure, green column). Finally, the 9% of the SCAN's cp cases were classified by Natali as "marine or continental" (Fig. 4.6, right part of the figure, dark blue column).

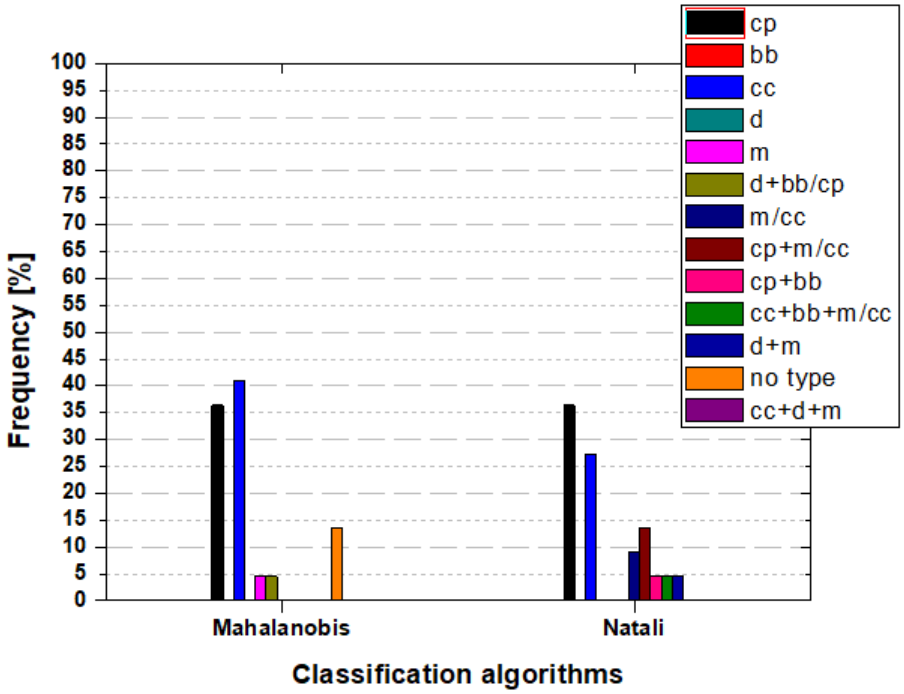

Figure 4.6 *Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted".*

Natali and Mahalanobis have a 59% agreement in their results. Simultaneously, both Mahalanobis and Natali result with a high frequency of "clean continental" aerosol type for the layers classified as "continental polluted" by SCAN. Finally, the 100% of Mahalanobis "no type" results (concerning the cp cases by SCAN), classified by Natali as a mixture.

In Table 4.2 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted" by SCAN for this study.

Table 4.2 *Mean values of intensive optical properties of the "continental polluted" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	1.30 ± 0.25
CI355/532	1.29 ± 0.24
CI532/1064	0.99 ± 0.14
CR355/532	1.70 ± 0.17
CR532/1064	2.02 ± 0.21
LR355 [sr]	49 ± 5
LR532 [sr]	52 ± 5
LPDR532 [%]	3 ± 1

4.1.2. Clean Continental

The Mahalanobis algorithm classified as "continental polluted" the 33% of the SCAN "clean continental" cases (Fig. 4.7, left part of the figure, black

column). At the same time, couldn't classify the 67% of them (Fig. 4.7, left part of the figure, orange column).

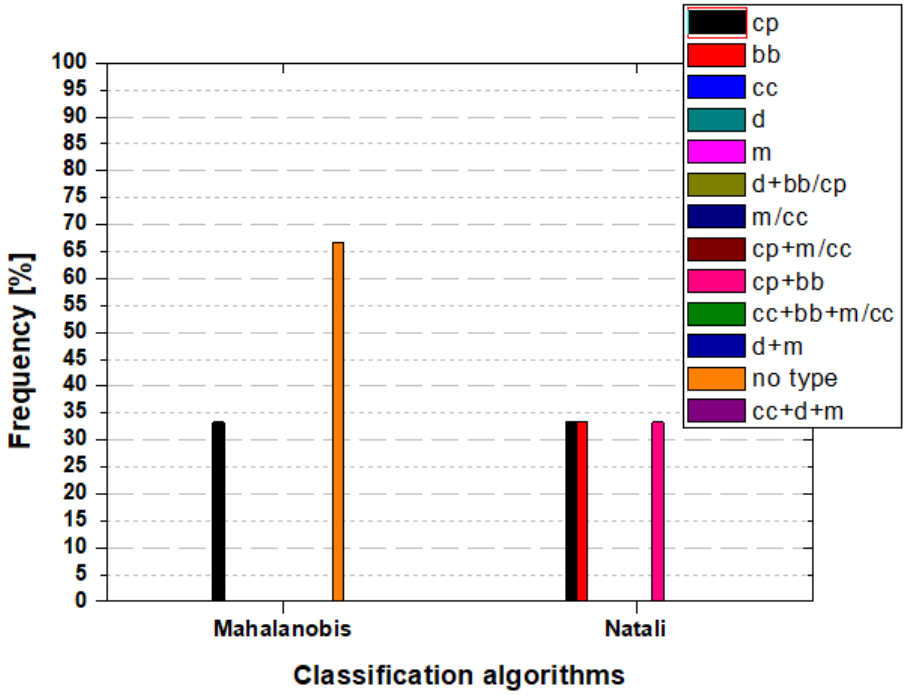

Figure 4.1 *Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Clean Continental".*

Natali algorithm classified as "continental polluted" the 33% (Fig. 4.7, right part of the figure, black column) of the same cases mentioned above ("clean continental") and as "smoke" the other 33% of them (Fig. 4.7, right part of the figure, red column). Finally, the last 33% of the SCAN's cc cases were classified by Natali as "continental polluted and smoke" (Fig. 4.7, right part of the figure, pink column).

Natali and Mahalanobis algorithm have 33% agreement in their results concerning the SCAN's cc type.

Mahalanobis has a high percentage of "no type" results (67%), 50% of them classified by Natali as a mixture.

Table 4.3 *Mean values of intensive optical properties of the "clean continental" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	1.06 ± 0.21
CI355/532	2.15 ± 0.18
CI532/1064	0.99 ± 0.10
CR355/532	2.43 ± 0.15
CR532/1064	1.98 ± 0.14
LR355 [sr]	50 ± 4
LR532 [sr]	73 ± 6
LPDR532 [%]	2 ± 1

In Table 4.3 we introduce the mean values of the intensive optical properties for the layers classified as "clean continental" by SCAN for this study.

4.1.3. Smoke

The Mahalanobis algorithm classified as "clean continental" the 100% of the SCAN "smoke" cases (Fig. 4.8, left part of the figure, blue column).

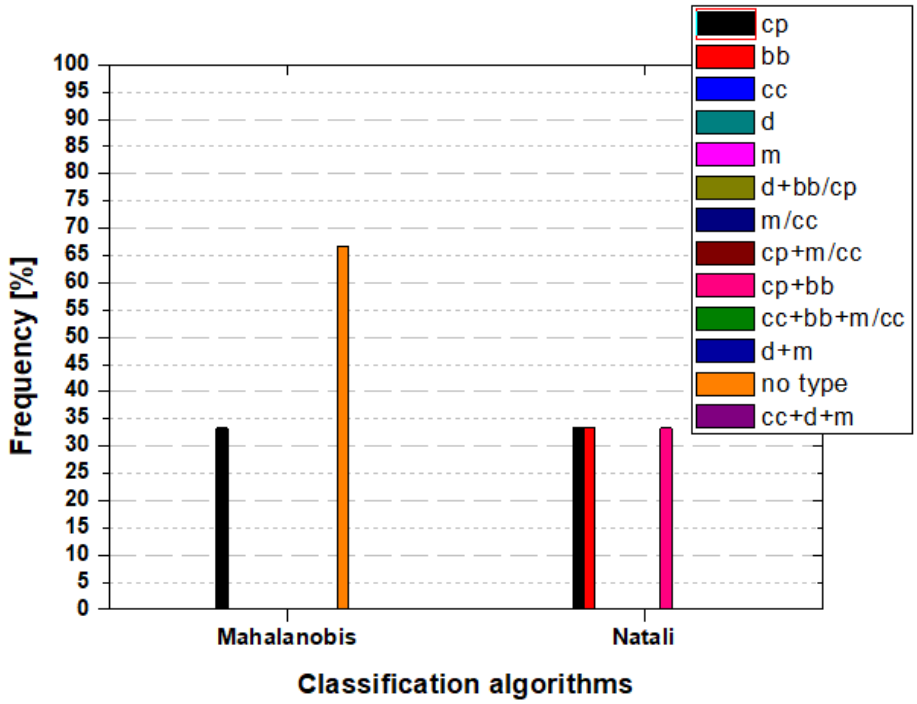

Figure 4.8 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Clean Continental".***

Natali algorithm classified as "continental polluted" the 20% (Fig. 4.8, right part of the figure, black column) of the "smoke" cases of SCAN and as "clean continental" the other 40% (Fig. 4.8, right part of the figure, blue column). Finally, the last 40% of the SCAN's bb cases were classified by Natali as "continental polluted and marine" (Fig. 4.8, right part of the figure, brown column).

We can observe that Natali and Mahalanobis have 40% agreement in their results concerning the SCAN's bb type.

Moreover, Mahalanobis has a high percentage of "clean continental" results (100%).

In Table 4.4 we introduce the mean values of the intensive optical properties for the layers classified as "smoke" by SCAN for this study.

Table 4.4 ***Mean values of intensive optical properties for the "smoke" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	1.55 ± 0.30

CI355/532	0.92 ± 0.25
CI532/1064	1.33 ± 0.15
CR355/532	1.47 ± 0.15
CR532/1064	2.58 ± 0.29
LR355 [sr]	46 ± 8
LR532 [sr]	37 ± 4
LPDR532 [%]	3 ± 1

4.1.4. Dust

The Mahalanobis algorithm classified as "clean continental" the 50% of the SCAN dust cases (Fig. 4.9, left part of the figure, blue column). At the same time, Mahalanobis classified as "dust and smoke" or "clean continental" the 25% of the same cases (Fig. 4.9, left part of the figure, olive column), while the last 25% classified it as "dust and marine" (Fig. 4.9, left part of the figure, medium blue column).

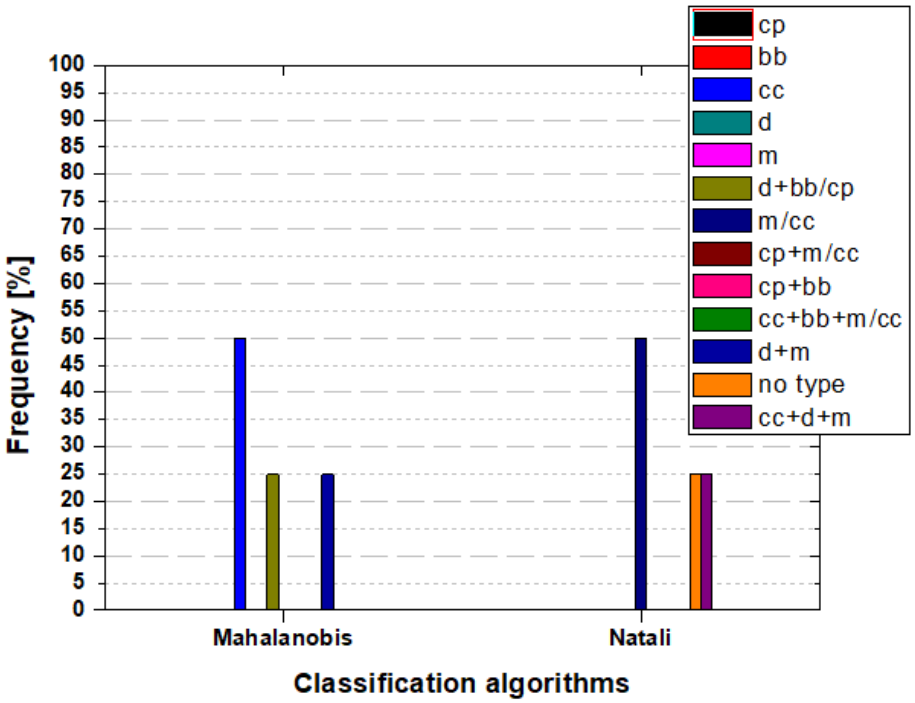

Figure 4.2 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "dust".***

For the same cases, Natali algorithm classified as "marine" or "clean continental" the 50% (Fig. 4.9, right part of the figure, dark blue column) and as "clean continental, dust and marine" the other 25% (Fig. 4.9, right part of the figure, purple column). Finally, the last 25% of the SCAN's "dust" cases Natali was not able to classify it (Fig. 4.9, right part of the figure, orange column).

Natali and Mahalanobis algorithm have 25% agreement at their results concerning the SCAN's d type.

Moreover, Mahalanobis has a relatively high percentage of "clean continental" results (50%).

In Table 4.5 we introduce the mean values of the intensive optical properties for the layers classified as "dust" by SCAN for this study.

Table 4.5 *Mean values of intensive optical properties for the "dust" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	0.94 ± 0.22
CI355/532	0.43 ± 0.22
CI532/1064	0.82 ± 0.13
CR355/532	1.25 ± 0.12
CR532/1064	1.78 ± 0.16
LR355 [sr]	39 ± 3
LR532 [sr]	32 ± 3
LPDR532 [%]	11 ± 3

4.1.5. Marine

The Mahalanobis algorithm couldn't classify the 100% of the SCAN "marine" cases (Fig. 4.10, left part of the figure, orange column), while Natali classified the 100% of the SCAN "marine" cases as "continental polluted and marine" or "clean continental" (Fig. 4.10, right part of the figure, brown column).

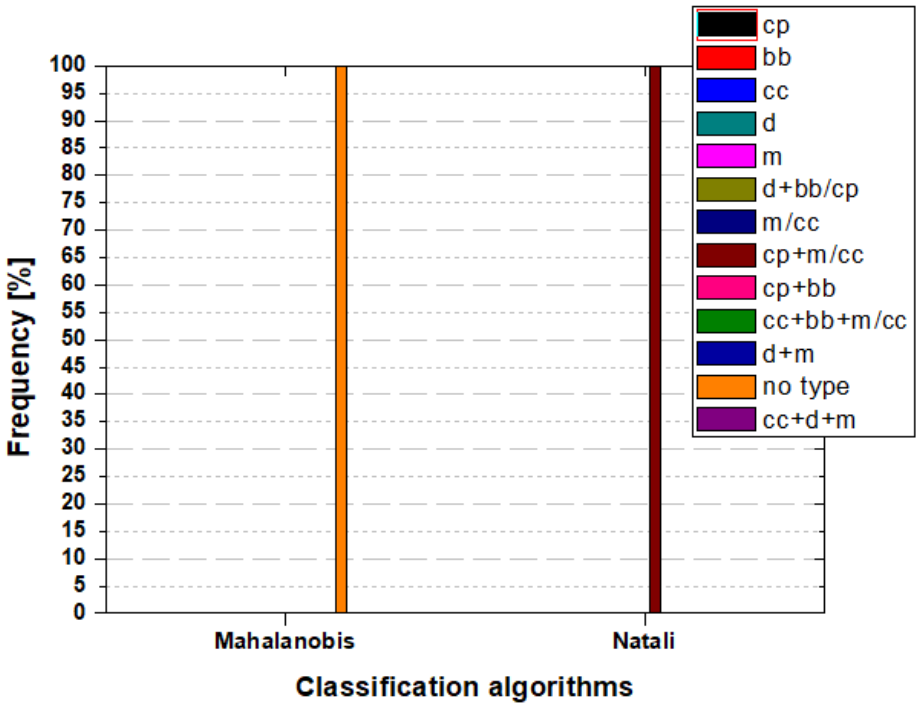

Figure 4.3 *Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "marine".*

We should mention here that there is only one case classified by SCAN as a pure "marine" type.

In Table 4.6 we introduce the mean values of the intensive optical properties for the layers classified as "marine" by SCAN for this study.

Table 4.6 *Mean values of intensive optical properties for the "marine" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	0.42 ± 0.29
CI355/532	1.14 ± 0.29
CI532/1064	0.88 ± 0.17
CR355/532	1.59 ± 0.18
CR532/1064	1.85 ± 0.21
LR355 [sr]	29 ± 3
LR532 [sr]	39 ± 5
LPDR532 [%]	1 ± 1

4.2. Mixtures of 2 aerosol types

The "mixture of 2" major category is divided into 5 minor subcategories: "Continental Polluted and Smoke", "Continental Polluted and Dust", "Continental Polluted and Clean Continental", "Continental Polluted and Marine" and "Clean Continental and Marine".

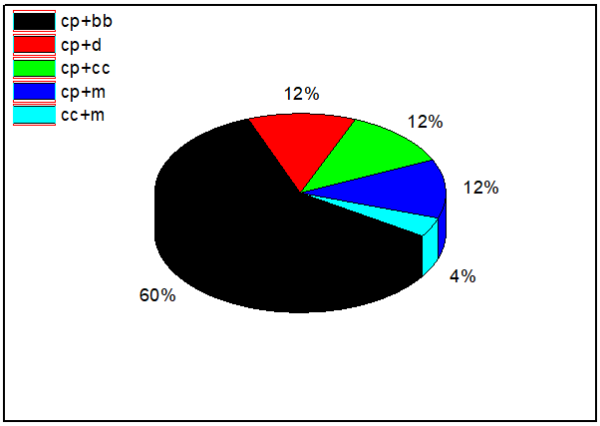

Figure 4.4 *Minor categories of SCAN's results for major category "mixture of 2".*

The "Continental Polluted and Smoke" category (black slice of the pie, Fig. 4.11) represents the 60% (15 cases) of the mixture of 2 aerosol types detected by SCAN. The "Continental Polluted and Dust" category (red slice of the pie, Fig. 4.11) represents the 12% (3 cases) of the mixture of 2 aerosol types detected by SCAN. The "Continental Polluted and Clean Continental" category (green slice of the pie, Fig. 4.11) represents the 12% (3 cases) of the mixture of 2 aerosol types detected by SCAN. The "Continental Polluted and Marine" category (blue slice of the pie, Fig. 4.11) represents the 12% of the mixture of 2 aerosol types detected by SCAN. Finally, the "Clean Continental and Marine" category (cyan slice of the pie, Fig. 4.11) represents the 4% (1 case) of the mixture of 2 aerosol types detected by SCAN.

4.2.1. Continental Polluted and Smoke

The Mahalanobis algorithm classified as "continental polluted" the 13% of the SCAN "continental polluted and smoke" cases (Fig. 4.12, left part of the figure, black column). At the same time, Mahalanobis classified as "clean continental"

the 20% of the same cases (Fig. 4.12, left part of the figure, blue column), while it couldn't classify the 67% of the SCAN's cp+bb cases (Fig. 4.12, left part of the figure, orange column).

Natali algorithm classified as "continental polluted" the 33% (Fig. 4.12, right part of the figure, black column) of the same cases mentioned above ("continental polluted and smoke"), and as "clean continental" the 40% of them (Fig. 4.12, right part of the figure, blue column). The 13% of the SCAN's cp+bb cases were classified by Natali as mixtures, specifically as "continental polluted and marine or clean continental" (Fig. 4.12, right part of the figure, brown column). The 7% of the SCAN's cp+bb cases were classified by Natali as "marine or continental" (Fig. 4.12, right part of the figure, dark blue column) and the last 7% as "no type".

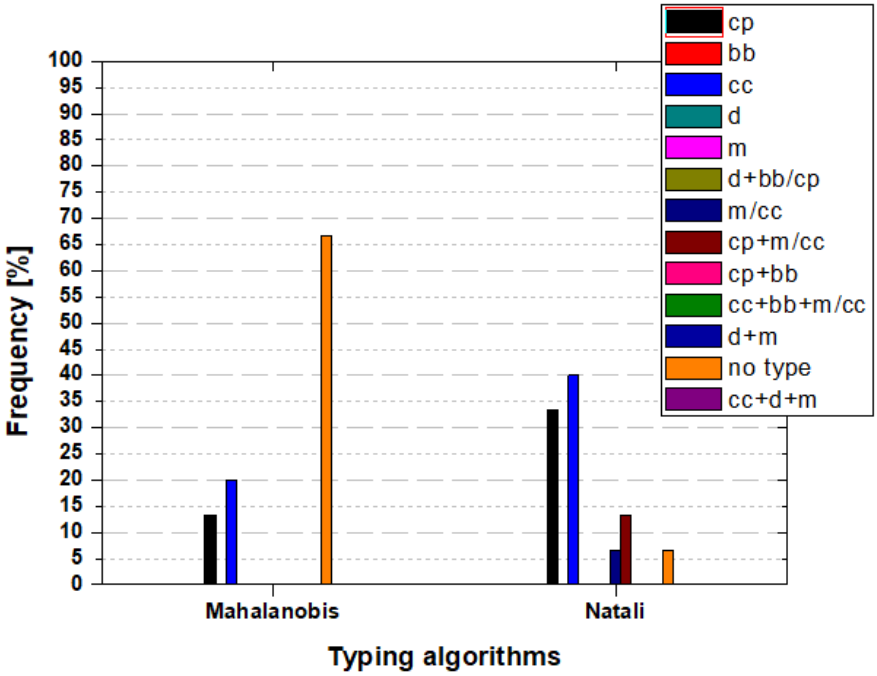

Figure 4.5 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted and Smoke".***

In Table 4.7 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted and smoke" by SCAN for this study.

Table 4.7 ***Mean values of intensive optical properties for the "continental polluted and smoke" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	1.63 ± 0.40
CI355/532	0.78 ± 0.35
CI532/1064	1.31 ± 0.21
CR355/532	1.41 ± 0.20
CR532/1064	2.67 ± 0.40
LR355 [sr]	52.73 ± 7.95

LR532 [sr]	45.87 ± 6.96
LPDR532 [%]	4.02 ± 1.94

4.2.2. Continental Polluted and Dust

The Mahalanobis algorithm classified as "continental polluted" the 33% of the SCAN "continental polluted and dust" cases (Fig. 4.13, left part of the figure, black column). At the same time, Mahalanobis classified as "clean continental" the 33% of the same cases (Fig. 4.13, left part of the figure, blue column), while it could not classify the 33% of the SCAN's cp+d cases (Fig. 4.13, left part of the figure, orange column).

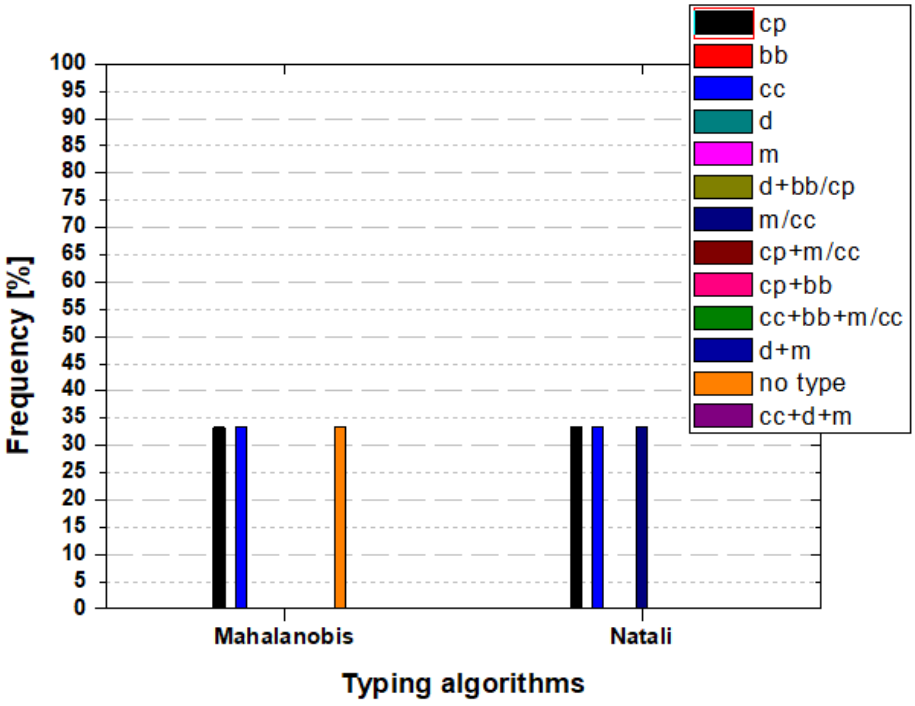

Figure 4.6 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted and Dust".***

Natali algorithm classified as "continental polluted" the 33% (Fig. 4.13, right part of the figure, black column) of the same cases mentioned above ("continental polluted and dust"), and as "clean continental" the 33% of them (Fig. 4.13, right part of the figure, blue column). The 33% of the SCAN's cp+d cases were classified by Natali as "marine or continental" (Fig. 4.13, right part of the figure, dark blue column).

In Table 4.8 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted and dust" by SCAN for this study.

Table 4.8 ***Mean values of intensive optical properties for the "continental polluted and dust" classified layers by SCAN.***

Intensive optical properties	Mean ± error

AE355/532	1.04 ± 0.28
CI355/532	0.83 ± 0.28
CI532/1064	0.93 ± 0.17
CR355/532	1.41 ± 0.17
CR532/1064	1.91 ± 0.22
LR355 [sr]	44.67 ± 5.07
LR532 [sr]	43.33 ± 5.13
LPDR532 [%]	4.33 ± 1.68

4.2.3. Continental Polluted and Clean Continental

The Mahalanobis algorithm classified as "continental polluted" the 33% of the SCAN "continental polluted and clean continental" cases (Fig. 4.14, left part of the figure, black column). At the same time, Mahalanobis classified as "no type" the 67% of the same cases (Fig. 4.14, left part of the figure, orange column).

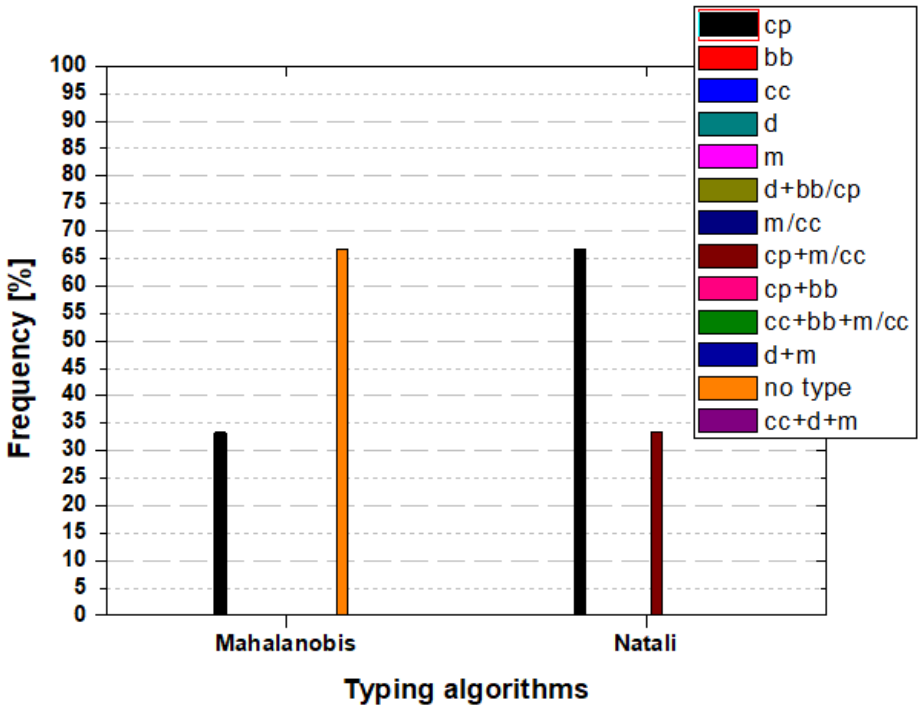

Figure 4.7 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted and Clean Continental".***

Natali algorithm classified as "continental polluted" the 67% (Fig. 4.14, right part of the figure, black column) of the same cases mentioned above ("continental polluted and clean continental"), and as "continental polluted and marine or clean continental" the 33% of them (Fig. 4.14, right part of the figure, brown column).

In Table 4.9 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted and clean continental" by SCAN for this study.

Table 4.9 ***Mean values of intensive optical properties for the "continental polluted and clean continental" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	1.04 ± 0.28
CI355/532	0.83 ± 0.28
CI532/1064	0.93 ± 0.17
CR355/532	1.41 ± 0.17
CR532/1064	1.91 ± 0.22
LR355 [sr]	44.67 ± 5.07
LR532 [sr]	43.33 ± 5.13
LPDR532 [%]	4.33 ± 1.68

4.2.4. Continental Polluted and Marine

The Mahalanobis algorithm classified as "marine" the 50% of the SCAN "continental polluted and marine" cases (Fig. 4.15, left part of the figure, magenta column). At the same time, Mahalanobis classified as "no type" the other 50% of the same cases (Fig. 4.15, left part of the figure, orange column).

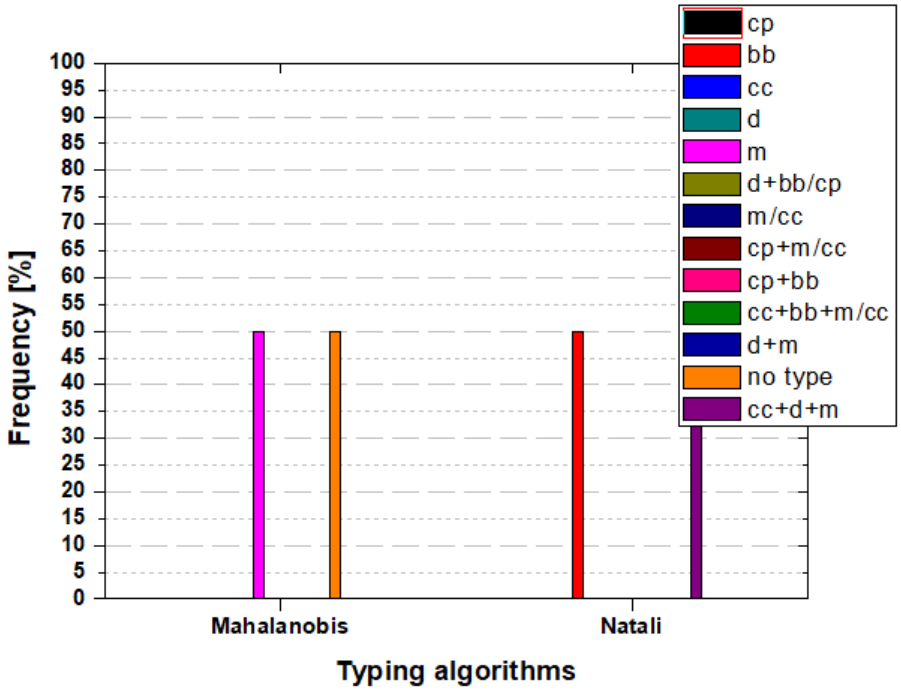

Figure 4.8 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted and Marine".***

Natali algorithm classified as "smoke" the 50% (Fig. 4.15, right part of the figure, red column) of the same cases mentioned above ("continental polluted and marine"), and as "clean continental, dust and marine" the other 50% of them (Fig. 4.15, right part of the figure, purple column).

In Table 4.10 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted and marine" by SCAN for this study.

Table 4.10 *Mean values of intensive optical properties for the "continental polluted and marine" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	0.51 ± 0.32
CI355/532	0.33 ± 0.32
CI532/1064	0.74 ± 0.18
CR355/532	1.22 ± 0.16
CR532/1064	1.68 ± 0.22
LR355 [sr]	47.00 ± 6.14
LR532 [sr]	49.33 ± 6.58
LPDR532 [%]	8.22 ± 4.01

4.2.5. Clean Continental and Marine

The Mahalanobis algorithm classified as "clean continental" the 100% of the SCAN "clean continental and marine" cases (Fig. 4.16, left part of the figure, blue column).

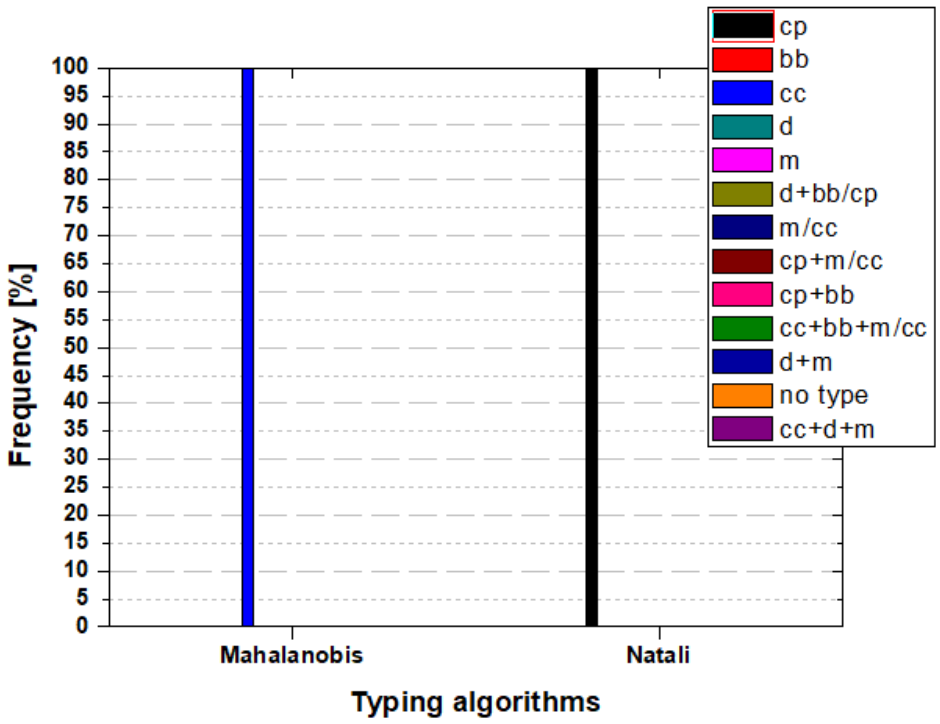

Figure 4.9 *Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Clean Continental and Marine".*

Natali algorithm classified as "continental polluted" the 100% (Fig. 4.16, right part of the figure, red column) of the same cases mentioned above ("continental polluted and marine").

In Table 4.11 we introduce the mean values of the intensive optical properties for the layers classified as "clean continental and marine" by SCAN for this study.

Table 4.11 *Mean values of intensive optical properties for the "clean continental and marine" classified layers by SCAN.*

Intensive optical properties	Mean ± error
AE355/532	0.69 ± 0.29
CI355/532	0.14 ± 0.29
CI532/1064	0.73 ± 0.17
CR355/532	1.78 ± 0.21
CR532/1064	1.66 ± 0.19
LR355 [sr]	51.00 ± 5.88
LR532 [sr]	63.00 ± 7.26
LPDR532 [%]	1.99 ± 0.81

4.3. Mixtures of 3 aerosol types

The "mixture of 3" major category is divided into 3 minor subcategories: "Continental Polluted, Clean Continental and Marine", "Continental Polluted, Smoke and Marine" and "Clean Continental, Smoke and Dust".

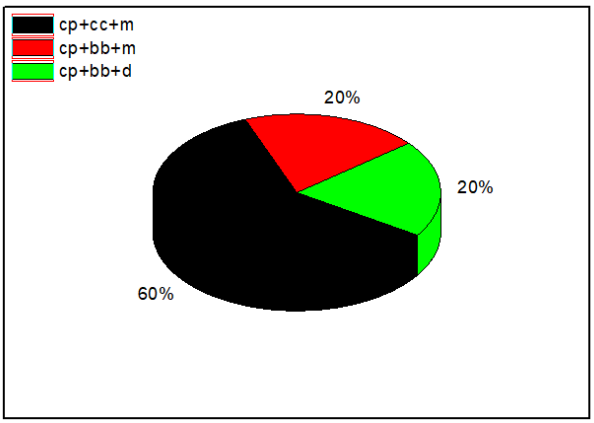

Figure 4.10 *Minor categories of SCAN's results for major category "mixture of 3".*

The "Continental Polluted, Clean Continental and Marine" category (black slice of the pie, Fig. 4.17) represents the 60% (3 cases) of the mixture of 3 aerosol types detected by SCAN. The "Continental Polluted, Smoke and Marine" category (red slice of the pie, Fig. 4.17) represents the 20% (1 cases) of the mixture of 3 aerosol types detected by SCAN. Finally, the "Clean Continental, Smoke and Dust" category (green slice of the pie, Fig. 4.17) represents the last 20% (1 case) of the mixture of 3 aerosol types detected by SCAN.

4.3.1. Continental Polluted, Clean Continental and Marine

The Mahalanobis algorithm classified as "continental polluted" the 33 % of the SCAN "continental polluted, clean continental and marine" cases (Fig. 4.18, left part of the figure, black column). At the same time, Mahalanobis classified as "marine" the 33% of the same cases (Fig. 4.18, left part of the figure, magenta column). Finally, Mahalanobis wasn't able to classify the last 33% (Fig. 4.18, left part of the figure, orange column).

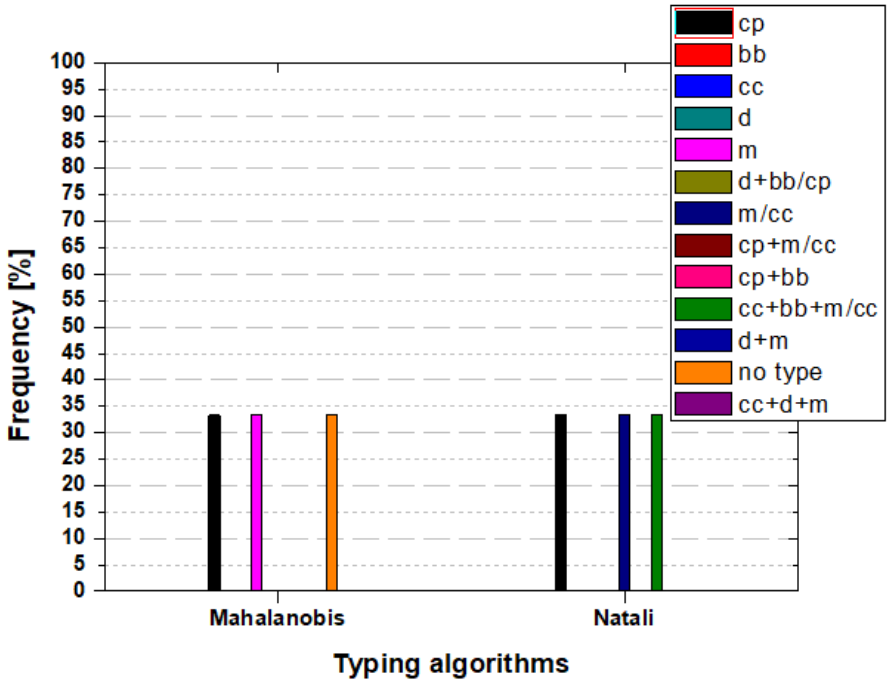

Figure 4.11 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted, Clean Continental and Marine".***

Natali algorithm classified as "continental polluted" the 33% (Fig. 4.18, right part of the figure, black column) of the same cases mentioned above ("continental polluted, clean continental and marine"), as "marine or clean continental" the 33% of them (Fig. 4.18, right part of the figure, dark blue column) and as "clean continental, smoke and marine" the last 33% of the cases SCAN classified as "continental polluted, clean continental and marine".

In Table 4.12 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted, clean continental and marine" by SCAN for this study.

Table 4.12 ***Mean values of intensive optical properties for the "continental polluted, clean continental and marine" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	0.77 ± 0.18
CI355/532	1.30 ± 0.18
CI532/1064	0.98 ± 0.10
CR355/532	1.70 ± 0.12
CR532/1064	2.01 ± 0.13
LR355 [sr]	43.00 ± 3.24
LR532 [sr]	55.67 ± 4.78
LPDR532 [%]	2.89 ± 1.17

4.3.2. Continental Polluted, Smoke and Marine

The Mahalanobis algorithm classified as "clean continental" the 100% of the SCAN "continental polluted, smoke and marine" cases (Fig. 4.19, left part of the figure, blue column).

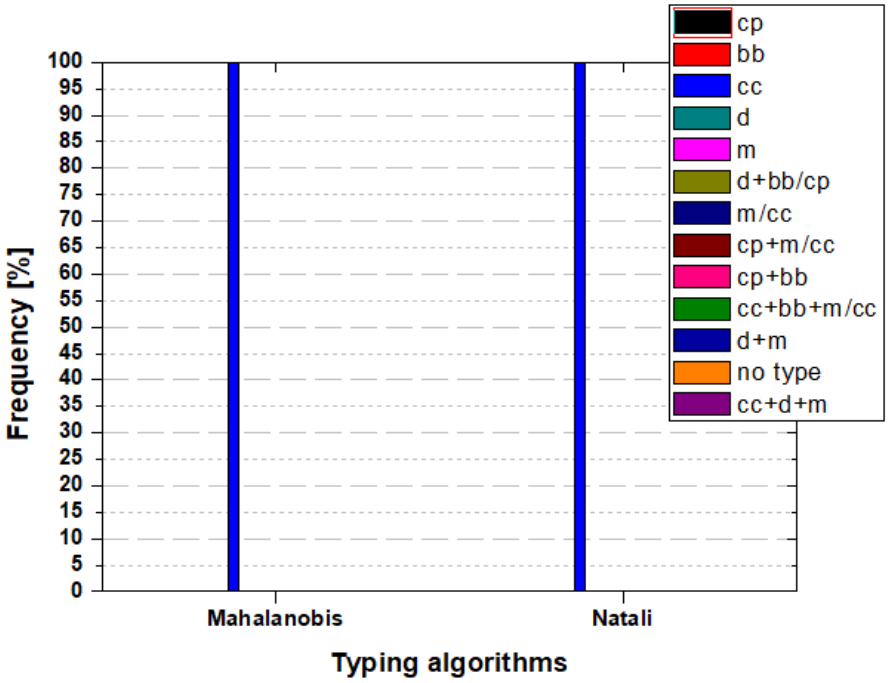

Figure 4.12 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted, Smoke and Marine".***

At the same time, Natali algorithm classified as "clean continental" the 100% (Fig. 4.19, right part of the figure, blue column) of the same cases mentioned above ("continental polluted, smoke and marine").

Table 4.13 ***Mean values of intensive optical properties for the "continental polluted, smoke and marine" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	0.98 ± 0.13
CI355/532	1.60 ± 0.13
CI532/1064	0.92 ± 0.08
CR355/532	1.91 ± 0.10
CR532/1064	1.90 ± 0.10
LR355 [sr]	42.00 ± 2.20
LR532 [sr]	54.00 ± 2.85
LPDR532 [%]	3.48 ± 0.66

In Table 4.13 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted, smoke and marine" by SCAN for this study.

4.3.3. Continental Polluted, Smoke and Dust

The Mahalanobis algorithm classified as "clean continental" the 100% of the SCAN "continental polluted, smoke and dust" cases (Fig. 4.20, left part of the figure, blue column).

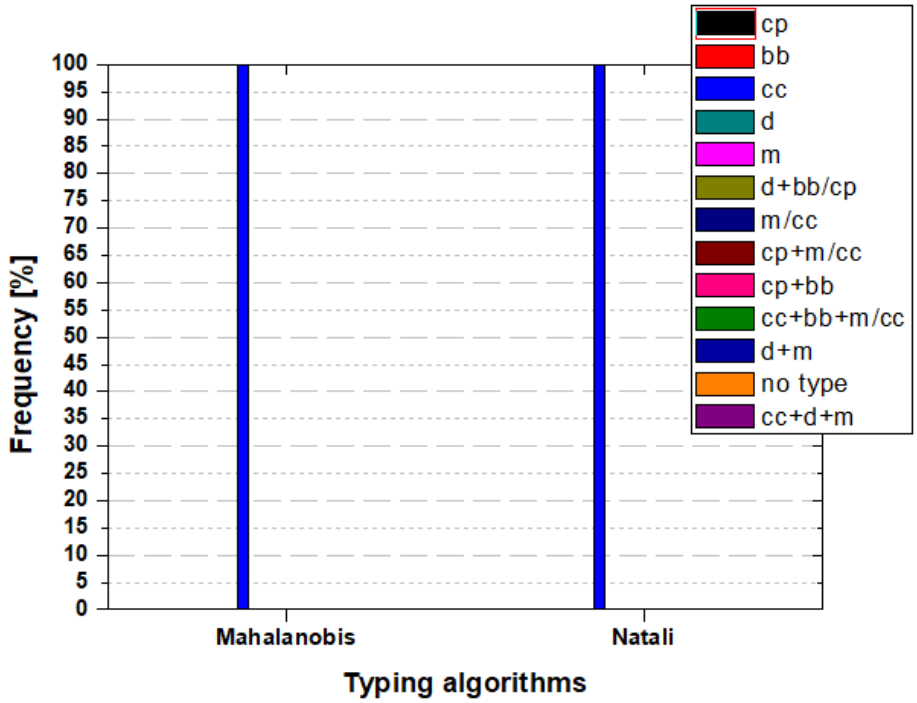

Figure 4.13 ***Frequencies of observed aerosol types by Mahalanobis and Natali for the layers classified by SCAN as "Continental Polluted, Smoke and Dust".***

At the same time, Natali algorithm classified as "clean continental" the 100% (Fig. 4.20, right part of the figure, blue column) of the same cases mentioned above ("continental polluted, smoke and dust").

Table 4.14 ***Mean values of intensive optical properties for the "continental polluted, smoke and dust" classified layers by SCAN.***

Intensive optical properties	Mean ± error
AE355/532	1.25 ± 0.25
CI355/532	1.01 ± 0.25
CI532/1064	0.95 ± 0.14
CR355/532	1.50 ± 0.15
CR532/1064	1.93 ± 0.19
LR355 [sr]	51.00 ± 5.09
LR532 [sr]	47.00 ± 4.71
LPDR532 [%]	4.08 ± 1.44

In Table 4.14 we introduce the mean values of the intensive optical properties for the layers classified as "continental polluted, smoke and dust" by SCAN for this study.

5. Conclusions and Future Work

In this study we developed an improved algorithm, for automated aerosol optical property layer identification based on backward trajectory analysis and satellite data in combination with the application of a number of criteria. SCAN uses the HYSLPIT (6-days) back-trajectories and takes into account the atmospheric layer height above the aerosol source regions (<1km for marine, <2 km for continental polluted, clean continental and dust, 3km for smoke) and distance from the potential fire spot (<8km) as well as fire spot confidence (>80%). Its classification results are based on the amount of time that the air parcel spends above certain already characterized aerosol source region. It is capable of handling a considerable amount of layers in a negligible time. The innovation of SCAN is that it can be used from stations that have not the full set of optical properties in order to characterize the observed aerosol layers above them. However, the above criteria may not serve well all the layers under study.

For the first time, a comparison between three different aerosol classification methods, named: (1) "Mahalanobis distance automatic aerosol type classification", (2) "Neural Network Aerosol Typing Algorithm" (NATALI) and (3) "Source Classification Analysis" (SCAN) is shown.

The lidar station selection depended on the availability of the vertical profiles of the full set ($3\beta + 2\alpha + \delta$) of aerosol optical properties: backscatter coefficient (b_{355}, b_{532}, b_{1064}), extinction coefficient (e_{355}, e_{532}), Lidar Ratio (LR_{355}, LR_{532}), Ångström exponent ($AE_{355/532}$, $CI_{355/532}$, $CI_{532/1064}$) and Linear Particle Depolarization Ratio ($LPDR_{532}$) at the EARLINET database during the period 2014-2018. The 4 lidar stations from which we obtained the aerosol optical properties data : Kuopio (Finland), Bucharest (Romania), Leipzig (Germany) and Potenza (Italy).

Moreover, 48 dates of lidar observations by multiwavelength Raman and depolarization lidars from the aforementioned lidar stations have been studied in the frame of this Thesis. For some dates, more than one layer has been observed increasing the number of layers to 97.

Four different categories were created depending on the results from SCAN: "pure", "mixture of 2", "mixture of 3" and "no type". Each of these categories was, then, divided into subcategories depending on the typing results of SCAN. The "pure" category was divided into 5 subcategories "Continental Polluted" (cp), "Clean Continental" (cc), "Smoke" (bb), "Dust" (d) and "Marine" (m).

The “mixture of 2” category was divided into 5 subcategories, “Continental Polluted and Smoke”, “Continental Polluted and Clean Continental” (cp+cc), “Continental Polluted and Marine” (cp+m), “Clean Continental and Marine” (cc+m) and “Continental Polluted and Dust” (cp+d). Finally, the “mixture of 3” category was divided into 3 subcategories “Continental Polluted, Clean Continental and Marine” (cp+cc+m), “Continental Polluted, Smoke and Marine” (cp+bb+m) and “Continental Polluted, Smoke and Dust” (cp+bb+d).

The performance of the Mahalanobis and Natali algorithms was studied for each of the above categories. The mean values of aerosol optical properties for each of these categories have been calculated.

Natali is an automatic aerosol optical property dependent classification algorithm. It is able to identify pure aerosol types, mixtures of two aerosol types and mixtures of three aerosol types. Natali takes as inputs the profiles of aerosol optical properties uploaded to the EARLINET database and gives all the classification results together with the mean values and their uncertainties of all aerosol optical properties of the layers under study. We showed that this algorithm is capable to well classify the pure continental polluted aerosol layers but it has a difficulty to classify the pure clean continental and pure smoke aerosol types probably due to the overlapping mean values of aerosol optical properties of these aerosol types, compared to the results obtained from the SCAN classification algorithm.

Mahalanobis is an automatic aerosol optical property dependent classification algorithm. It is able to identify pure aerosol types and mixtures of two aerosol types. Mahalanobis takes as inputs the mean values of the aerosol optical properties of each layer under study. We showed that this algorithm is capable to classify the pure continental polluted aerosol layers but it has a difficulty to classify the pure clean continental and pure smoke aerosol types probably due to the overlapping mean values of aerosol optical properties of these aerosol types, just like Natali, compared to the results obtained from the SCAN classification algorithm. Finally, Mahalanobis is not able to classify aerosol layers which consist of more than two aerosol types with relatively equal contribution to the layer, compared to the results from SCAN classification algorithm. These layers are classified as “no type” by Mahalanobis algorithm.

Concerning the mean values of aerosol optical properties, we found relatively smaller values compared to those of the literature, especially to the “lidar ratios” and the “linear particle depolarization ratios“. This behavior might be a result of the dependence of the optical property on the time the trajectory spent above the source area, concerning the pure types.

In the future, we plan to study the role of the atmospheric humidity onto the aerosol transformation and the change that this contribution induces to the aerosol optical properties. Moreover, we plan to further investigate the role of the predominant aerosol type to the optical properties of the aerosols, through the percentages that Mahalanobis and Scan bring upon.

We would like to perform the same methodology but without the use of the linear particle depolarization ratio parameter to investigate how this parameter contributes to the final classification of the two aerosol optical property dependent algorithms. Finally, but most importantly, we plan to perform our methodology to a considerably larger amount of layers so that the statistical analysis will be statistically significant.

References

Althausen, D., et al. "Portable Raman Lidar PollyXTfor Automated Profiling of Aerosol Backscatter,Extinction, and Depolarization." *Journal of Atmospheric and Oceanic Technology* 2009: 2366-2378.

Amiridis, V., et al. "Optical characteristics of biomass burning aerosols overSoutheastern Europe determined from UV-Raman lidarmeasurements." *Atmos. Chem. Phys.* 2009: 2431–2440.

—. "Smoke injection heights from agricultural burning in Eastern Europe as seen by CALIPSO." *Atmospheric Chemistry and Physics* 2010: 11567-11567.

—. "Smoke injection heights from agricultural burning in EasternEurope as seen by CALIPSO." *Atmos. Chem. Phys.* 2010: 11567–11576.

Ansmann, A., et al. "European pollution outbreaks during ACE 2: Lofted aerosol plumes observed with Raman lidar at the Portuguese coas." *Journal of Geophysical Research* 2001: 20,725-20,733.

—. "Long-range transport of Saharan dust to northern Europe: The 11– 16 October 2001 outbreak observed with EARLINET." *Journal of Geophysical Research* 2003: 4783.

Ansmann, A., M. Riebesell and C. Weitkamp. "Measurement of atmospheric aerosol extinction profiles with a Raman lidar." *Optics Letters* 1990: 746–748,.

Avila, G., et al. "Ro-vibrational Raman Cross Sections of Water Vaporin the OH Stretching Region." *Journal of Molecular Spectroscopy* 1999: 77–92.

Baars, H., et al. "Aerosol profiling with lidar in the Amazon Basin during the wetand dry season." *Journal of Geophysical Research* 2012: D21201.

—. "An overview of the first decade of PollyNET: an emergingnetwork of automated Raman-polarization lidars forcontinuous aerosol profiling." *Atmos. Chem. Phys.* 2016: 5111–5137.

Balis, D., et al. "Raman lidar and sunphotometric measurements of aerosoloptical properties over Thessaloniki, Greece duringa biomass burning episode." *Atmospheric Environment* 2003: 4529–4538.

Barrie, L.A. "Arctic air pollution: An overview of current knowledge." *Atmospheric Environment* 1986: 643-663.

Belegante, L., et al. "Retrieval of the boundary layer height from active and passive remote sensors. Comparison with a NWP model." *Acta Geophysica* 2014: 276–289.

Biele, J., G. Beyerle and G. Baumgarten. "Polarization lidar: Corrections of instrumental effects." *Optical Society of America* 2000: 427-435.

Binietoglou, I., et al. "A methodology for investigating dust model performance using synergistic EARLINET/AERONET dust concentration retrievals." *Atmospheric Measurement Techniques Discussions* 2015: 3605–3666.

Burton, S.P., et al. "Aerosol classification from airborne HSRL and comparisons withthe CALIPSO vertical feature mask." *Atmos. Meas. Tech.* 2013: 1397–1412.

—. "Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples." *Atmos. Meas. Tech.* 2012: 73-98.

Chaikovsky, A., et al. "Lidar-Radiometer Inversion Code (LIRIC) for the retrieval ofvertical aerosol properties from combined lidar/radiometer data:development and distribution in EARLINET." *Atmos. Meas. Tech.* 2016: 1181–1205.

Davies, D. K., et al. "Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data." *IEEE Transactions on Geoscience and Remote Sensing* 2009: 72 - 79.

Dawson, K.W., et al. "Spaceborne observations of the lidar ratio of marine aerosols." *Atmos. Chem. Phys.* 2015: 241–3255.

Draxler, R.R. and G.D. Hess. "An overview of the HYSPLIT_4 modelling system for trajectories, dispersion and deposition." *Aust. Met. Mag.* 1998: 295-308.

Engelmann, R., et al. "The automated multiwavelength Raman polarization andwater-vapor lidar PollyXT: the neXT generation." *Atmos. Meas. Tech.* 2016: 1767–1784.

Fernald, F. G. "Analysis of atmospheric lidar observations: some comments." *Applied Optics* 1984: 253-254.

Fitzgerald, J. W. "Marine aerosols: A review." *Atmospheric Environment* 1991: 533-545.

Freudenthaler, V., et al. "Depolarization ratio profiling at severalwavelengths in pure Saharan dust during SAMUM2006." *Tellus* 2009: 165-179.

Giannakaki, E., et al. "Optical and microphysical characterization of aerosol layers over South Africa by means of multi-wavelength depolarization and Raman lidar measurements." *Atmos. Chem. Phys.* 2016: 8109–8123.

—. "Optical properties of different aerosol types: seven years of combined Raman-elastic backscatter lidar measurements in Thessaloniki, Greece." *Atmos. Meas. Tech.* 2010: 569–578.

Giglio, L., et al. "An Enhanced Contextual Fire Detection Algorithm for MODIS." *Remote Sensing of Environment* 2003: 273–282.

Gross, S., et al. "Aerosol classification by airborne high spectral resolution lidar observations." *Atmos. Chem. Phys.* 2013: 2487–2505.

Guibert, S., et al. "The vertical distribution of aerosol over Europe-synthesis of one year of EARLINET aerosol lidar measurements and aerosol transport modeling with LMDzT-INCA." *Atmospheric Environment* 2004: 2933-2943.

Heese, B., et al. "Aerosol Properties over Southeastern China from Multi-Wavelength Raman and Depolarization Lidar Measurements." *EPJ Web of Conferences* 2016: 23018.

Hess, M., Koepke, P., and Schult, I. "Optical properties of aerosols and clouds: The software package OPAC." *B. Am. Meteorol. Soc.* 1998: 831–844.

Hitchcock, D. R., L. L. Spiller and W. E. Wilson. "Sulfuric acid aerosols and HCl release in coastal atmospheres: Evidence of rapid formation of sulfuric acid particulates." *Atmospheric Environnment* 1980: 165-182.

Hobbs, P. Hobbs, P. *Aerosol- Cloud- Climate Interactions*. SEATTLE, WASHINGTON: ACADEMIC PRESS, INC., 1993.

"IPCC." 2014.

Jacobson, M. *FUNDAMENTALS OF ATMOSPHERIC PARTICLES*. Cambridge: © Cambridge University Press, 2005.

Jaenicke, R. and L. Schutz. "Comprehensive Study of Physical and Chemical Properties of the Surface Aerosols in the Cape Verde Islands Region." *Journal of Geophysical Research* 1978: 3585-3599.

Jaenicke, R. "Tropospheric Aerosols." *International Geophysics* 1993: 1-31.

Klett, J. D. "Stable analytical inversion solution for processing lidar returns." *APPLIED OPTICS* 1981: 211-220.

Koepke, P., Hess, M., Schult, I., and Shettle. "Global Aerosol Data Set." *Max-Planck-Institut für Meteorologie* 1997.

Kokkalis, P., et al. "Optical, microphysical, mass and geometrical properties of agedvolcanic particles observed over Athens, Greece, during theEyjafjallajökull eruption in April 2010 through synergy ofRaman lidar and sunphotometer measurements." *Atmos. Chem. Phys.* 2013: 9303–9320.

Leskinen, A., et al. "Overview of the research activities and results at Puijo semi-urban measurement station." *Boreal Environment Research* 2009: 576-590.

Long, D.A. *Raman spectroscopy*. New York: 1-12, 1977.

Madonna, F., et al. "CIAO: the CNR-IMAA advanced observatory for atmosphericresearch." *Atmos. Meas. Tech.* 2011: 1191–1208.

Mahowald, N., et al. "The size distribution of desert dust aerosols and its impact on the Earth system." *Aeolian Research* 2014: 53-71.

Mamouri, R. E. "Low Arabian dust extinction-to-backscatter ratio." *GEOPHYSICAL RESEARCH LETTERS* 2013: 4762–4766 .

Mamouri, R.E., et al. "Validation of CALIPSOspace-borne-derived aerosol verticalstructures using a ground-based lidar inAthens, Greece." *Atmos. Meas. Tech.* 2009: 561–587.

Matthias, V., et al. "Vertical aerosol distribution over Europe: Statistical analysis of Raman lidar data from 10 European Aerosol Research Lidar Network (EARLINET) stations." *Journal of Geophysical Research: Atmospheres* 2004: D18201.

Mattis, I., et al. "Ten years of multiwavelength Raman lidar observationsof free-tropospheric aerosol layers over central Europe: Geometricalproperties and annual cycle." *Journal of Geophysical Research* 2008: D20202.

—. "Volcanic aerosol layers observed with multiwavelength Ramanlidar over central Europe in 2008–2009." *Journal of Geophysical Research* 2010: D00L04.

McLachlan, G J. "Mahalanobis Distance ." *RESONANCE* 1999: 20-26.

Mishchenko, M.I., L.D. Travis and D.W. Mackowski. "T-matrix computations of light scattering by nonspherical particles: A review." *J. Quant. Spectrosc. Ra.* 1996: 535-575.

Mona, L., et al. "EARLINET dust observations vs. BSC-DREAM8b modeledprofiles: 12-year-long systematic comparison at Potenza, Italy." *Atmos. Chem. Phys.* 2014: 8781–8793.

—. "Multi-wavelength Raman lidar observations of the Eyjafjallajökull volcanic cloud over Potenza, southern Italy." *Atmos. Chem. Phys.* 2012: 2229-2244.

—. "One year of CNR-IMAA multi-wavelength Raman lidarmeasurements in coincidence with CALIPSO overpasses: Level 1products comparison." *Atmos. Chem. Phys.* 2009: 7213–7228.

Muller, D. "Aerosol-type-dependent lidar ratios observed withRaman lidar." *JOURNAL OF GEOPHYSICAL RESEARCH* 2007.

Muller, D., et al. "Closure study on optical and microphysicalproperties of a mixed urban and Arctic haze air massobserved with Raman lidar and Sun photometer." *Journal of Geophysical Research* 2004: D1306.

Myhre, G., et al. "Aerosols and their Relation to Global Climate and Climate Sensitivity." *Nature Education Knowledge* 2013.

—. "Modelled radiative forcing of the direct aerosol effect." *Atmos. Chem. Phys* 2009: 1365–1392.

NASA Earth Observatory : Home. 2019.

Nemuc, A., et al. "Assessment of aerosol's mass concentrations from measured linearparticle depolarization ratio (vertically resolved) and simulations." *Atmos. Meas. Tech.* 2013: 3243–3255.

—. "Optical Properties of Aerosols from Lidar Data and Other Ground-based Instruments near Bucharest." *researchgate.net* 2008.

Nicolae, D., et al. "A neural network aerosol-typing algorithm based on lidar data." *Atmos. Chem. Phys.* 2018: 14511–14537.

—. "Characterization of fresh and aged biomass burning events using multiwavelength Raman lidar and mass spectrometry." *Journal of Geophysical Research: Atmospheres* 2015: 2956–2965.

—. "Characterization of fresh and aged biomass burning events using multiwavelength Raman lidar and mass spectrometry." *Journal of Geophysical Research: Atmospheres* 2013: 2956-2965.

—. "LIDAR monitoring of aerosols loading over Bucharest." *Journal of Optoelectronics and Advanced Materials* 2006: 238-242.

—. "Strenghts and limitations of the Natali code for aerosol typing from multiwavelength Raman lidar observations." *EPJ Web of Conferences* 2018.

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.-Atmos.* 2013: 2956–2965.

Nisantzi, A., et al. "Middle East versus Saharan dust extinction-to-backscatter ratios." *Atmos. Chem. Phys.* 2015: 7071–7084.

Omar, A., et al. "The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm." *Jouran of Atmospheric and Oceanic Technology* 2009: 1994-2014.

Ortiz-Amezcua, P., et al. "Microphysical characterization of long-range transported biomassburning particles from North America at three EARLINET stations." *Atmos. Chem. Phys.* 2017: 5931–5946.

Ottar, B. "Arctic air pollution: A Norwegian perspective." *Atmospheric Environment* 1989: 2349-2356.

Pandolfi, M., et al. "Three years of Raman Lidar Measurements of Tropospheric Aerosol over Potenza in the framework of EARLINET." 2004.

Papagiannopoulos, N., et al. "An automatic observation-based aerosol typing methodfor EARLINET." *Atmos. Chem. Phys.* 2018: 15879–15901.

—. "CALIPSO climatological products: evaluation and suggestionsfrom EARLINET." *Atmos. Chem. Phys.* 2016: 2341–2357.

Papayannis, A. "Optical, size and mass properties of mixed type aerosols in Greece and Romania as observed by synergy of lidar and sunphotometers in combination with model simulations: A case study." *Science of the Total Environment* 2014: 277-294.

Papayannis, A., et al. "Optical properties and vertical extension of aged ash layers over the EasternMediterranean as observed by Raman lidars during the Eyjafjallajökull eruptionin May 2010." *Atmospheric Environment* 2012: 56-65.

—. "Systematic lidar observations of Saharan dust over Europe in theframe of EARLINET (2000–2002)." *Journal of Geophyical Research* 2008: D10204.

Pappalardo, G., et al. "EARLINET: towards an advanced sustainable European aerosollidar network." *Atmos. Meas. Tech.* 2014: 2389–2409.

—. "Four-dimensional distribution of the 2010 Eyjafjallaj¨okull volcaniccloud over Europe observed by EARLINET." *Atmos. Chem. Phys.* 2013: 4429–4450.

—. "Raman lidar observations of aerosol emitted during the 2002 Etna." *Geophysical Research Letters* 2004: L05120.

Penning de Vries, M.J.M., et al. "A global aerosol classification algorithm incorporating multiplesatellite data sets of aerosol and trace gas abundances." *Atmos. Chem. Phys.* 2015: 10597–10618.

Poschl, U. "Atmospheric Aerosols: Composition, Transformation, Climate and Health Effects." *Angew. Chem. Int. Ed.* 2005: 7520 – 7540.

Prospero, J.M. and D.L. Savoie. "Effect of continental sources on nitrate concentrations over the Pacific Ocean." *Letters to Nature* 1989: 687-689.

Rahn, K.A. "RELATIVE IMPORTANCES OF NORTH AMERICA AND EURASIA AS SOURCES OF ARCTIC AEROSOL." *Atmospheric Environment* 1981: 1447-1455.

Raman, C.V. and K.S. Krishnan. "A new type of secondary radiation." *Nature* 1990: 746-748.

Remer, L.A., et al. "Global aerosol climatology from the MODIS satellite sensors." *Journal of Geophysical Research: Atmospheres* 2008.

Roberts, G., et al. "Retrieval of biomass combustion rates and totals from fire radiativepower observations: Application to southern Africa using geostationary SEVIRI imagery." *Journal of Geophysical Research* 2005: D21111.

Schmauss, A. and A. Wigand. *Die Atmosphäre als Kolloid*. Braunschweig: Vieweg, 1929.

Schwarz, A. "Aerosol typing over Europe and its benefits for." *University of Leipzig* 2016.

Seinfeld, J.H. and S.N. Pandis. *ATMOSPHERIC CHEMISTRY AND PHYSICS From Air Pollution to Climate Change*. Hoboken, New Jersey: John Wiley & Sons, Inc., 2006.

Sicard, M., et al. "Monitoring of the Eyjafjallaj ̈okull volcanic aerosol plume over theIberian Peninsula by means of four EARLINET lidar stations." *Atmos. Chem. Phys.* 2012: 3115–3130.

Soupiona, O., et al. "Long-term systematic profiling of dust aerosol optical properties using the EOLE NTUA lidar system over Athens, Greece (2000–2016)." *Atmospheric Environment* 2018: 165-174.

Tesche, M., et al. "Ground-based validation of CALIPSO observations of dust and smoke in the Cape Verde region." *Journal of Geophysical Research: Atmospheres* 2013: 2889–2902.

Valenzuela, A., et al. "Aerosol transport over the western Mediterranean basin: Evidence of the contribution of fine particles to desert dust plumes over Alborán Island." *Journal of Geophysical Research: Atmospheres* 2014: 14,028-14,044.

Veselovskii, I., et al. "Application of randomly oriented spheroids for retrieval of dustparticle parameters from multiwavelength lidar measurements." *Journal of Geophysical Research* 2010: D21203.

Wandinger, U. "Raman Lidar." 2005. 241-271.

Wandinger, U., et al. "HETEAC:The Aerosol Classification Model for Earth CARE." *EPJ Web of Conferences* 2016: 01004.

Waterman, P.C. "Symmetry, Unitarity, and Geometry in Electromagnetic Scattering." *PHYSICAL REVIEW D* 1971: 825–839.

Weitkamp, C. *LIDAR, Range Resolved Optical Remote Sensing of the Atmosphere*. Germany: Springer, 2005.

Wiegner, M., et al. "The May/June 2008 Saharan dust event over Munich:Intensive aerosol parameters from lidar measurements." *Journal of Geophysical Research* 2011: D23213.

Wilks, S. "Multivariate statistical outliers." *Sankhya: The Indian Journal of Statistics* 1963: 407–426.