Biomass burning events measured by lidars in EARLINET. Part I. Data analysis methodology.

Mariana Adam[1], Doina Nicolae[1], Iwona S. Stachlewska[2], Alexandros Papayannis[3], Dimitris Balis[4]

[1]National Institute for R&D in Optoelectronics, Magurele, 077225, Romania
[2]Institute of Geophysics, Faculty of Physics, University of Warsaw, 02093, Poland
[3] National Technical University of Athens, Department of Physics, Athens, 15780, Greece
[4] Laboratory of Atmospheric Physics, Aristotle University of Thessaloniki, Thessaloniki, 54124, Greece

Correspondence to: Mariana Adam (mariana.adam@inoe.ro)

Abstract. The methodology of analysing the biomass burning events recorded in the database of the European Aerosol Research Lidar Network in the frame of Aerosol Cloud and Trace Gases Research Infrastructure is presented. The period of 2008-2017 was chosen to analyse all of the events stored in the database under Forest Fire category for a number of 14 stations available. The data provided ranged from complete data sets (particle backscatter, extinction and linear depolarization ratio profiles) to single profiles (particle backscatter coefficient profile). Smoke layers geometry was evaluated and the mean optical properties within each layer were computed. The backtrajectories technique was used to double check the source of all pollution layers. The biomass burning layers were identified taking into account the presence of the fires along the backtrajectory. The biomass burning events are analysed by the means of the intensive parameters. The analysis was structured in three directions: I) common biomass burning source (fire) recorded by at least two stations, II) long-range transport from North America, III) analysis over four geographical regions (South-East Europe, North-East Europe, Central Europe and South-West Europe). Based on backtrajectories calculations and fires' location, the lidar measurements can be labelled either as measurements of 'single fire' or 'mixed fires' (case I), measurements of North America fires or measurements of mixed North America and local fires (case II). The histogram of the fires' location reveals the smoke sources for each region. For each region, statistics on intensive parameters is performed. The sources' origin of the intensive parameters is categorized based on the continental origin of the air-mass (European, African, Asian, North American or a combination of them). The methodology presented here is meant to provide a perspective to explore large amount of lidar data and deliver novel approaches to analyse the intensive parameters based on the assigned biomass burning sources. A thorough consideration of all potential fires' sources reveals that most of the time the lidar measurements characterise the smoke from a mixture of fires. A comprehensive discussion over all results (based on the intensive parameters and the sources' location), will be given in a companion paper submitted to ACP EARLINET special issue.

1 Introduction

Biomass burning (BB) represents one of the major sources of atmospheric particles (aerosols). The sources of BB are both natural (wild fires) and anthropogenic (controlled fires). The direct effect on radiative transfer can be both negative and positive as a consequence of the opposite effects of the two main components: black carbon (absorption, positive effect) and organic carbon (scattering, negative effect) (Fiebig et al, 2002; Bond et al., 2013; Myhre et al., 2013) with a general radiative forcing close to zero (https://www.ipcc.ch/site/assets/uploads/2018/02/WG1AR5_Chapter08_FINAL.pdf, IPCC, 2013, last access 20200203). As indirect effect, BB can act as cloud condensation nuclei (e.g. Yu, 2000) or ice nuclei (e.g. Prenni et al, 2012). At ground level or within the Planetary Boundary Layer (PBL), strong pollution events produce a large reduction of the visibility over various regions (e.g. Adam et al., 2004), which can affect the traffic and more importantly can cause serious health issues for humans (e.g. Pahlow et al., 2005; Sapkota et al., 2005, Alonso-Blanco et al., 2018). Numerous wild forest fires events occur each year in Europe (refer to annual reports from European Commission on "Forest Fires in Europe, Middle East and North Africa YYYY" at http://effis.jrc.ec.europa.eu/reports-and-publications/annual-fire-reports/ available since 2000, where YYYY represents the year; last access 20191126). Therein, the studies report the impact of forest fires in Europe, providing the number of fires and burnt area by country. The 2017 report (San-Miguel-Ayanz et al, 2018) shows that the number of fires was higher as comparted with the average over 2000-2017 period while the burnt area was even larger, especially for Portugal. For the Southern States (Portugal, Spain, France, Italy and Greece), most effected by fires, the highest number of fires occurred in Portugal (44%) and Spain (29%) while the largest burn area occurred in Portugal (59%) and then Spain (19%) and Italy (18%). As mentioned in the report, climate change affects forest fires through the weather conditions, and through its effects on vegetation and fuel (combustible material). Besides the European fires which can act locally or through short range transport within continent, Europe is also affected by long range transport, the most common being the smoke transported from Russia or even from North America (aged smoke). Note that over Poland the BB is the most important source of long range transported particles even in boundary layer (e.g. Wang et al., 2019). In a recent study by Nicolae et al. (2019) it was shown that the smoke is the most predominant aerosol type in Europe for long-range transport, as revealed by lidar measurements over 2008-2018 period over 17 stations equipped with multiwavelength lidars. The inventory of fires is provided by European Commission through EFFIS (European Forest Fire Information System) at European level and by NASA through FIRMS (Fire Information for Resource Management System) at global level (https://firms.modaps.eosdis.nasa.gov/, last access 20191126) (Davies et al., 2009). Both databases are based on MODIS (Moderate Resolution Imaging Spectroradiometer) satelitare data.

Lidars provide measurements of the smoke aerosol being able to deliver the boundaries of the smoke layers as well as the optical properties of the smoke aerosol in the layers. Both layers geometry and optical properties can be used as a validation of the transport models (e.g. van Drooge et al., 2016). For a detailed analysis of the data requirements by models, see Benedetti et al., 2018 and references herein. Referring to BB aerosol, the study mentions the usefulness of lidar retrieval of the plume height. Layer geometry can also be used to constrain specific satellite profiles retrievals which, as known, have less sensitivity

towards the ground (e.g. Bond et al., 2013). The optical properties can be used (closure studies) to model the microphysical properties of the aerosol (e.g. Fiebig et al., 2002; Fiebig et al., 2003). Regarding the impact of BB aerosol on weather forecast, the study by Zhang et al. (2016) suggests that the BB' effect is seen for AOD (aerosol optical depth) at 550nm larger than 1. Ultimately the researchers try to establish the BB effect on the radiative forcing (e.g. Fiebig et al., 2002; Lolli et al., 2019; Markowicz et al., 2016).

Within these applications, the EARLINET (European Aerosol Research Lidar Network) can provide spatial and temporal coverage of the BB transport over Europe. EARLINET is part of Aerosol Cloud and Trace Gases Research Infrastructure (ACTRIS) (http://actris.nilu.no/, last access 20191126), which provides ground-based, insitu or remote sensing. EARLINET provides measurements of the biomass burning aerosol with first lidar profiles dating from 2002. More systematic measurements over the network can be seen after 2012 (closely related with the increase of the number of stations in the network). A review of the lidar studies over BB will be given at a later stage in this paper.

The current study envisages the research on the biomass burning aerosol as measured by 14 stations in EARLINET over the period 2008-2017. This paper presents the methodology of data analysis, whereas a companion paper (submitted to ACP, EARLINET special issue) focuses in detail on the discussion of the outcome. The EARLINET is shortly presented in Section 2. In Section 3 we briefly discuss the extra undertaken data quality checks. Data analyses follows in Section 4. The data interpretation (Section 5) is based on aerosol intensive parameters (lidar ratio - LR, extinction Ångström exponent - EAE, backscatter Ångström exponent - BAE, linear particle depolarization ratio - PDR). Examples are given for the following research directions: I) fires events (same source) as observed by two lidar systems, II) long range transport (LRT) from North America, III) study based on four geographical measurements regions including the histogram of the fire sources locations for each region. Summary and conclusions follow in Section 6. A short introduction to the present methodology was given at ILRC29 (Adam et al., 2019). A list of acronyms used in the current work is given in the Supplement (Table S1).

2 EARLINET database

The EARLINET database consists currently (October 2019) of 30 lidar stations covering most of Europe and one location outside Europe (Dushanbe, Tajikistan). There are three other non-permanent stations as well as six newly joining stations. Twelve stations are not active (https://www.earlinet.org/index.php?id=105, last access 20191126). A review of the EARLINET network is given by Pappalardo et al. (2014). The data submitted to the database belongs to one or more of the following categories: climatology (measurements taken at specific times on Mondays and Thursdays), Calipso (measurements taken during Calipso overpasses), Saharan dust (measurements taken during Saharan dust intrusions), volcanic eruptions (measurements taken during volcanic eruptions), forest fires (measurements taken during forest fires events), diurnal cycles (measurements taken continuously over the whole day), Cirrus (measurements of Cirrus clouds), photosmog (measurements during photosmog episodes), rural/urban (measurements taken under rural or urban environments) and stratosphere (measurements at stratospheric levels). Most of the measurements are taken under Climatology category followed by Calipso

and Saharan dust categories (Pappalardo et al. 2014). The largest data set regarding the optical properties was submitted by each station following its own retrieval procedure. Note that while Climatology and Calipso measurements are mandatory in the network, all the other categories are built based on voluntary bases. Currently it is envisaged the use of Single Calculus Chain (SCC) developed in EARLINET for all raw data submitted by each station (see D'Amico et al, 2015; D'Amico et al, 2016; Mattis et al, 2016). There are two types of files submitted to the database for the optical properties: b-files (containing particles backscatter coefficient) and e-files (containing particles extinction coefficient and sometimes particles backscatter coefficient retrieved using both elastic and Raman channel). Other variables may be reported as well (see Pappalardo et al., 2014). For systems with depolarization capability, the retrieval of the particles linear depolarization ratio (PDR) is recorded in b-files except Warsaw, which records it in both b-files and e-files.

2.1 EARLINET Forest Fire category

For the present study, the data submitted under Forest Fire (FF) category was considered. Over the period 2008-2017 (when most of the data was submitted), a number of 3759 b-files and e-files were available at 20 stations corresponding to the emission wavelengths 355, 532 and 1064 nm. Another 256 files were submitted at different emission wavelengths (313, 351, 510 and 694 nm). In addition, the Hysplit (Hybrid Single-Particle Lagrangian Integrated Trajectory model) model was available since September 2007 (Stein et al., 2015; Rolph et al., 2017). Hysplit model is used to double check the origin of the aerosols from various layers (as discussed later on). We also considered only the data whose emission wavelength was at 355, 532 and 1064 nm as these are the most used in the lidar community and allows the direct intercomparing of the optical properties. After preliminary quality control checks (discussed in chapter 3), 14 stations were selected over 2008-2017 period, delivering 2341 files (~ 65 % of total), quality checked. The geographical location of the stations is shown in Fig. 1. Also shown, the four geographical regions analysed separately further on (section 5.3). An overview over the number of stations reporting data for FF category as compared with their total number of files submitted (as for March 2018), reveal that the following. FF category represented ~ 40% of the total measurements for Thessaloniki, Athens and Warsaw, ~ 20 % for Minsk and ~ 10% for Bucharest and Granada.

Several stations (Athens, Barcelona, Bucharest, Potenza, Observatory Hohenpeißenberg, Thessaloniki, Warsaw) were committed to send either reprocessed data or to process additional data using SCC, which were not yet included yet in EARLINET (most of them for 2017) as for March 2018. Currently (20200625) all the data are uploaded in Earlinet database. Within these circumstances, the number of files increased to 3589 (QC for 2341 files already on EARLINET database). The diagram of the methodology is shown in Fig. 2. Following Ist step in methodology (Fig. 2), the number of input files (b-files and e-files) for each station is shown in Fig. 3 and Table 1. As a first remark, we can observe that we handle a large variety of lidar systems. Thus, we have one system with 3b+2e+2d (Warsaw), three systems with 3b+2e+1d (Bucharest, Observatory Hohenpeißenberg and Potenza), five systems with 3b+2e (Athens, Cabauw, Evora, Granada, and Thessaloniki), four systems with 3b (Barcelona, Belsk, Leipzig and Minsk) and one system with 1b (Sofia). The total number of b-files and e-files is shown in Table 1. The total number of particles backscatter coefficient, extinction coefficient and linear depolarization ratio after the

first quality checks (EARLINET checks) is: #βp355=969, #βp532=864, #βp1064=830, #κp355=524, #κp532=503, #PDR355=131 and #PDR532=320 (total of 4141 profiles). The distribution by stations follows closely Fig. 3. The stations with the largest contribution for Forest Fire category are Athens, Bucharest, Granada, Thessaloniki and Warsaw.

Note that the number of particle extinction coefficient profiles equals the number of "e" files while the number of particles backscatter coefficient at 1064nm equals the number of b1064 files. In the case of particles backscatter coefficient at 355nm and 532nm, their number can be larger than the number of "b" files as retrievals of backscatter coefficient is reported also in "e" files, employing the use of Raman channel in the retrieval of backscatter coefficient (using either Raman method or aerosol backscattering ratio method). As seen in Table 1, this is the case for the stations Bucharest, Granada, Potenza and Warsaw. For the same time stamp, the profile from "e" file is kept. The reason behind is that the backscatter coefficient will be further used to calculate the lidar ratio, and the same spatial resolution as the extinction coefficient is desirable. The occurrence in time of the retrievals is shown in Fig. 4. Only the cases for which the layers have a fire origin are shown (smoke layers). As observed, there are sporadic measurements reported for the years 2008-2011. Starting with 2012, more stations started reporting biomass burning pollution events. The total number of time stamps for which we have at least one profile with an optical property is shown in the last column of Table 1.

2.2 Overview of the metadata

Metadata refers to those parameters which give information about the instrument and site location, data processing or analysing, and are saved as general attributes in the "e" and "b" netCDF files. The fixed parameters refer to the measurement position, system name, emission and detection wavelengths. The dynamic parameters change in time and space: date and time of measurement, raw resolution, time resolution (shots averaged), detection mode, final resolution, evaluation method. Another two parameters (Input Parameters and Comments) can give additional information (e.g. values used for Rayleigh calibration). Detection mode (type of data acquisition) can be analog, photon counting or both (glued signal). Evaluation method (approach used to retrieve the optical properties) can be Fernald-Klett method (used in b-files to calculate particle backscatter coefficient), Raman method (used in e-files to determine particles extinction coefficient), aerosol backscatter ratio to obtain particles backscatter coefficient using a combination of both elastic and Raman channels (recorded in e-files). The information for parameters is provided in a free style comment format and we could not perform the quantification for all parameters (e.g. we could not determine which was the final spatial resolution nor the method used for smoothing and error calculation). In general, there are three approaches reported: 1) keep the errors within an established threshold and thus vary the spatial resolution, 2) maintain a fine spatial resolution (with high statistical errors), 3) perform variable smoothing (e.g. lower spatial resolution above PBL). As expected, the backscatter profiles are less smooth than the extinction profiles.

The summary of the statistics (based on entire dataset input) over the Detection mode, Evaluation method, Raw resolution, Shots averaged and Zenith angle for each backscatter and extinction coefficient is given in Table S2 (Fig. 2, stage III in Methodology). The main features are the following. For b355 and b532 files, the detection mode the most employed is the glued signal, while for b1064 it is the analog signal. For e-files, the detection mode the most employed is the photon counting.

The most used evaluation approach for b-files is Klett-Fernald (for particles backscatter coefficient) while for e-files is Raman (for particles extinction coefficient). Please note that there are 139 cases reported for b1064 which uses as evaluation method Raman (Warsaw station). The station used a PollyXT system and the algorithm employed to calculate the particles backscatter coefficient at 1064 nm involves the extinction coefficient determined at 607 nm from the Raman signal (no reference is available). Concerning the raw resolution, most of the stations used 3.75 m for backscatter and 60 m for extinction. The zenith angle was mostly 0. The shots averaged were in general of tens thousands. This overview shows that we deal with a variety of approaches to compute the mean profile and associated uncertainty.

3 Data quality control

The first check on data quality consisted of verifying the data which passed the quality control a.t. EARLINET criteria (Fig. 2, stage II in Methodology). The description of the checks can be found on EARLINET website (https://www.earlinet.org/index.php?id=125, last access 20191126). Basically, there are two types of checks. First check consists on technical checks on the raw data submitted (applied on-fly since June 2018). Here the conformity of the file content is checked w.r.t. EARLINET file structure and quality control (QC) procedures. With on fly version, an automatic feedback is provided to the Data Originator reporting all the problems incurred for each rejected file, fostering the prompt resubmission of the data (see 20180614 report on https://www.earlinet.org/index.php?id=293, last access 20191126). Also, the files not providing the associated error are eliminated. The second check (physical), applied off-line, first verifies if the data was submitted to the right category. Then, the errors are checked for negative values occurrence. Specific checks on backscatter and extinction coefficients are performed (e.g. if larger than a threshold, if integrated quantities such as AOD are within common limits) as well as on intensive variables (where possible), such as LR. The current short description corresponds to the documents issued by 14 June 2018. Three text files are generated with the name of the files rejected by technical criteria (QC_0.0.txt) or passing technical criteria but failing physical criteria (QC_1.0), the files passing both tests being shown in the file QC_2.0. After the EARLINET QC tests, there were 2341 out of 3579 initial files (available from 14 stations) that there were passing both tests from (i.e. ~ 65 %). After adding reprocessed data/newly processed data, the total number of input files increased to 3589.

Several in house checks (most of them manually performed) were further applied for the selection of good quality data. Recall that the additional and revised profiles sent by several stations were not on-fly QC by the EARLINET procedures. Note that the additional quality checks are performed along various steps during data analysis and are briefly discussed here. The off-line QC was manually performed for each profile before applying the algorithm for the layers' detection (section 4.1). Thus, profiles showing atypical behaviour (biased towards negative values, unrealistically increasing above a certain altitude, with PDR outside the [0 1] range) were dismissed. Sometimes, there were profiles which did not show a clearly defined layer and thus they were eliminated (section 4.1.). The profiles for which no fire was found along the backtrajectory were also eliminated (section 4.3). The optical properties in a specific layer were dismissed if the number of available points were less than 90%,

and thus, no mean value was calculated (section 4.4). Values of the mean optical properties for which SNR was less than 2 were also dismissed (section 4.4). Finally, the values of the intensive parameters with SNR < 2 and outside the imposed limits are also excluded (section 4.4).

4 Data analysis

4.1 Calculation of the aerosol layers boundaries

Prior to evaluation of the layers' boundaries (Fig.2, stage IV in Methodology), a manual QC of the profiles of the optical properties was performed. Thus, a number of optical properties profiles which did not show expected realistic behaviour were eliminated. The most common behaviours of the profiles being eliminated are:

- Profiles showing a large range of negative values due to a systematic error resulting in a bias towards negative values.
- Profiles displaying an increase above a certain altitude, most probably due to a non-accurate calibration region.
- Depolarization profiles displaying values outside [0, 1] region.
- To noisy profiles (SNR < 1).

The algorithm to describe the layers boundaries is described in Section 2 of the Supplement.

A few examples of boundaries estimation of the (smoke) pollution layers are shown in Fig. S1 in Supplement. All the optical properties profiles are shown (left - particle backscatter coefficients, middle - particle extinction coefficients, left - particle linear depolarization ratio), in order to get a glimpse of how the layers look like for all profiles. First plots (a-e) show examples of automatic selection of the layers based on developed algorithm. First four are based on b1064 signal, while the fifth is based on b355 signal. The next plots (f-i) show examples where one or more boundaries are manually modified. The last plot is based on b532 signal. In example f) a layer automatically selected around 6500 m was dismissed (considered as not substantial) while in example g) the uppermost layer was manually added. The layers are shown by the grey areas.

Most of the layers detected are situated between 1000 and 5000 m altitude (typically above PBL). However, the minimum layer bottom was found at 257.5 m while the highest layer top was found at 19,8 km. Minimum, maximum and the mean layer thickness were 300, 6862.5 and 1337.5 m. Please note that not all the layers shown here have BB origin (as this check is not performed yet).

The optical profiles shown in Fig. S1 to illustrate the layer estimation show also various questionable patterns for different optical variables. Our quality checks were meant to eliminate profiles (or parts of profiles) where suspicions arise. The examples shown in a), b) and h) were eliminated as they were considered of non-BB origin (as discussed later). For the profiles in c), all layers have BB origin. Various QC did not allow the estimation of the IPs based on non-reliable backscatter coefficient at 355 nm for second and third layer while PDR@355 was dismissed as well. For d) case, both layers have BB origin. QC did not allow the retrieval of various IPs based on non-reliable backscatter coefficient at 355 nm. For g) case, the first layer was considered as having BB origin. QC did not allow the computation of the mean PDR in the layers. For i) case, all three layers

have BB origin. However, the QC allowed the estimation of both LR and EAE for first layer and only EAE for the second layer.

4.2 Calculation of the back-trajectories

Even if the lidar data were stored within Forest Fires category in EARLINET/ACTRIS database, in many cases there are situations when several aerosol layers are present and not all of them have a BB origin. Thus, we would like to differentiate these cases and eliminate the non-smoke layers from analysis.

Matlab and Python routines were used to automatically obtain the back-trajectories using Hysplit. The settings for Hysplit backward run are the following: input altitude (middle of the layer) is provided about sea level (a.s.l.), run time is 240 h, meteorological model is GDAS at 0.5° resolution, vertical motion is chosen as Model vertical velocity (Fig.2, stage V in Methodology). The terrain height is saved in the output txt file. For 23 cases (during August 2010, 2012 and 2013), the GDAS meteorology was not available at 0.5° resolution and thus, a manual run was performed using a resolution of 1°. For the situations when more than three layers were available for a profile, Hysplit was run more than one time. There were 75 such cases. We performed 1036 Hysplit runs for 1901 layers corresponding to 960 time stamps.

Hysplit backtrajectory is the most used tool to track the airmass origin. For the literature review on BB, 30 papers made use of Hysplit (references 5, 6, 10-12, 14-21, 23, 24, 28-44, 46 on S4 Table, Sicard et al, 2019), whereas five studies use Flexpart (refs. 2-4, 24, 26 on S2 Table). The running time used in Hysplit simulations reported in literature varies from two to ten days.

4.3 Layers identification based on back-trajectories and fires emissions

The output from the Hysplit text files and the information about fires taken from FIRM MODIS (https://firms.modaps.eosdis.nasa.gov/, last access 20191126) were used to produce the combined plots containing information about both the fires' presence and the back-trajectory (Fig. 2, stage VI in Methodology). The fires are represented on the plots as a function of their fire radiative power (FRP), using different colours and sizes (see legend). The location of the trajectory, each 24 h backwards, is shown by the number of hours. The lower plot shows the altitude a.s.l. of each back-trajectory versus time. The fires are chosen within 100 km and within ±1 h around trajectory points (backtrajectory data are available at 1 h temporal resolution). In other words, we assume that those fires are more likely to contribute to the transported smoke, recorded by the lidars. When no fire is recorded along a trajectory, we catalogue that layer as of non-biomass burning origin (dismissed from further analysis). As a consequence of this criterion, a number of 283 time stamps were considered as having layers with non-BB origin. For many cases, for the same time stamp, there were equally layers of BB origin and non-BB origin. We obtained a number of 678 Hysplit-FIRMS plots (one for each time stamp). Note that the Hysplit-FIRMS plot (corresponding to a time stamp) contains all available layers with fires detected along trajectory. Examples of such plots are given at a later stage, when discussing specific events (Figs. 12 and 14). Based on these criteria, the number of time stamps and corresponding number of layers with BB origin are shown in columns 5 and 6 in Table 3 (compare with columns 3 and 4).

It is worth mentioning that the Hysplit model does not provide the uncertainty. In order to get a possible uncertainty of an individual trajectory, a trajectory ensemble is suggested (Rolph et al., 2017). We may assume that high uncertainties in the airmass location may occur particularly over long periods of time (e.g. ten days), which in conjunction with fires location may mean a missed fire or a fire detection that was not contributing to the measurement. Drexler

5 (https://www.arl.noaa.gov/hysplit/hysplit-frequently-asked-questions-faqs/faq-hg11/, last access 20191126) mentions that the uncertainty is between 15% and 30%. On the other hand, FIRMS may miss some fires (especially during cloudy atmosphere). According to Giglio et al. (2016), the collection 6 MODIS has a smaller commission error (false alarm) as compared with Collection 5 (1.2% versus 2.4% respectively). The probability of fire detection (regionally) increased by 3% in boreal North America, while staying almost the same in regions as Europe or North Africa. We have been using fires with a confidence

10 level larger than 70%. We did not investigate the injection height based on FRP in order to estimate if the smoke of a particular fire reached indeed the altitude of the backtrajectory. We would like to emphasize that due to the satellites polar orbit, the same geographical location can be seen four times a day at the equator and more times as the latitude increases (due to orbits overlap). Thus, we may miss a certain number of fires (which burn less than few hours, between the two orbits). However, we may consider those short life fires as not significant in smoke production.

15 FIRMS database was used in several studies to identify the BB origin. However, all fires occurring over certain periods (for which the backtrajectories were calculated) are typically accounted for. Thus, there were reported fires occurring over the whole day (e.g. Nicolae et al., 2013; Stachlewska et al., 2018; Janicka et al., 2017), or several days (e.g. Mylonaki et al., 2017; Heese and Wiegner, 2008; Tesche et al., 2011). By contrast, our novel approach accounts only for those fires which were occurring around backtrajectory (100 km radius) at the time of air masses passage ($\pm$ 1 h).

20 **4.4 Calculation of the mean optical properties and intensive parameters inside the layers**

After determining the aerosol layers and their boundaries (sections 4.1 – 4.3), the mean value of the optical properties is computed for each layer (Fig. 2, stage VII in Methodology). Note that the first and last 50 m of each layer were not considered for the calculation (cf. Nicolae et al., 2018). The uncertainty for each mean value was computed following the error propagation. The mean value in the layer was calculated if there were at least 90% of the points available (the ratio between

25 the layer depth and the resolution in the layer). There were many cases for which the extinction coefficient or the linear particle depolarization ratio could not be calculated as typically their profiles have a shorter extend than that of a backscatter coefficient. A visual check is performed as well, and thus, where the profiles were suspicious, the mean values in the layers were manually set to NaN.

The signal to noise ratio (SNR) was computed as the ratio of the mean value to its uncertainty (e.g. Nicolae et al., 2018). The

30 values of each optical property for which SNR < 2 were dismissed. In Table 3 (columns 7 and 8), we show the number of the time stamps and the layers with a good SNR for optical properties. As a result of this criterium, only one time stamp was dismissed (Athens) and three layers (one for each of Athens, Bucharest and Observatory Hohenpeißenberg). Please keep in

mind that the number of dismissed optical properties in layers is larger than the number of dismissed layers, as one layer is dismissed only when all optical properties in the layers are discharged.

Figure S2 shows an example (for Athens) of the number of layers selected and the corresponding number of optical properties evaluated in each layer. For this example, we have 171 / 172 time-stamps from which we could determine 250 layers. It was just one layer for which we could not determine the optical properties due to SNR constraint (see columns 5-8 in Table 3).

An example of the mean optical properties computed in the layers (versus measurement time) is given in Fig. 5 for Bucharest station. The range of values taken by different variables is large. Another aspect is the lower number of values reported or retrieved for extinction and depolarization (these features are specific to all stations). The figures for all the stations will be shown in the companion paper. Systematic measurements and more intensive parameters were available for Athens, Bucharest, Thessaloniki and Warsaw stations while over the entire analysed period, systematic measurements were provided by Athens, Bucharest and Thessaloniki.

Once the mean optical properties were calculated, the intensive parameters were determined, where possible (Fig. 2, stage VIII in Methodology). All the IPs have SNR > 2. The imposed limits (data filtering) for IPs are the following: LR@355 = [20 150] sr, LR@532 = [20 150] sr, EAE = [-1 3], BAE@355/532 = [-1 3], BAE@532/1064 = [-1 3], PDR@355 = [0 0.3] and PDR@532 = [0 0.3] (following closely Burton et al., 2012; Nicolae et al., 2018). The figures representing the intensive parameters for all the stations will be shown in the companion paper. The collected records about intensive parameters found in literature (46 reference values from 39 cited papers) are shown in Fig. S3 and Table S3 in supplement. Table S4 shows the cited references. The extreme values for LR@355 are 21 sr (Müller et al. 2005) and 130 sr (Tesche et al. 2011) while for LR@532 are 26 sr (Müller at al. 2005) and 147 sr (Mariano et al. 2010). For EAE we have 0 (Müller at al. 2005) and 2.4 (Giannakaki et al., 2016). For BAE@355/532 we found 0.35 (Tesche et al., 2011) and 2.8 (Giannakaki et al., 2010) while for BAE@532/1064 we found 0.29 (Tesche et al., 2011) and 2.85 (Gross et al., 2013). For PDR@355 the extreme values were 1% (Janicka et al., 2019) and 31% (Vaughan et al., 2018) while for PDR@532 the extreme values were 0.3% (Stachlewska et al., 2018) and 20% (Hu et al., 2018). Figure S4 shows an example (for Athens) of the number of layers selected and the corresponding number of intensive parameters retrieved. One can compare with Fig. S2 over the difference between available number of optical properties and the final number of intensive parameters. An example for the intensive parameters is shown in Fig. 6 for Warsaw station. For Warsaw site, all profiles for extinction and backscatter were calculated using the classical Raman evaluation (Ansmann et al., 1992). The lines in magenta and cyan represent the minimum and the maximum as reported in literature.

In general, LR, EAE, both BAE and PDR are spread well within the reported limits, although the latter is rather in its lower range. Lower PDR may be explained as due to specific location of Warsaw being much closer to BB sources in Ukraine, Belarus and Russia than other EU countries, and therefore much more exposed to faster and more direct BB transport to this site. In such case, differences of properties with respect to other measurement sites in Europe are expected and revealed within our work. Moreover, Warsaw is much less exposed to mineral dust intrusions, in comparison with many EU sites at which the BB measurements are more likely to be affected by slight contamination of dust, e.g. Spanish sites. On the other hand, Warsaw is an urban site, thus for BB observed in layers at lower ranges a slight contamination of the BB with the local urban pollution

is possible. For higher layers, industrial pollution form Silesia region in Poland (Stachlewska et al. 2018) or Ruhr region in Germany (e.g. case here on 20160704 at 07:30) is possible. This can explain a few values outside the literature range observed in Warsaw, i.e. for EAE (2 values), for both BAE (33 values) and for PDR355 (8 values). The extreme values for EAE observed on 20160704 07:30 and 20170619 20:30 (2.6 and 2.8 respectively along with $CR_{LR} < 1$) correspond to fires in Germany / Belgium and United Kingdom respectively. The fires occurred in less than 40 h before smoke measurements and thus we can consider the smoke relatively fresh.

The mean, median, minimum and maximum values taken by the intensive parameters for all the stations providing at least one parameter (all stations but Sofia) are discussed in the companion paper. The final number of selected layers and the corresponding time stamps are shown in Table 3, columns 9 and 10. The range of values taken by a specific parameter is large. The number of outliers dismissed based on predefined ranges of acceptable values for each intensive parameter is small (3.7% per total). For each IP we have the following numbers: 8/305 (2.6%) for LR@355nm, 8/253 (3.2%) for LR@532nm, 18/243 (7.4%) for EAE, 39/642 (6.1%) for BAE355/532, 21/706 (3%) for BAE532/1064, 0/132 (0%) for PDR@355 and 0/242 (0%) for PDR@532. Please note that a thorough investigation of the outliers was not performed. Conversely, we focused on a semi-automatic evaluation by applying different criteria for outliers' elimination (filtering). One can draw three possible reasons for the presence of outliers: a) the smoothing applied on the Raman signals, which induces a very smooth profile for extinction coefficients, b) a shift of the profile towards higher altitude (most probably due to non-accurate calibration range) and, c) a slight difference between the peaks of the backscatter versus extinction or a small difference in the slope (most probably related with the smoothing of the Raman signals). All values of the intensive parameters will be shown in the companion paper. After outliers' elimination we start the data (IP) interpretation.

5 Data interpretation

We have focused on several directions to investigate and interpret the measurements by means of the intensive parameters. Here we show examples for each direction, while the comprehensive analysis will be performed in the following paper.

The following results were found in literature regarding the IPs values versus travel distance (time travel):

- effective radius: According to Müller at al. (2005, 2007, 2016), the effective radius of the BB particles increases with time (distance), most probably due to coagulation and aggregation of the particles (Reid et al., 2005). Measurements and retrieved values of effective radius of BB particles showed that the fresh particles have a mean radius around 150 nm, while the aged particles have a radius around 300 - 400 nm (Müller et al., 2007). Particle size depends on many factors such as the combustion type in fires (in-flame or smouldering) or type of vegetation while the smoke aerosol undergoes different physical-chemical processes in atmosphere. (Reid and Hobbs, 1998). While the heaviest particles will sediment during transport, the particles reaching measurement site are still larger than the emitted ones.

- EAE decreases with time (distance) (e.g. Müller at al., 2005, 2007, 2016)

- Lidar ratio: It was shown that for aged BB particles (big travel time) LR@532 > LR@355 (e.g. Wandinger et al., 2002, Murayama et al., 2004; Müller et al., 2005; Sugimoto et al., 2010, Nicolae et al., 2013). Thus, the colour ratio (CR) for LR, i.e. $CR_{LR} > 1$ for aged smoke (LR@532 increases more with aging smoke)
- PDR@532 decreases with time (Nisantzi et al., 2014).

5 The general correlations between IPs were shown:

- In general, LR532 increases with increasing LR355 (e.g. Nicolae et al., 2018; Janicka and Stachlewska, 2019).
- In general, PDR532 increases with increasing PDR355 (e.g. Stachlewska et al., 2018; Janicka and Stachlewska, 2019).

The following differences between smoke properties in different regions were reported:

- smoke particles from Brazil absorb more and scatter less the solar radiation as compared with smoke in North
10   America (Reid and Hobbs, 1998).
- differences between North America and Siberian smoke over Germany: particles from North America showed smaller size and higher EAE (Müller et al., 2005)

The following studies showed different behaviour of BAE versus EAE:

- Veselovskii et al. (2015) tried to relate BAE with EAE and showed that while EAE depends mainly on particle size,
15   BAE depends both on particle size and refractive index, being very sensitive to the latter (see their Fig. 20). On one hand, when the real part of the refractive index increases, the imaginary part is constant, the effective radius for coarse mode and the high fine mode fraction are constant while the effective radius of the fine particles decreases, BAE@532/1064 decreases (BAE@355/532 increases) with increasing EAE (over ~ 0.5 - 1.5 region) (see their Figs. 20 and 22). On the other hand, for a constant refractive index, a fixed effective radius for coarse mode and an increase
20   of the fine mode fraction, both BAE increase with increasing EAE (over ~ 0.5 - 1.5 region). This is visible for different effective radius for the fine particles (see their Fig. 19).
- Su et al. (2008) showed few scenarios for the relationship between BAE@532/1064 and EAE@553/855, which depend on RH (relative humidity) and the contribution of the fine mode fraction. For a high fine mode fraction in the smoke composition and RH < 85%, BAE@532/1064 increases with increasing EAE@553/855 (EAE > 2). For a
25   dominant coarse mode and RH < 85% BAE@532/1064 decreases with increasing EAE@553/855 (EAE < 0.5) (see their Fig. 8). For RH > 85 there is no clear pattern of BAE with respect to EAE (BAE increasing and then decreasing with increasing EAE).

Concluding the findings so far, we expect $CR_{LR} > 1$, small values of EAE (< 1.4) and smaller PDR@532 for aged smoke (especially for long range transport from North America). The studies showed no straightforward pattern for CR_{BAE} and CR_{PDR}
30 evolution with time. Studies by Veselovskii et al. (2015) and Su et al. (2008) may help interpreting the scatter plots between BAE and EAE. We did not investigate the RH field though.

5.1 Fire events seen by several stations. Example for Bucharest and Thessaloniki stations.

Over the ten years period, we have found five events (with retrieved intensive parameters) recorded at two stations which have the same smoke origin (Fig. 2, stage IX in Methodology). The events belong to local fires in Europe. Two events are seen by Athens and Thessaloniki stations (2014 and 2017) and one event is seen by Bucharest and Thessaloniki (2014), Minsk and Warsaw (2015), Athens and Observatory Hohenpeißenberg (2016) stations. Unfortunately, we did not obtain common intensive parameters for all events. Only two events had one common intensive parameter (BAE@532/1064). One event is shortly presented below, emphasizing the methodology.

The event was recorded during 20140909-20140910 at Bucharest and Thessaloniki stations (see time frame and altitude on Fig. 7 a). The backtrajectories revealed the same fire origin. Figure 8 shows the air backtrajectories for the layers detected at the two stations. The fires' location (within 100 km and +/- 1h) is shown as well. The colour and size of the fires correspond to their FRP (see legend). The Hysplit back trajectories in Fig. 8 show that the main smoke source was the fires over eastern Ukraine. Figure 8 a) (Bucharest station) indicates as second smoke source from the fires over eastern Romania (~ 72 h back) and North Ukraine (~ 144 h back). The smoke particles were transported from Ukraine to Bucharest station over a period of five days descending from approximately 4200 m on 2014/09/04 to 3000 m on 2014/09/09. For the Thessaloniki station the main source was the fires over southwestern Ukraine (Fig. 8 b). The smoke particles were identified on 2014/09/10 around 2100 m. Note that for Thessaloniki case, the 'common fire' was detected at ~ 3500 m. In Fig. 7 a-b we show the location of the fires which contributes to all measurements on 9 - 10th of September and their histogram occurrence on 1x1 grids. From such histograms we pick the grids where we have fires contributing to both stations. See grid which has both colours, between 36 and 37 East and 47 and 48 North (also marked by a square on the upper plot). The location (longitude and latitude) of these fires from the "common" grid are shown versus fires occurrence time as well as versus measurement time at each station (Fig. 7 c-d). The common fire revealed through the back-trajectory occurred at 10:39 on 4th of September (latitude = 47.537, longitude = 36.275) and it was recorded at ~ 11:00 on 9th at Bucharest and at ~11:00 on 10st of September at Thessaloniki (lower layer left). Two layers were detected at Thessaloniki station (where one corresponds to the common fire) and one in Bucharest station. We analysed if there were other fires contributing to the same measurement as in general, along backtrajectories, we encounter many fires at several locations. In the case of Bucharest there were 24 other fires (detected 46 times) contributing, located in Ukraine as well as in North-East Romania (occurring during 3rd, 4th and 6th of September). For Thessaloniki we counted 35 other fires (detected 66 times), located at East Ukraine (occurring during 2nd and 4th of September). See Fig. 8 b). Please note that function of the air trajectory, one fire can be seen more than once (e.g. during a cyclone or anticyclone) or due to slow air motion (spatial-temporal stationarity over the 100 km area and 1 hour).

We can conclude that we have a mixed smoke recorded in both Bucharest and Thessaloniki. However, the mixtures are different for Bucharest and Thessaloniki stations.

Figure 7 right plots shows the layers location (marked by a square in front of the middle of the layer) and intensive parameters for the layers. Note that the retrieval of the intensive parameters was not possible at all times for all layers. The two layers detected in Bucharest and Thessaloniki were located at 3061 m and 2123 m respectively.

For the measurements at Bucharest (20140909 10:45 UTC) and Thessaloniki (20140910 11:15 UTC) locations, which correspond to the common fire in East Ukraine (47.537 North, 36.275 East) on 20140904 10:39 (as well as other additional fires for each station), there were the following intensive parameters calculated.

Bucharest, 09/09 10:45, 3061 m: BAE@355/532 = 1.82 ± 0.0001, BAE@532/1064 = 1.32 ± 0.0001, PDR@532 = $6.8 \pm 0.03\%$.

Thessaloniki, 10/09 11:15, 2123 m: BAE@532/1064 = 0.51 ± 0.02.

The values for BAE@532/1064 were 1.32 and 0.51 for Bucharest and Thessaloniki respectively. The difference between the two values may be explained by the different mixture of smoke (originating from different fires). All the contributing fires for Thessaloniki besides the common fire are located in East Ukraine while the contributing fires for Bucharest are located in East Romania, East Ukraine and North-East Ukraine. According to backtrajectories for Thessaloniki, the other fires contributing to measurements are further located in time than the common fire. Lower value for BAE@532/1064 in Thessaloniki reveals a relatively larger contribution of the big particles to backscatter. On the contrary, in Bucharest the smallest particles show the highest backscatter coefficient.

For the other event with common IP for the same source (20170529-20170602), the smoke was labelled as of 'single fire' as no other fires were identified along the backtrajectory. The common fire occurred on 26th of May at midnight in Ukraine (48.171 North, 30.622 East) and it was recorded in Thessaloniki and Athens on 29/05 and 31/05 respectively. BAE@532/1064 value in Thessaloniki was less than half of that in Athens, while BAE@355/532 was larger for Thessaloniki. High BAE corresponds to higher backscatter at smaller wavelengths, which indicates a higher number of small particles. The values in Thessaloniki correspond to a higher number of small size particles (at 355 nm) and with a higher proportion of large particles (at 1064 nm) compared with Athens. CR_{BAE} (colour ratio of the backscatter Ångström exponents) increases from Thessaloniki to Athens (0.22 to 0.78, respectively), which suggests an increase with travel distance (time). As CR_{LR} (colour ratio of the lidar ratios) and EAE (extinction Ångström exponent) were not available to characterize the smoke in terms of age, we classified the smoke as aged based on the duration of the travel time.

Overall, we conclude that the number of common events as well as the number of the common IPs is limited and, thus, no thorough examination of these events is possible. The most important feature of this analysis is that it enables us to quantify the smoke as of 'single fire' or 'mixed' and hence explain various IP values. This kind of analysis can be successfully applied in the future, when more data become available.

5.2 Long-range transport from North America. Example for Athens station.

We have identified a number of 24 events with long range transport from North America (Fig. 2, stage X in Methodology) for which we have at least one intensive parameter retrieved. The events are reported in 2009 and over 2012-2017 period. As in previous section, for each event, we plot the Hysplit trajectory and the location of the fires along it, the histogram with the

number of fires in each 1° x 1° geographical grid, the geographical location (longitude and latitude) versus occurrence time of the fires and versus measurement time at the stations(s). Last, we show the layers altitude (error bar signifies the layer' thickness) and the intensive parameters in the layers.

We have identified eight measurements periods (events) when the smoke is arriving solely from North America ("pure N

5   America"). The other cases represent measurements of mixed smoke, where the smoke is coming from both North America and local fires (mostly European) ("mixed"). In two cases we have mixed smoke from North America and North Africa or Middle East. Usually, during measurements period, there were also recordings of BB with other origin (Europe). For each layer we check if the smoke is coming from a single fire or more fires (count their number), and quantify the locations. We have one event when there were measurements taken at three stations and one event with measurements taken at two stations.

10   All others cases represent measurements recorded by a single station. As the number of intensive parameters determined for each station varies considerably, in general we cannot compare directly all IPs for the same event. As a consequence, a statistical analysis will be performed over entire set of parameters. Below we show one example of long range transport recorded in Athens which provided more measurements and more intensive parameters. The event recorded over three days in July 2013 will be discussed in the companion paper.

15   **5.2.1 Smoke event recorded on 20170713**

During this day we recorded several measurements in Athens (see time and layers altitude in Fig. 9 e). According to backtrajectory and the fire occurrence along it, we determined three layers of North America smoke origin (their location is marked with a black square at the left). As seen in the backtrajectory (Fig. 10), the fires' location in North America is different for the three layers (different source).

20   The first layer of smoke origin was detected at 14:27 at ~ 2900 m altitude. Our calculations show that there were different fires contributing to this measurement. Thus, we have four fires (counted eight times) from which only one fire (counted twice) was of North America origin. This can be seen in Fig. 9 d) as well as in the backtrajectory in Fig. 10 a) where we see the fires detected in the first two days backwards (local fires, in Greece and Italy) as well as the fire found after almost eight days back in North America. The fire from North America occurred at 20:14 on 5 July in the longitude by latitude grid [-111 -110] x [50

25   51]. The location of the fire is shown by blue arrows in Fig. 9 (a-e).

The second layer was measured at 18:49 at 2102 m altitude. 12 fires were found (counted 19 times) to have contributed to the measurement from which one fire (detected once) was of North America origin. See Figs. 9 and 10. The fire from North America was detected at 19:16 on 6 July in the longitude by latitude grid [-84 -83] x [42 43]. The location of the fire is shown by green arrows in Fig. 9 (a-e).

30   The last layer was measured at 19:34 at 3872 m altitude. We found two fires (each detected once) of North America origin. The North America fires occurred at 09:47 on 6 July and were located in the longitude by latitude grid [-109 -108] x [47 48]. The location of the fire is shown by magenta arrows in Fig. 9 (a-e). The trajectory layer in Fig. 10 c) is the higher one (light blue). The local fires observed in the plot are detected by the first layer (at 2012 m).

Thus, we can label the first two layers as 'mixed' smoke while the third one as "pure N America" smoke. We observe that for the mixed fires, the contribution from local fires is much larger as we have 6/8 and 18/19 fires counts in Europe for first and second layer respectively. The intensive parameters for the three layers are:

- layer 1 @ 14:27 2942 m, mixed: BAE@355/532 = 1.42 ± 0.03, BAE@532/1064 = 1.39 ± 0.01
- layer 2 @ 18:49 2102 m, mixed: LR@355 = 48.4 ± 1.04 sr, LR@532 = 33.7 ± 3.5 sr, BAE@355/532 = 1.72 ±0.05, BAE@532/1064 = 1.85 ± 0.02
- layer 3 @ 19:34 3872 m, North America: LR@355 = 58.5 ± 2.63 sr, LR@532 = 67.2 ± 4.79 sr, EAE = 1.02 ± 0.14, BAE@355/532 = 1.36 ± 0.15, BAE@532/1064 = 2.11 ± 0.08

We observe that BAE@355/532 has the smallest value for 'pure N America' while BAE@532/1064 is the highest. Relatively small value for EAE for the third layer corresponds to bigger (coarse mode) particles, associated with aged smoke. On the other hand, values of LR (LR@355 < LR@532) suggest the presence of aged aerosol as well. The layer's altitude is the highest of all three. The values of LR for the second layer (LR@355 > LR@532) suggest fresh aerosol (of smaller size, fine mode particles). This can be supported by the contribution of the local fires (detected a few hours back). If we compare the colour ratio (CR) of BAE we obtain values of 0.98, 1.08 and 1.55 for the three layers. CR_{LR} for the second and third layers were 0.7 (fresh smoke) and 1.15 (aged aerosol). Based on all reported values in literature (Table S3), we found the following values for CR_{LR} and EAE for fresh, aged and North America case (particular case of aged smoke). CR_{LR} was 0.88, 1.08 and 1.23 while EAE was 1.47, 1.2, 0.95. CR_{BAE} had the values 0.76, 0.98 and 0.98. We may speculate that CR_{BAE} may increase with time (distance).

Please note that the time difference (an hour at most) between right plots and bottom left plot comes from the fact that Hysplit back-trajectories start at sharp hours. For example, for the measurement at 18:49 the starting point on back-trajectory is 18. Statistics over LRT from North America will be shown in the companion paper. We encountered 168 measurements over the 24 periods (over 2009 – 2017 period). From these measurements, 77 have a North America origin while the other 91 have a different BB origin (local). The LRT events from North America is analysed differencing between 'pure N America' fires (sensed smoke comes solely from North America) and 'mixed' fires (measured smoke is a mixture of North America smoke and local/European smoke).

5.3 Analysis over geographical regions

5.3.1. Geographical regions

Taking into account the position of the 14 stations (Fig. 1), four geographical regions are chosen as follows: South-East Europe (Potenza, Athens, Thessaloniki, Sofia, Bucharest), South-West Europe (Granada, Barcelona and Evora), North-East Europe (Belsk, Warsaw, Minsk) and Central Europe (Cabauw, Leipzig, Observatory Hohenpeißenberg). This corresponds to stage XI from Methodology (Fig. 2). The statistics over the measurements for each individual station show the following. Belsk and Cabauw stations focused on LRT smoke from North America with ~ 99 % of the measurements (1149 / 1159 and 1001 / 1013 North America fires / total number of fires). Leipzig measured 86.2 % (156/181 North America fires / total number of fires)

LRT smoke. The others have measured mostly in Europe as follows (# fires in EU / # total fires): Athens 81.7 % (1657 / 2028), Barcelona 91 % (172 / 189), Bucharest 93.82 % (1883 / 2007 cases), Granada 68.32 % (1643 / 2405), Minsk 98.04 % (1051 / 1072), Observatory Hohenpeissenberg 74.84 % (116 / 286), Potenza 100% (142 / 142), Sofia 100% (16 / 16), Thessaloniki 87.38 % (1807 / 2068), Warsaw 85.4 % (5036 / 5897). For Evora station we encountered 44.39 % (99 /223) fires in North America and 48.43 % in Europe (108 / 223). Per regions, we have detected fires from North America as 4.5 % (282 / 6261), 8.56 % (241 / 2817), 87.55 % (1181 /1349) and 24.13 % (1961 / 8128) for South-East, South-West, Central and North-East region respectively. The number of fires detected in Europe was the following: 87.93 % (5505 / 6261), 68.26 % (1923 / 2817), 11.34 % (153 / 1349) and 75.01 % (6097 / 8128) for South-East, South-West, Central, and North-East region respectively.

In Fig. 11 we show the location of the fires seen by the stations from South-East region. Note that the grid size is 1 x 1 degree latitude and longitude. The first remarks for each cluster are the following. As mentioned, for South-East region, we have a number of 282 fires located in North America (4.5 %) and 5979 elsewhere (total of 6261 fires), most of them being located in East Europe (5505). The longitude region with most of the fires is between 20 East and 30 East while the latitude region is 37 - 46 North. This corresponds to the Balkan region, covering parts of Romania, Bulgaria, Macedonia, Greece. Most of the measurements were taken at Bucharest, Athens and Thessaloniki.

5.3.2. Intensive parameters by geographical regions

A statistical investigation over the intensive parameters was performed for each geographical region in order to identify the main features. The analysis was performed by separating the intensive parameters based on their continental source origin. The following origin were considered: Europe (EU), Africa (AF), Asia (AS), North America (NA) and a combination of any of them (EUAF=EU+AF, EUAS=EU+AS, EUNA=EU+NA, EUAFAS=EU+AF+AS, etc). The statistics was performed over all available cases. Here we present the results for South-East region (for consistency). The other three regions will be discussed in the following paper as well as the overall assessment.

Figure 12 shows the scatter plots for some of the combinations between two IPs. Note that the number of pair points available for each combination is different. The following features are revealed (the mean values are discussed). In average, there is a linear correlation between the two LR if we dismiss the value for NA. The mean LR@532 is slightly larger than the mean LR@355 (CR_{LR} > 1) which suggests the presence of aged aerosol in general. For EU source region, PDR@532 is below 7% (except three values) which corresponds to a low depolarization. For EAE features, we observe low value for EUAF source region (~ 0.65), while the other three source regions (EU, EUAS and EUNA) have similar values, around 1.5 (based on EAE vs BAE@355/532 plot). Large value of EAE (~1.5) suggests smaller size aerosol, specific to fresh smoke particles. The values for the EUAS and EUNA mixtures suggest that the contribution of EU fires to the mixture is large. On the other hand, the small value for EUAF region (corresponding to bigger particles) may be due to the major contribution of the AF region (as the value is not close to the EU value) corresponding to relatively big smoke travel time. Further analysis will be performed in part II (Adam et al., 2020), where both CR_{LR} and EAE will be accounted for when interpreting the smoke (fresh versus aged) based on the same measurements. The scatter plots between EAE and the two BAE are opposite. While EAE increases with

increasing BAE@355/532, EAE decreases with increasing BAE@532/1064 (also reported by Veselovskii et al., 2015, in special conditions). A general decreasing trend of EAE versus LR@532 is observed. The comparison between the two BAE, based on source origin, did not show a specific relationship. Based on standard deviation, we observe a big overlap among the values for all source regions. In conclusion, based on the current dataset, there is no clear separation between sources for South-East region but for average values, specific trends are observed for some scatter plots.

As the dataset is limited, we cannot conclude at this stage about the existence of a clear feature with respect to the continental source origin. As mentioned on literature, EAE exhibits a decay versus time (while smoke effective radius increases with time). In the present example, for South-East region we obtained quite large values for mixtures which suggest a large contribution of the European fires. On the other hand, for EUAF mixture, the value was quite small, which suggests a relatively big travel time. Medium absorption as shown by LR for both 355 nm and 532 nm (~ 50 sr) was observed for EUAS, EUAF and EU regions. The smallest / largest value for LR@355 / LR@532 was observed for NA region, suggesting less / more absorption for small / medium size particles. The LR values for EUNA source are close to those for EU source which suggest once more the big contribution of EU fires to the mixture.

The complete analysis of IPs and CRs will be performed in the companion paper. Based on a larger IP dataset, we expect that the complete set of CR for LR, BAE and PDR, along with EAE will better characterize the measured aerosol. The statistics over all four regions will eventually bring more insights about the IPs and CRs trends versus continental source origin.

6. Summary and conclusions

The current study focuses on developing a methodology to analyse large amounts of biomass burning measurements by lidars. The current lidar dataset is from EARLINET database, Forest Fire category.

First, we would like to mention the current challenges when performing such analyses. Besides tackling with high amount of data, there is different data processing used by stations (using either their own algorithms or more recently SCC). The data still need quality control checks and the best way is the manual check. For more accurate results, the algorithm developed to detect the pollution layers needs a manual check as well, and thus, the boundaries are manually corrected where required. If the investigation is based on intensive parameter results, one has to bear in mind that their number is a limited subsample of the initial data set. This is due to two main factors. First, not all of the lidar systems provide the complete sets of the intensive parameters. Second, the quality checks remove a large amount of data (more than half of the initial dataset).

The possible sources of uncertainty during such an analysis may be the following. A small change in the input to Hysplit may give a different output (e.g. use of altitude a.g.l. versus a.s.l., use of GDAS0.5 versus GDAS1). See for example Su et al. (2015). Various algorithms employed to estimate the layer geometry may give slightly different values over the mean values in the layers. Thus, the direct comparison with other reports over the same event should be carefully performed. Uncertainties in Hysplit backtrajectories as well as in FIRMS database are not considered. Least but not last, the imperfect data quality control (including the present methodology) may contribute as well.

The current methodology (Fig. 2) describes various criteria involved in order to assure a quality-controlled dataset and the steps preceding the computation of the mean intensive parameters in the layers. The algorithm to select the layers of BB origin, employing the information from both Hysplit and FIRMS within the current criteria (*novelty*) allowed us to identify only the fires which most probably contributed to the measurements. Further, we were able to determine if the smoke measurements

5   were originating from a single fire source or from many fire sources by identifying all the fires along backtrajectory (*novelty*). Therefore, we could quantify the measurements as having a 'single fire' source or 'mixed fires' sources, having a North America origin ('pure N America') or a 'mixed origin' (North America and Europe). The number of fires occurring along a backtrajectory as well as the number of fires' counts was calculated (*novelty*). Further, we proposed few directions for BB study by means of the intensive parameters.

10   The first direction is to study the same BB event through the measurements taken at several stations. For the current dataset we found five events as observed by two stations. The common fire source is precisely identified while the number of other fires contributing to the smoke measurement was quantified (*novelty*). For two cases, BAE@532/1064 could be compared. In the current example, the measurements represent smoke originating from several fires (mixed smoke). We found that BAE@532/1064 had a smaller value for the station which recorded smoke transported for a longer time for the common fire.

15   The differences between the two values can be attributed to the fact that the mixed smoke measured in the two locations originated from different fires (besides the common fire).

The second direction was the study of LRT of smoke from North America. 24 events were available. We have identified that the LRT from North America can be of North America origin only or a mixture of both North America and local fires (*novelty*). Note that we did not consider other potential aerosol sources to the smoke mixture. The quantification of the smoke as mixed

20   explained various values for IPs in cases where the values were closer either to North America type or to European type. In the example shown here, we identified three layers where two were labelled as 'mixed' and one as 'pure N America'. For the later, EAE (1) and $CR_{LR} > 1$ suggested aged smoke, larger particle size. For one of the mixed smokes we had $CR_{LR} < 1$ suggesting fresh aerosol (no EAE available). This can be explained by the contribution of the local fires detected. For the 'pure N America' smoke layer we obtained the smallest BAE@355/532 and the larger value for BAE@532/1064. CR_{BAE} provides

25   the biggest value for 'pure N America' smoke.

The third direction was based on consideration of four geographical regions (South-East, South-West, North-East and Central Europe), analysed individually. Histograms of fires' locations as detected by individual stations (through HYSPLIT and FIRMS) are presented (*novelty*), showing the predominant type of measurements taken by each station (local versus LRT).

Statistics over intensive parameters were performed in the following manner. For each geographical region, the scatter plots

30   between various IPs were drawn and the mean values for each IP were computed function of continental BB origin. In the present example for South-East region, we could observe the following (Fig. 12). LR@532 versus PDR@532 for EU source region suggest the presence of medium size particles with low depolarization and relatively high absorption. A linear dependence of LR@532 versus LR@355 was observed (as reported in literature) with colour ratios larger than one, suggesting the presence of aged smoke in average. EAE had the smaller value for EUAF source region, suggesting relatively high particle

size (aged smoke). EAE values for EU, EUAS and EUNA had similar values, relatively large (~1.5) suggesting relatively small particle size (rather fresh smoke). High EAE values for mixtures (EUAS and EUNA) can be explained by the large contribution of EU fires to the mixtures. Based on South-East results, we found the following trends. EAE increases with increasing BAE@355/532 and decreases with increasing BAE@532/1064. A slight EAE decrease with increasing LR@532 was observed as well. The relationship between the two BAE does not show a clear feature. Note that for the current dataset, the standard deviation is large for all the means and there is an overlap among most of the values.

One of the important outcomes of this study is the quantification (within the existing assumptions) of the fires which contribute to the smoke measurement. As observed, in most of the cases, the smoke measured have several fire sources ('mixed smoke'). Note that in the part II paper more discussion is given on the statistical analysis based on the four geographical regions considered where the data are interpreted function of continental source origin.

The current paper presents first ever such widely applied approach for the analyses of the BB optical properties derived form the lidar measurements. Full methodology was developed, from QC via selection of the layers to obtaining mean optical properties with uncertainty analyses. Comparison of optical properties of BB aerosol depending on its origin was possible and revealed interesting and distinctly different results for various regions of Europe. In general, we measure mixed smoke and thus, we do not recommend to associate precisely a measurement with a specific source without a careful check over other possible sources. Based on favourable meteorological conditions and the fires' sources locations, the analysed stations or regions provides different measurements. Thus, Central Europe mostly measures LRT smoke from North America, South-West Europe mostly measure smoke originating from Iberic Peninsula and North Africa, while South-East and North-East Europe mostly measure smoke originating in East Europe. The current study provides a reference for further research, including algorithm testing and aerosol typing. As for the limitations, we show that although enormous efforts are undertaken on the EARLINET-ACTRIS regular long-term observations, still the availability of the optical properties profiles in the database is limited, which is mainly due to the fact that for many stations still manual evaluation of profiles is needed. Therefore, we see a strong need for both the further development of the SCC data evaluation automated chain as well as the use of this evaluation chain by as many as possible stations for processing of lidar data. The present methodology can reveal more insights regarding the specifics of different measurement regions versus various sources when a larger number of IPs is available.

In order to increase the number of input data and further the number of intensive parameters as well as the BB study through measurements taken at several locations, we recommend to have coordinated measurements in EARLINET during BB events. On the other hand, there are numerous events discussed in literature while the data is not reported in EARLINET yet. The data should be treated in the same manner (use SCC). The data quality control through SCC output still needs improvement. We acknowledge the large effort put to improve SCC which is envisaged to be fully operational in a few years. In the data subset under investigation, we are not confident with some of the provided uncertainties of the optical properties' products stored in the EARLINET data base, as in many cases it was not possible to determine the final spatial resolution nor the method used for smoothing and error calculation. Regarding the EARLINET QC over the existing datasets we recommend the improvement of the quality assurance over the uncertainties' estimates. Also, taking into account the various methods used by the in-house

algorithms to estimate the uncertainties for the profiles of the optical properties, we advise to use the standard deviation as associated uncertainty when computing the mean values for the optical properties in the layers.

Future investigations envisage several important features to be accounted for. A more detailed analysis on grouping the sources' location using cluster analysis is envisaged, where a larger number of clusters should be chosen in order to pack more homogeneous regions with similar vegetation type. Thus, a more accurate correlation between the source type and the measurements is envisioned. The time travel should be in some way considered. FRP will be considered to estimate the injection height and thus have more confidence that the smoke reaches the backtrajectory' altitude. The biggest challenge remains the quantification of the contribution of different fires.

Author contributions. MA developed the methodology and wrote the paper. All authors but MA are PIs of the stations contributing the most to EARLINET Forest Fire category. All authors contributed to revisions of the paper.

Competing interests. The authors declare that they have no conflict of interest.

Special issue statement. This article is part of the special issue "EARLINET, the European Aerosol Research Lidar Network. It is not associated with a conference.

Acknowledgements: *We acknowledge the use of data and imagery from LANCE FIRMS operated by the NASA/GSFC/Earth Science Data and Information System (ESDIS) with funding provided by NASA/HQ. The authors gratefully acknowledge the NOAA Air Resources Laboratory (ARL) for the provision of the HYSPLIT transport and dispersion model and/or READY website (http://www.ready.noaa.gov) used in this publication. The authors acknowledge the EARLINET-ACTRIS community for provision of the aerosol lidar profiles used in this study, in particular the PIs of all stations that provided data to the Forest Fire category in the EARLINET-ACTRIS database: A. Papayannis (Athens), A. Comeron (Barcelona), A. Pietruczuk (Belsk), D. Nicolae (Bucharest), A. Apituley (Cabauw), D. Bartoli (Evora), L. Alados-Arboledas (Granada), U. Wandinger (Leipzig), A. Chaikovsky (Minsk), I. Mattis (Observatory Hohenpeissenberg), A. Amodeo (Potenza), D. V. Stoyanov (Sofia), D. Balis (Thessaloniki), and I. S. Stachlewska (Warsaw). We acknowledge Wojciech Kumala, Lucja Janicka, Dominika Szczepanik, Krzysztof Markowicz, and Rafal Fortuna (University of Warsaw) for their technical support at the ACTRIS site in Warsaw, Livio Belegante, Cristi Radu, Dragos Ene and Alexandru Dandocsi (INOE 2000) for their technical support at the ACTRIS site in Magurele.*

Funding: *The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 262254, as well as the H2020 ACTRIS-2 grant n° 654109. It was also supported with following national funding: the Romanian National contracts 18N/08.02.2019 and 19PFE/17.10.2018 as well as with the European Space Agency (ESA-ESTEC) funding: The Technical assistance for Polish Radar and Lidar Mobile Observation System (POLIMOS 4000119961/16/NL/FF/mg). The research was partially funded by the European Regional Development*

Fund through the Competitiveness 613 Operational Programme 2014-2020, POC-A.1-A.1.1.1- F- 2015, project Research Centre for Environment and Earth 614 Observation CEO-Terra, SMIS code 108109, contract No. 152/2016.

Data access: *The aerosol lidar profiles used in this study are available upon registration from EARLINET webpage* https://data.earlinet.org/earlinet/login.zul*, last access: 20191126). The FIRMS data used in the study is available upon request*

5 *from* https://firms.modaps.eosdis.nasa.gov/*.*

References

Adam, M., Pahlow, M., Kovalev, V. A., Ondov, J. M., Parlange, M. B., and Nair, N.: Aerosol optical characterization by nephelometer and lidar: The Baltimore Supersite experiment during the Canadian forest fire smoke intrusion, J. Geophys. Res., 109, D16S02, doi:10.1029/2003JD004047, 2004.

10 Adam, M., Nicolae, D., Belegante, L., Stachlewska, I. S., Szczepanik, D., Mylonaki, M., Papanikolaou, C. A., Siomos, N. , Voudouri, K. A., Apituley, A., Alados-Arboledas, L., Bravo-Aranda, J. A., Pietruczuk, A., Chaikovski, A., Sicard, M., Muñoz-Porcar, C., Mattis, I., Papagiannopoulos, N., Mona, L., Baars, H., Wandinger, U., Bortoli, D., Grigorov, I., Peshev, Z., Antonescu, B.: Biomass burning measurements in EARLINET, ILRC29, S25-17, Hefei, China, 2019.

Adam, M., Nicolae, D., Belegante, L., Stachlewska, I. S., Janicka, L., Szczepanik, D., Mylonaki, M., Papanikolaou, C. A.,

15 Siomos, N., Voudouri, K. A., Alados-Arboledas, L., Bravo-Aranda, J. A., Apituley, A., Papagiannopoulos, N., Mona, L., Mattis, I., Chaikovsky, A., Sicard, M., Muñoz-Porcar, C., Pietruczuk, A., Bortoli, D., Baars, H., Grigorov, I., and Peshev, Z.: Biomass burning events measured by lidars in EARLINET. Part II. Results and discussions, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2020-647, in review, 2020.

Alonso-Blanco, E., Castro, A., Calvo, A. I., Pont, V., Mallet, M., Fraile, R.: Wildfire smoke plumes transport under a

20 subsidence inversion: Climate and health implications in a distant urban area, Sci. Total Environ., 619-620, 988-1002, https://doi.org/10.1016/j.scitotenv.2017.11.142, 2018.

Ansmann, A., Riebesell, M., Wandinger, U., Weitkamp, C., Voss, E., Lahmann, W. and Michaelis, W.: Combined Raman elastic-backscatter LIDAR for vertical profiling of moisture, aerosol extinction, backscatter, and LIDAR ratio, *Appl. Phys.*, B 55, 18-28, 1992.

25 Ansmann, A., Baars, H., Tesche, M., Müller, D., Althausen, D., Engelmann, R., Pauliquevis, T., and Artaxo, P.: Dust and smoke transport from Africa to South America: Lidar profiling over Cape Verde and the Amazon rainforest, Geophys. Res. Lett., 36, L11802, doi:10.1029/2009GL037923, 2009.

Benedetti, A., Reid, J. S., Knippertz, P., Marsham, J. H., Di Giuseppe, F., Rémy, S., Basart, S., Boucher, O., Brooks, I. M., Menut, L., Mona, L., Laj, P., Pappalardo, G., Wiedensohler, A., Baklanov, A., Brooks, M., Colarco, P. R., Cuevas, E., da

30 Silva, A., Escribano, J., Flemming, J., Huneeus, N., Jorba, O., Kazadzis, S., Kinne, S., Popp, T., Quinn, P. K., Sekiyama, T. T., Tanaka, T. and Terradellas, E.: Status and future of Numerical Atmospheric Aerosol Prediction with a focus on data requirements, Atmos. Chem. Phys., 18, 10615–10643, https://doi.org/10.5194/acp-18-10615-2018, 2018.

Bond, T. C., Doherty, S. J., Fahey, D. W., Forster, P. M., Berntsen, T., DeAngelo, B. J., Flanner, M. G., Ghan, S., Kärcher, B., Koch, D., Kinne, S., Kondo, Y., Quinn, P. K., Sarofim, M. C., Schultz, M. G., Schulz, M., Venkataraman, C., Zhang, H., Zhang, S., Bellouin, N., Guttikunda, S. K., Hopke, P. K., Jacobson, M. Z., Kaiser, J. W., Klimont, Z., Lohmann, U., Schwarz, J. P., Shindell, D., Storelvmo, T., Warren, S. G., and Zender, C. S.: Bounding the role of black carbon in the climate system: A scientific assessment, J. Geophys. Res., 118, 5380-5552, doi:10.1002/jgrd.50171, 2013.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, Atmos. Meas. Tech., 5, 73–98, doi:10.5194/amt-5-73-2012, 2012.

Burton, S. P., Hair, J. W., Kahnert, M., Ferrare, R. A., Hostetler, C. A., Cook, A. L., Harper, D. B., Berkoff, T. A., Seaman, S. T., Collins, J. E., Fenn, M. A., and Rogers, R. R.: Observations of the spectral dependence of linear particle depolarization ratio of aerosols using NASA Langley airborne High Spectral Resolution Lidar, Atmos. Chem. Phys., 15, 13453–13473, doi:10.5194/acp-15-13453-2015, 2015.

D'Amico, G., Amodeo, A., Baars, H., Binietoglou, I., Freudenthaler, V., Mattis, I., Wandinger, U., and Pappalardo, G.: EARLINET Single Calculus Chain – overview on methodology and strategy, Atmos. Meas. Tech., 8, 4891–4916, doi:10.5194/amt-8-4891-2015, 2015.

D'Amico, G., Amodeo, A., Mattis, I., Freudenthaler, V., and Pappalardo, G.: EARLINET Single Calculus Chain – technical – Part 1: Pre-processing of raw lidar data, Atmos. Meas. Tech., 9, 491–507, doi:10.5194/amt-9-491-2016, 2016.

Ancellet, G., Pelon, J., Totems, J., Chazette, P., Bazureau, A., Sicard, M., Di Iorio, T., Dulac, F., and Mallet, M.: Long-range transport and mixing of aerosol sources during the 2013 North American biomass burning episode: analysis of multiple lidar observations in the western Mediterranean basin, Atmos. Chem. Phys., 16, 4725–4742, doi:10.5194/acp-16-4725-2016, 2016.

Davies, D. K., Ilavajhala, S., Wong, M. M., and Justice, C. O.: Fire Information for Resource Management System: Archiving and Distributing MODIS Active Fire Data, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, VOL. 47, NO. 1, 72-79, doi:10.1109/TGRS.2008.2002076, 2009.

Fiebig, M., Petzold, A., Wandinger, U., Wendisch, M., Kiemle, C., Stifter, A., Ebert, M., Rother, T., and Leiterer, U.: Optical closure for an aerosol column: Method, accuracy, and inferable properties applied to a biomass-burning aerosol and its radiative forcing, J. Geophys. Res, VOL. 107, NO. D21, 8130, doi:10.1029/2000JD000192, 2002.

Fiebig, M., Stohl, A., Wendisch, M., Eckhardt, S., and Petzold, A.: Dependence of solar radiative forcing of forest fire aerosol on ageing and state of mixture, Atmos. Chem. Phys., 3, 881–891, 2003, www.atmos-chem-phys.org/acp/3/881/.

Giannakaki, E., D. S. Balis, V. Amiridis, and C. Zerefos, Optical properties of different aerosol types: seven years of combined Raman-elastic backscatter lidar measurements in Thessaloniki, Greece, Atmos. Meas. Tech., 3, 569–578, doi:10.5194/amt-3-569-2010, 2010.

Giannakaki, E., P. G. van Zyl, D. Müller, D. Balis, and M. Komppula, Optical and microphysical characterization of aerosol layers over South Africa by means of multi-wavelength depolarization and Raman lidar measurements, Atmos. Chem. Phys., 16, 8109–8123, doi:10.5194/acp-16-8109-2016, 2016.

Giglio, L., Schroeder, W., Justice, C. O.: The collection 6 MODIS active fire detection algorithm and fire products, *Rem. Sens. Environ.*, 178, 31-41, http://dx.doi.org/10.1016/j.rse.2016.02.054, 2016.

Gross, S., Esselborn, M., Weinzierl, B., Wirth, M., Fix, A., and Petzold, A.: Aerosol classification by airborne high spectral resolution lidar Observations, Atmos. Chem. Phys., 13, 2487–2505, doi:10.5194/acp-13-2487-2013, 2013.

Haarig, M., Ansmann, A., Baars, H., Jimenez, C., Veselovskii, I., Engelmann, R., and Althausen, D.: Depolarization and lidar ratios at 355, 532, and 1064 nm and microphysical properties of aged tropospheric and stratospheric Canadian wildfire smoke, Atmos. Chem. Phys., 18, 11847-11861, https://doi.org/10.5194/acp-18-11847-2018, 2018.

Heese, B., and Wiegner, M.: Vertical aerosol profiles from Raman polarization lidar observations during the dry season AMMA field campaign, J. Geophys. Res., 113, D00C11, doi:10.1029/2007JD009487, 2008.

Hu, Q., Goloub, P., Veselovskii, I., Bravo-Aranda, J.-A., Popovici, I., Podvin, T., Haeffelin, M., Lopatin, A., Pietras, C., Huang, X., Torres, B., and Chen, C.: A study of long-range transported smoke aerosols in the Upper Troposphere/Lower Stratosphere, Atmos. Chem. Phys., 19, 1173–1193, https://doi.org/10.5194/acp-19-1173-2019, 2019.

IPCC: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, Nauels, A., Xia, Y., Bex, V. and Midgley, P.M. (eds.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp, 2013.

Janicka, L., Stachlewska, I. S., Veselovskii, I., Baars, H.: Temporal variations in optical and microphysical properties of mineral dust and biomass burning aerosol derived from daytime Raman lidar observations over Warsaw, Poland, *Atmos. Environ.*, 169, 162-174, http://dx.doi.org/10.1016/j.atmosenv.2017.09.022, 2017.

Janicka, L., Bockmann, C., Wang, D., Stachlewska, I. S.: Lidar derived fine scale resolution properties of tropospheric aerosol mixtures, ILRC29, S2-122, Hefei, China, 2019.

Janicka, L. and Stachlewska, I. S.: Properties of biomass burning aerosol mixtures derived at fine temporal and spatial scales from Raman lidar measurements: Part I optical properties, Atmos. Chem. Phys. Discuss., https://doi.org/10.5194/acp-2019-207, 2019.

Lolli, S., Khor, W. Y., Matjafri, M. Z. and Lim, H. S.: Monsoon Season Quantitative Assessment of Biomass Burning Clear-Sky Aerosol Radiative Effect at Surface by Ground-Based Lidar Observations in Pulau Pinang, Malaysia in 2014, Remote Sens., **11**, 2660, doi:10.3390/rs11222660, 2019.

Mariano, G. L., Lopes, F. J. S., Jorge, M. P. P. M., Landulfo, E.: Assessment of biomass burnings activity with the synergy of sunphotometric and LIDAR measurements in São Paulo, Brazil, Atmos. Res., 98, 486-499, doi:10.1016/j.atmosres.2010.08.025, 2010.

Markowicz, K. M., Chilinski, M. T., Lisok, J., Zawadzka, O., Stachlewska, I. S., Janicka, L., Rozwadowska, A., Makuch, P., Pakszys, P., Zielinski, T., Petelski, T., Posyniak, M., Pietruczuk, A., Szkop, A., Westphal, D. L.: Study of aerosol optical properties during long-range transport of biomass burning from Canada to Central Europe in July 2013, *J. Aerosol Sci.*, 101, 156-173, DOI: 10.1016/j.jaerosci.2016.08.006, 2016.

Mattis, I., Müller, D., Ansmann, A., Wandinger, U., Preißler, J., Seifert, P., and Tesche, M.: Ten years of multiwavelength Raman lidar observations of free-tropospheric aerosol layers over central Europe: Geometrical properties and annual cycle, J. Geophys. Res., 113, D20202, doi:10.1029/2007JD009636, 2008.

Mattis, I., D'Amico, G., Baars, H., Amodeo, A., Madonna, F., and Iarlori, M.: EARLINET Single Calculus Chain – technical – Part 2: Calculation of optical products, Atmos. Meas. Tech., 9, 3009–3029, doi:10.5194/amt-9-3009-2016, 2016.

Müller, D., Mattis, I., Wandinger, U., Ansmann, A., Althausen, D., and Stohl, A.: Raman lidar observations of aged Siberian and Canadian forest fire smoke in the free troposphere over Germany in 2003: Microphysical particle characterization, J. Geophys. Res., 110, D17201, doi:10.1029/2004JD005756, 2005.

Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., and Pisani, G.: Aerosol-type-dependent lidar ratios observed with Raman lidar, J. Geophys. Res., 112, D16202, doi:10.1029/2006JD008292, 2007.

Müller, D., Kolgotin, A., Mattis, I., Petzold, A., and Stohl, A., Vertical profiles of microphysical particle properties derived from inversion with two-dimensional regularization of multiwavelength Raman lidar data: experiment, Appl. Opt., 50, 2069-2079, 2011.

Müller, D., Böckmann C., Kolgotin, A., Schneidenbach, L., Chemyakin, E., Rosemann, J., Znak, P., and Romanov, A.: Microphysical particle properties derived from inversion algorithms developed in the framework of EARLINET, Atmos. Meas. Tech., 9, 5007–5035, https://doi.org/10.5194/amt-9-5007-2016,2016.

Murayama, T., Müller, D., Wada, K., Shimizu, A., Sekiguchi, M., and Tsukamoto, T.: Characterization of Asian dust and Siberian smoke with multiwavelength Raman lidar over Tokyo, Japan in spring 2003, Geophys. Res. Lett., 31, L23103, doi:10.1029/2004GL021105, 2004.

Myhre, G., Samset, B. H., Schulz, M., Balkanski, Y., Bauer, S., Berntsen, T. K., Bian, H., Bellouin, N., Chin, M., Diehl, T., Easter, R. C., Feichter, J., Ghan, S. J., Hauglustaine, D., Iversen, T., Kinne, S., Kirkevåg, A., Lamarque, J.-F., Lin, G., Liu, X., Lund, M. T., Luo, G., Ma, X., van Noije, T., Penner, J. E., Rasch, P. J., Ruiz, A., Seland, Ø., Skeie, R. B., Stier, P., Takemura, T., Tsigaridis, K., Wang, P., Wang, Z., Xu, L., Yu, H., Yu, F., Yoon, J.-H., Zhang, K., Zhang, H., and Zhou, C.: Radiative forcing of the direct aerosol effect from AeroCom Phase II simulations, Atmos. Chem. Phys., 13, 1853–1877, doi:10.5194/acp-13-1853-2013, 2013.

Mylonaki, M., Papayannis, A., Mamouri, R., Argyrouli, A., Kokkalis, P., Tsaknakis, G., and Soupiona, O.: Aerosol optical properties variability during biomass burning events observed by the EOLE-AIAS depolarization lidars over Athens, Greece (2007-2016), 28th ILRC, Bucharest, Romania, 2017.

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., 118, 2956–2965, doi:10.1002/jgrd.50324, 2013.

Nicolae, D., Vasilescu, J., Talianu, C., Binietoglou, I., Nicolae, V., Andrei, S., and Antonescu, B.: A Neural Network Aerosol Typing Algorithm Based on Lidar Data, Atmos. Chem. Phys., 18, 14511–14537, https://doi.org/10.5194/acp-18-14511-2018, 2018.

Nicolae, V., Talianu, C., Andrei, S., Antonescu, B., Ene, D., Nicolae, D., Dandocsi, A., Toader, V. E., Stefan, S., Savu, T. and Vasilescu, J.: Multiyear typology of long-range transported aerosols over Europe, Atmosphere, 10, 482, doi:10.3390/atmos10090482, 2019.

Ortiz-Amezcua, P., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Benavent-Oltra, J. A., Böckmann, C., Samaras, S., Stachlewska, I. S., Janicka, Ł., Baars, H., Bohlmann, S., and Alados-Arboledas, L.: Microphysical characterization of long-range transported biomass burning particles from North America at three EARLINET stations, Atmos. Chem. Phys., 17, 5931–5946, doi:10.5194/acp-17-5931-2017, 2017.

Pahlow, M., Kleissl, J., Parlange, M. B., Ondov, J. M. and Harrison, D.: Atmospheric boundary-layer strcture observed during a haze event due to forest-fire smoke, BLM, 114, 53-70, 2005.

Pappalardo, G., Amodeo, A., Apituley, A., Comeron, A., Freudenthaler, V., Linné, H., Ansmann, A., Bösenberg, J., D'Amico, G., Mattis, I., Mona, L., Wandinger, U., Amiridis, V., Alados-Arboledas, L., Nicolae, D., and Wiegner, M.: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmos. Meas. Tech., 7, 2389–2409, doi:10.5194/amt-7-2389-2014, 2014.

Preißler, J., Wagner, F., Guerrero-Rascado, J. L., and Silva, A. M., Two years of free-tropospheric aerosol layers observed over Portugal by lidar, J. GEOPHYS. RES., 118, 3676–3686, doi:10.1002/jgrd.50350, 2013.

Prenni, A., DeMott, P. J., Sullivan, A. P., Sullivan, R. C., Kreidenweiss, S. M., and Rogers, D. C.: Biomass burning as a potential source for atmospheric ice nuclei: Western wildfires and prescribed burns, Gephys. Res. Lett., 39, L11805, doi:10.1029/2012GL051915, 2012.

Reid, J. S., and Hobbs, P. V.: Physical and optical properties of young smoke from individual biomass fires in Brazil, J. Geophys. Res., 103, 32013-32030, 1998.

Reid, J. S., Eck, T. F. , Christopher, S. A., Koppmann, R., Dubovik, O., Eleuterio, D. P., Holben, B. N., Reid, E. A., and Zhang, J.: A review of biomass burning emissions part III: intensive optical properties of biomass burning particles, Atmos. Chem. Phys., 5, 827–849, 2005, www.atmos-chem-phys.org/acp/5/827/.

Rolph, G., Stein, A., Stunder, B.: Real-time Environmental Applications and Display sYstem: READY, Environ. Modelling Soft., 95, 210-228, https://doi.org/10.1016/j.envsoft.2017.06.025, 2017.

Samaras, S., Nicolae, D., Böckmann, C., Vasilescu, J., Binietoglou, I., Labzovskii, L., Toanca, F., Papayannis, A.: Using Raman-lidar-based regularized microphysical retrievals and Aerosol Mass Spectrometer measurements for the characterization of biomass burning aerosols, Journal of Computational Physics 299, 156–174, http://dx.doi.org/10.1016/j.jcp.2015.0, 2015.

San-Miguel-Ayanz, J., Durrant, T., Boca, R., Libertà, G., Branco, A., de Rigo, D., Ferrari, D., Maianti, P., Artés Vivancos, T., Costa, H., Lana, F., Löffler, P., Nuijten, D., Ahlgren, A. C., Leray, T: Forest Fires in Europe, Middle East and North Africa 2017. EUR 29318 EN, ISBN 978-92-79-92831-4, doi: 10.2760/663443, 2018.

Sapkota, A., Symons, J. M., Kleissl, J., Wang, L., Parlange, M. B., Ondov, J., Breysse, P. N., Diette, G. B., Eggleston, P. A., and Buckley, T.: Impact of the 2002 Canadian Forest Fires on Particulate Matter Air Quality in Baltimore City, Environ. Sci. Technol., 39, 24-32, 2005.

Sicard, M., Granados-Muñoz, M.J., Alados-Arboledas, L., Barragán, R., Bedoya-Velásquez, A.E., Benavent-Oltra, J.A., Bortoli, D., Comerón, A., Córdoba-Jabonero, C., Costa, M.J., del Águila, A., Fernández, A.J., Guerrero-Rascado, J.L., Jorba, O., Molero, F., Muñoz-Porcar, C., Ortiz-Amezcua, P., Papagiannopoulos, N., Potes, M. , Pujadas, M., Rocadenbosch, F., Rodríguez-Gomez, A., Román, R., Salgado, R., Salgueiro, V., Sola, Y., Yela, M.: Ground/space, passive/active remote sensing observations coupled with particle dispersion modelling to understand the inter-continental transport of wildfire smoke plumes, Remote Sens Environ, **232**, 111294, https://doi.org/10.1016/j.rse.2019.111294, 2019.

Stachlewska, I. S., Samson, M., Zawadzka, O., Harenda, K. M., Janicka, L., Poczta, P., Szczepanik, D., Heese, B., Wang, D., Borek, K., Tetoni, E., Proestakis, E., Siomos, N., Nemuc, A., Chojnicki, B. H., Markowicz, K. M., Pietruczuk, A., Szkop, A., Althausen, D., Stebel, K., Schuettemeyer, D. and Zehner, C.: Modification of Local Urban Aerosol Properties by Long-Range Transport of Biomass Burning Aerosol, Remote Sens., **10**, 412; doi:10.3390/rs10030412, 2018.

Stein, A.F., Draxler, R.R, Rolph, G.D., Stunder, B.J.B., Cohen, M.D., and Ngan, F.: NOAA's HYSPLIT atmospheric transport and dispersion modeling system, Bull. Amer. Meteor. Soc., **96**, 2059-2077, http://dx.doi.org/10.1175/BAMS-D-14-00110.1, 2015.

Su, W., Schuster, G. L., Loeb, N. G., Rogers, R. R., Ferrare, R. A., Hostetler, C. A., Hair, J. W., and Obland, M. D.: Aerosol and cloud interaction observed from high spectral resolution lidar data, J. Geophys. Res., 113, doi:10.1029/2008JD010588 (2008)

Su, L., Yuan, Z., Fung, J. C. H., Lau, A. K. H.: A comparison of HYSPLIT backward trajectories generated from two GDAS datasets, Sci. Total Environ., 506-507, 527-537, http://dx.doi.org/10.1016/j.scitotenv.2014.11.072, 2015.

Sugimoto, N., Tatarov, B., Shimizu, A., Matsui, I., and Nishizawa, T.: Optical Characteristics of Forest-Fire Smoke Observed with Two-Wavelength Mie-Scattering Lidars and a High-Spectral-Resolution Lidar over Japan, SOLA, 6, 093−096, doi:10.2151/sola.2010-024, 2010.

Tesche, M., Müller, D., Gross, S., Ansmann, A., Althausen, D., Freundenthaler, V., Weinzierl, B., Veira, A. and Petzold, A.: Optical and microphysical properties of smoke over Cape Verde inferred from multiwavelength lidar measurements, Tellus, 63B, 677-694, DOI: 10.1111/j.1600-0889.2011.00549.x, 2011.

van Drooge, B. L., Sicard, M., Stohl, A., Fontal, M., Bravo, N., Muñoz, A., Lange, D., Fernández, P., Grimalt, J. O.: Detection and simulation of wildfire smoke impacting a Mediterranean urban atmosphere, Atmos. Poll. Res., 7, 494-502, http://dx.doi.org/10.1016/j.apr.2015.12.003, 2016.

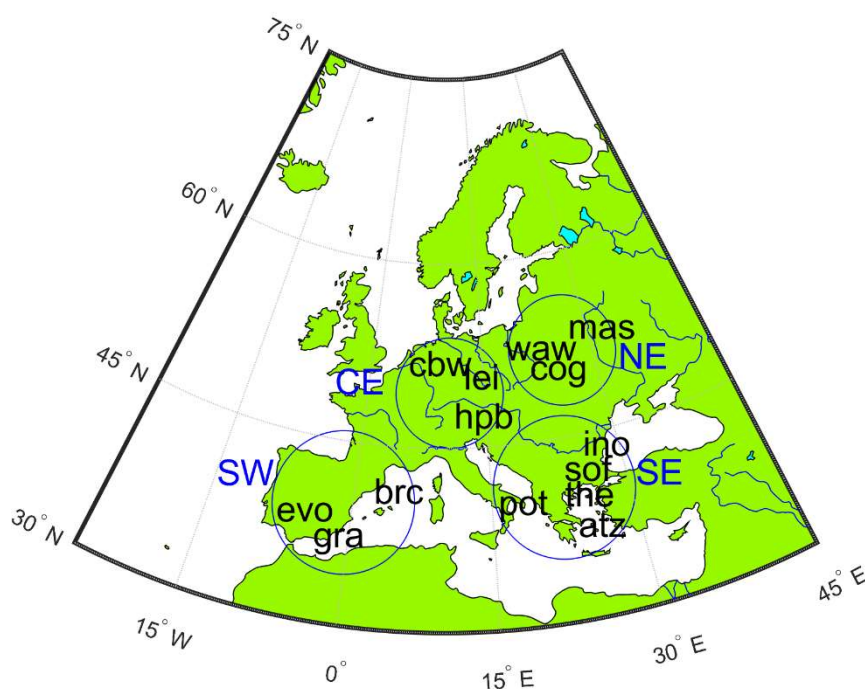
Vaughan, G., Draude, A. P., Ricketts, H. M. A., Schultz, D. M., Adam, M., Sugier, J. and Wareing, D. P.: Transport of Canadian forest fire smoke over the UK as observed by lidar, Atmos. Chem. Phys., 18, 11375–11388, https://doi.org/10.5194/acp-18-11375-2018, 2018.

Veselovskii, I, Whiteman, D. N, Korenskiy, M., Suvorina, A., Kolgotin, A., Lyapustin, A., Wang, Y., Chin, M., Bian, H., Kucsera, T. L., Pérez-Ramírez, D. and Holben, B.: Characterization of forest fire smoke event near Washington, DC in summer 2013 with multi-wavelength lidar, Atmos. Chem. Phys., 15, 1647–1660, doi:10.5194/acp-15-1647-2015, 2015.

Wandinger, U., Müller, D., Böckmann, C., Althausen, D., Matthias, V., Bösenberg, J., Weiß, V., Fiebig,, M. Wendisch, M., Stohl, A. and Ansmann, A.: Optical and microphysical characterization of biomassburning and industrial-pollution aerosols from multiwavelength lidar and aircraft measurements, J. Geophys. Res., 107, NO. D21, 8125, doi:10.1029/2000JD000202, 2002.

5 Wang D., Szczepanik, D. and Stachlewska, I. S.: Interrelations between surface, boundary layer, and columnar aerosol properties derived in summer and early autumn over a continental urban site in Warsaw, Poland, Atmos. Chem. Phys., 19, 13097-13128, https://doi.org/10.5194/acp-19-13097-2019, 2019.

Yu, S., Role of organic acids formic, acetic, pyruvic and oxalic in the formation of cloud condensation nuclei CCN: a review, Atmos. Res., 53, 185-217, 2000.

10 Zhang, J., Reid, J. S., Christensen, M. and Benedetti, A.: An evaluation of the impact of aerosol particles on weather forecasts from a biomass burning aerosol event over the Midwestern United States: observational-based analysis of surface temperature, Atmos. Chem. Phys., 16, 6475–6494, doi:10.5194/acp-16-6475-2016, 2016.

15 **Figure 1. The geographical location of the 14 stations proving data for Forest Fire category in EARLINET database over 2008-2017 period. The stations are located in Athens ("atz"), Barcelona ("brc"), Belsk ("cog"), Bucharest ("ino"), Cabauw ("cbw"), Evora ("evo"), Granada ("gra"), Leipzig ("lei"), Minsk ("mas"), Observatory Hohenpeissenberg ("hpb"), Potenza ("pot"), Sofia ("sof"), Thessaloniki ("the") and Warsaw ("waw"). The blue circles show the four European geographical regions: South-East (SE), South-West (SW), North-East (NE) and Central (CE).**

Step	Footnote
I) Read optical properties Input: b and e files[1]	
II) Quality control of the data[2]	
III) Overview of the metadata[3]	
IV) Layers estimation[4]	
V) Calculation of backtrajectories[5]	
VI) Smoke layers identification[6]	
VII) Calculation of mean optical properties in the layers[7]	
VIII) Calculation of intensive parameters[8]	
IX) Fires seen by two stations[9]	
X) LRT from N America[10]	XI) Geographical regions[12]
Histogram of fires' location[11]	
Histogram of fires' location[13]	
Histogram of backtrajectories[14]	
Intensive parameters by region[15]	

[1] 3589 input files; see Figs. 3, 4 and Table 1

[2] EARLINET/Forest Fire files checked following EARLINET criteria; additional in-house checks

[3] e.g. detection mode, evaluation method, raw resolution; see Table 2.

[4] algorithm developed; non-accurate estimates were manually corrected; 1901 layers calculated for 960 time stamps

[5] run Hysplit, 10 days backwards; use GDAS0.5° (23 cases on GDAS1°)

[6] use Hysplit backtrajectory and fires emissions (FIRMS); 1053 layers for 677 time stamps were considered as having BB origin; see Table 3, columns 5-6.

[7] the average in the layer is calculated if at least 90% data is available; the averages for which SNR<2 were dismissed; there were 1050 layers for 676 time stamps for which at least one average optical property was calculated; see Table 3, columns 7-8, e.g. Fig. 7.

[8] e.g. Fig. 9; literature review on intensive parameters (Fig. S1 and Table S1). Values outside the range of acceptable values are dismissed. Remaining IPs are considered further (795 layers, 526 time stamps). See Table 4, columns 9-10.

[9] based on backtrajectories, the common fires (smoke measured by two stations) were identified; compares IPs; measured smoke can be of "single fire" or "mixed fires". See Figs. 10-11.

[10] based on backtrajectories, the measured smoke can be "pure N America" or "mixed" (N America + local); smoke was measured by single station or several stations; e.g. smoke measured by one station Figs. 12-13; statistics over LRT from N America to be shown in second part of the paper.

[11] by station; e.g. Fig. 12.

[12] region: SE ("atz","ino","pot","the"), NE ("cog","mas","waw"), SW ("brc","evo","gra"), CE ("cbw","lei","hpb")

[13] by station; e.g. Figs. 14 for SE region.

[14] main circulation pattern revealed for some regions.

[15] IPs are classified based on smoke' continental origin (Europe, N America, Asia, Africa or a mixture of them). Mean values and scatter plots are considered. Fig. 15 for SE region.

Figure 2. Methodology diagram.

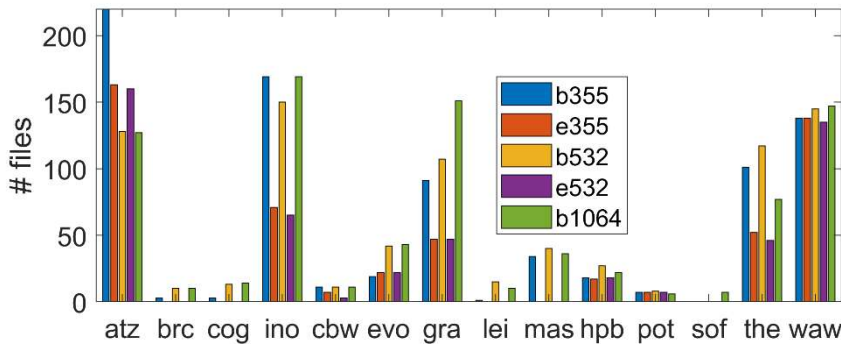



Figure 3. Number of "b" and "e" files available from 14 EARLINET stations providing data for Forest Fire over 2008-2017 period.

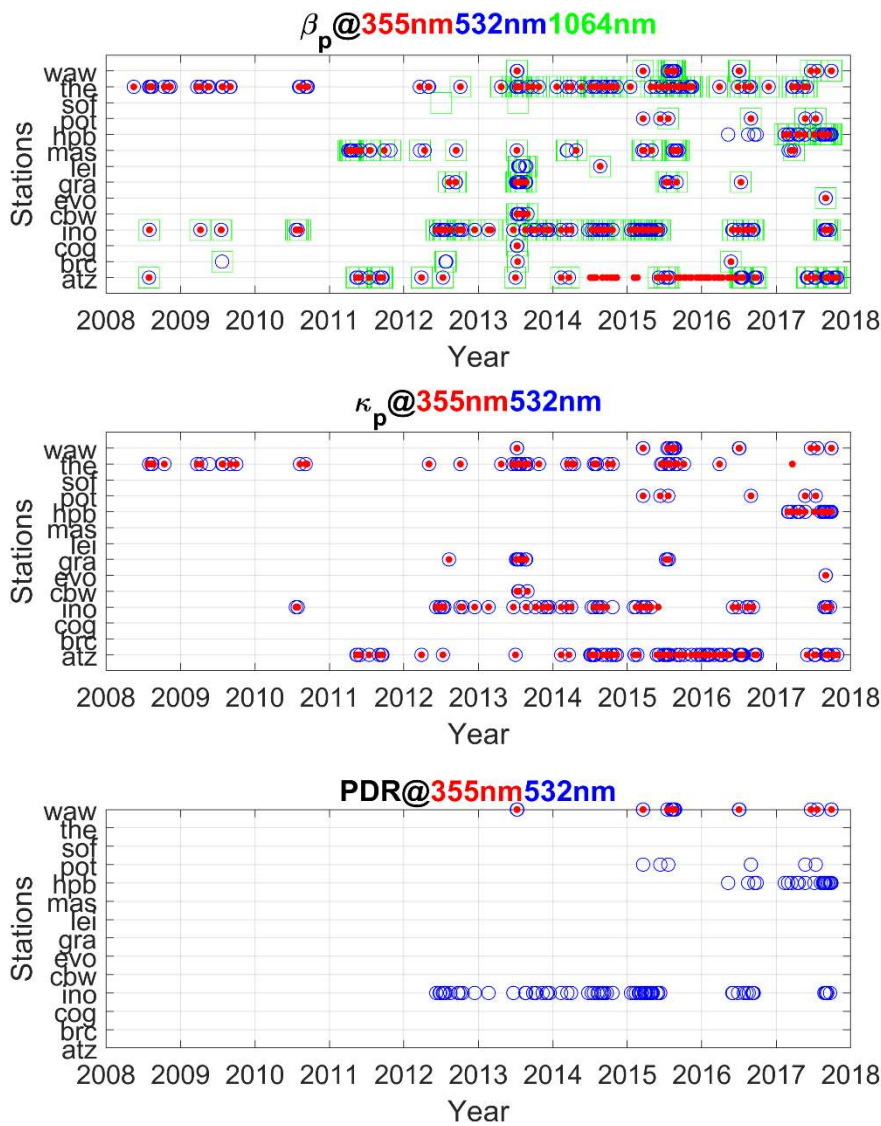



Figure 4. Time span of the available particles backscatter coefficient (top), extinction coefficient (middle) and depolarization ratio (bottom) for each station, over 2008-2017 period, for the smoke layers.

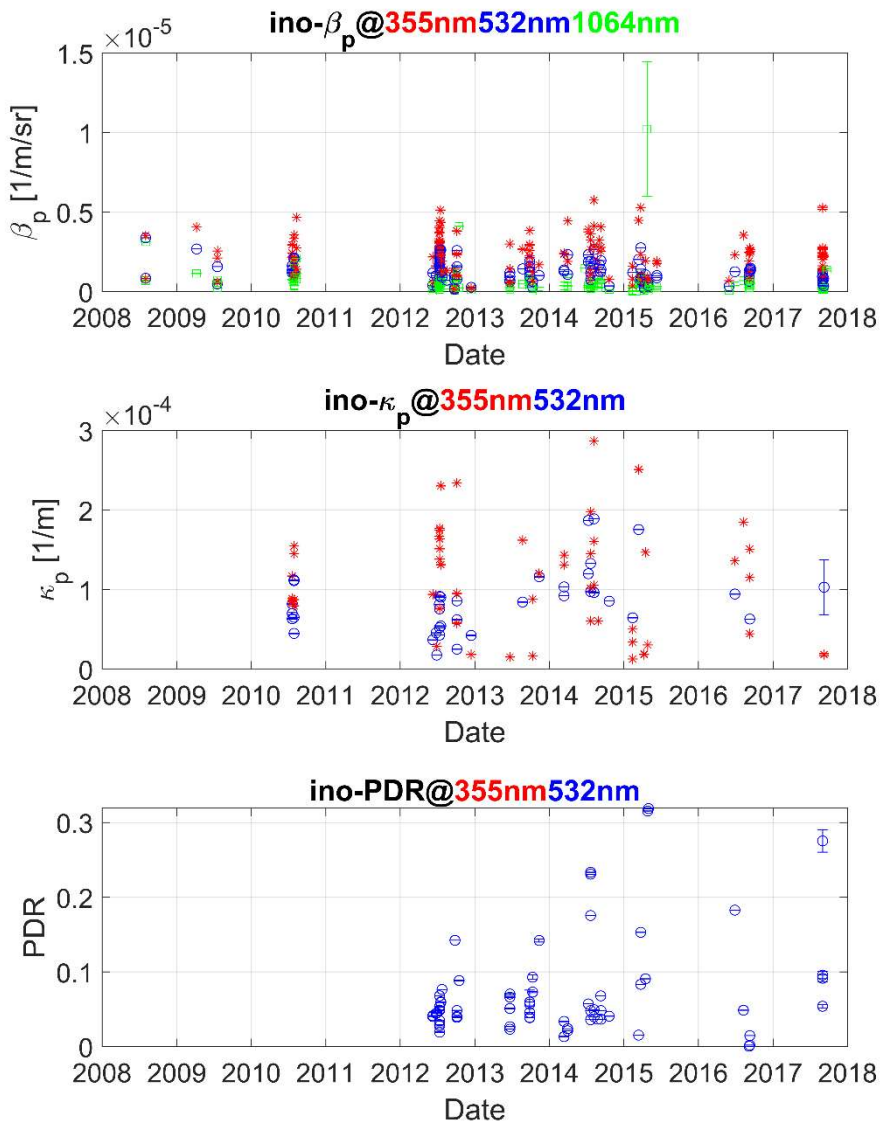



Figure 5. Values of the mean optical properties and associated uncertainties in the layers. Example for the Bucharest ("ino") station. Layers were identified as having a fire source. Note that with few exceptions the uncertainties are very small and thus undiscernible from the mean values.

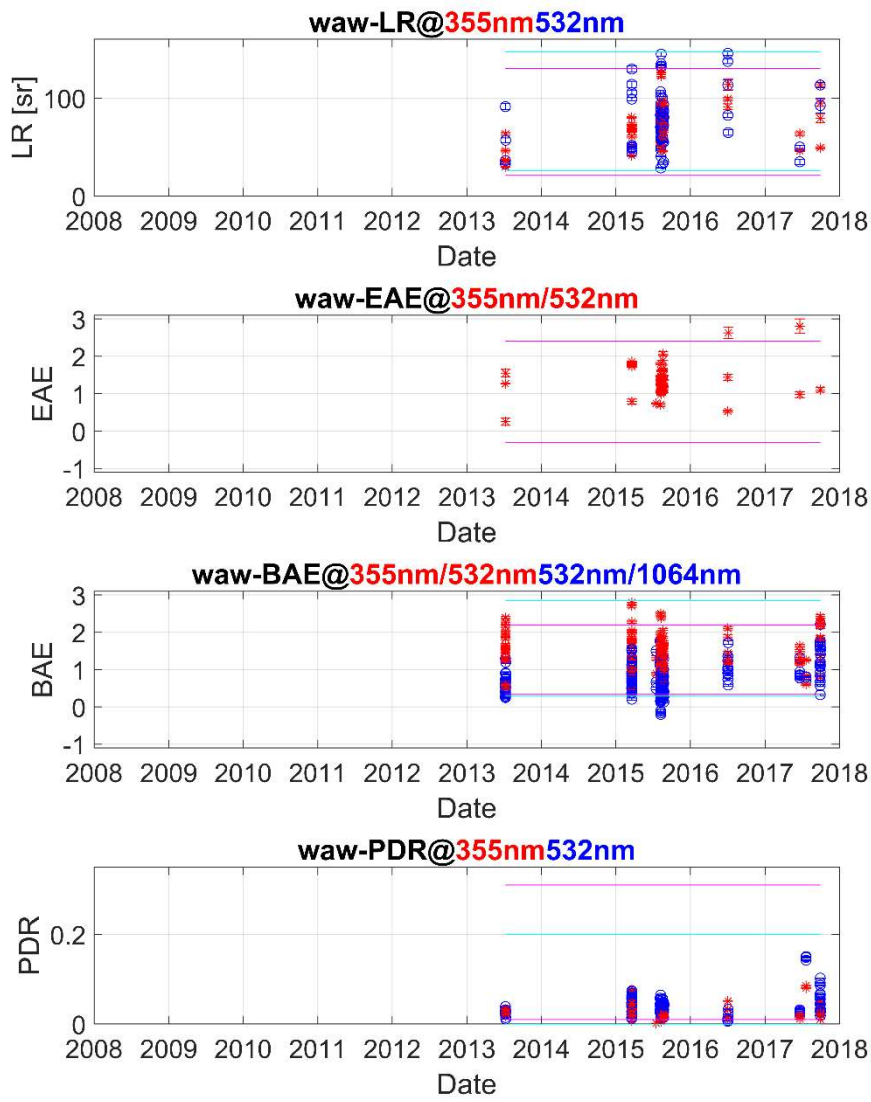



Figure 6. Intensive aerosol parameters derived in the layers. Example for Warsaw ("waw") station. The lines represent the minimum and maximum values reported in literature. Magenta lines represent the extreme values for the intensive parameters shown in red while the cyan lines represent the extreme values for the intensive parameters shown in blue. Layers have fire source origin.

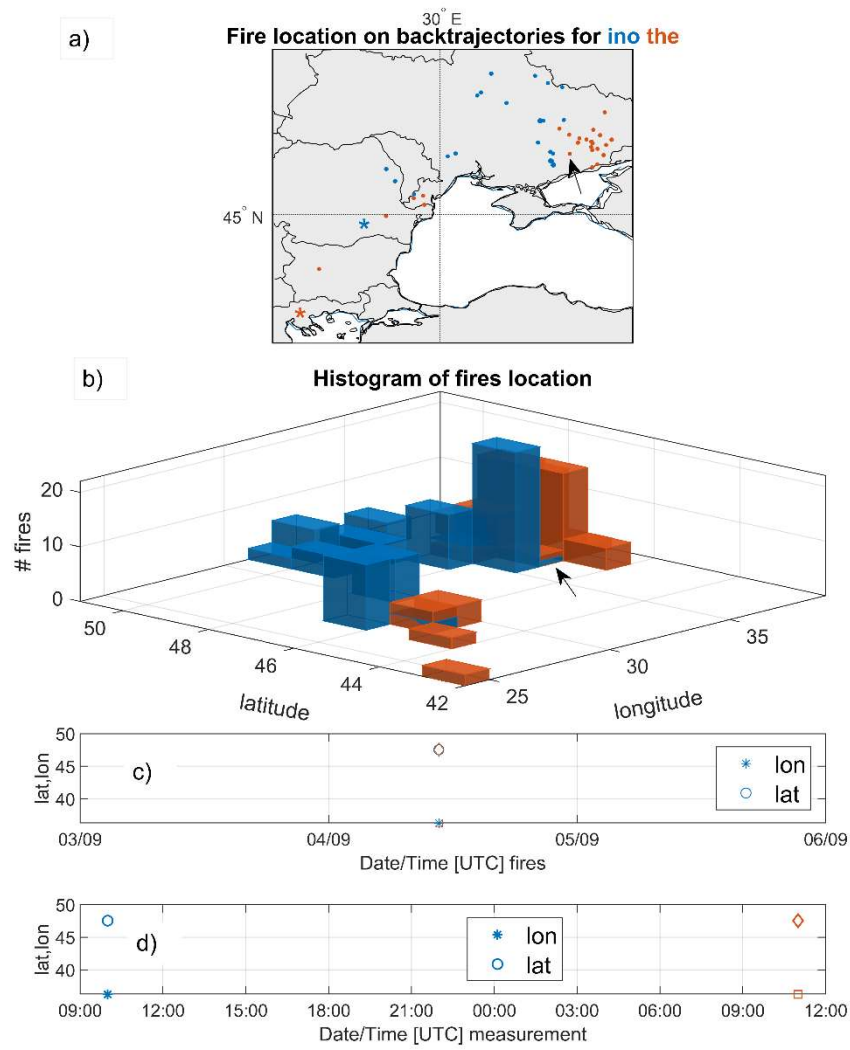
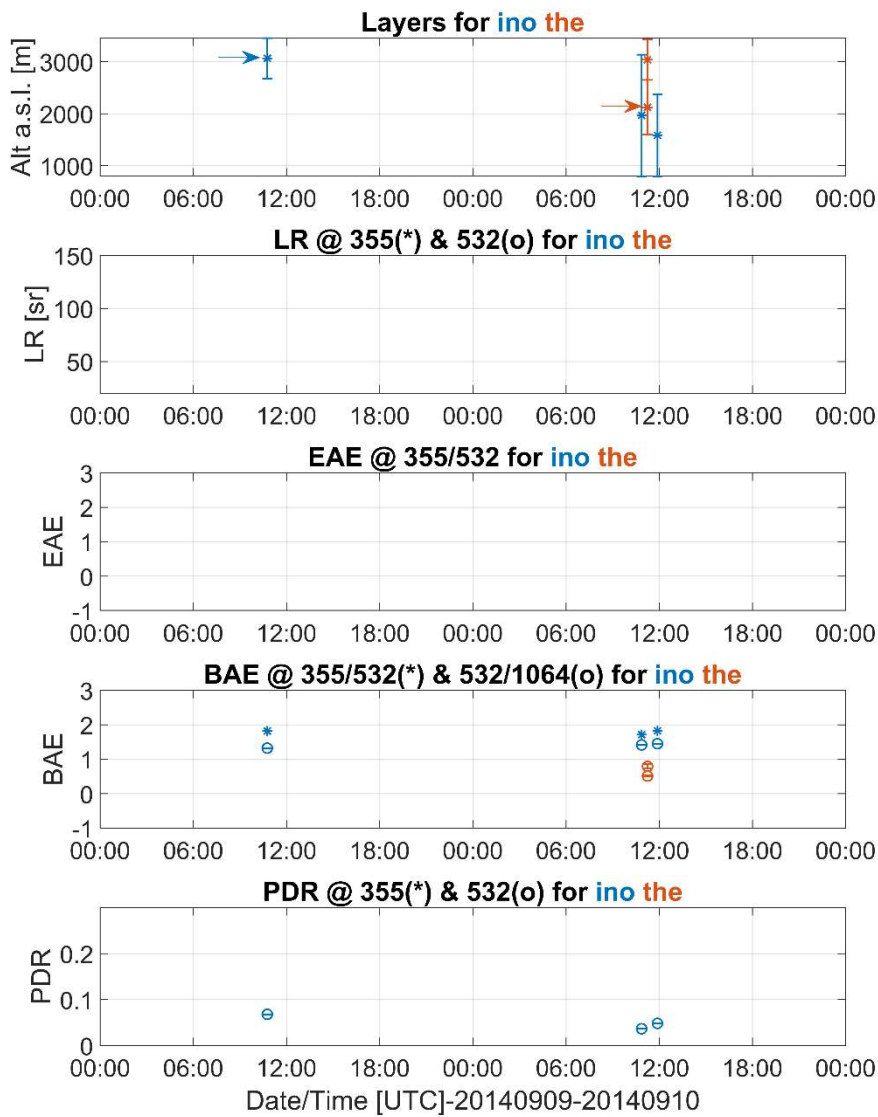



Figure 7. Measurements with the same source in Thessaloniki ("the") and Bucharest ("ino") during 20140909-20140910. a) Fires location seen during back-trajectories from each station (colour coded); b) Histogram of the fires occurrence in each geographical grid; c) Longitude and latitude of the fires' location versus fires' occurrence time; d) Longitude and latitude of the fires' location versus measurement time at the two locations; e–i) layers altitude and intensive parameters for each station. Layers measuring the common fire are shown by arrows. The geographical location of the common fire is shown on histogram by an arrow.

a)

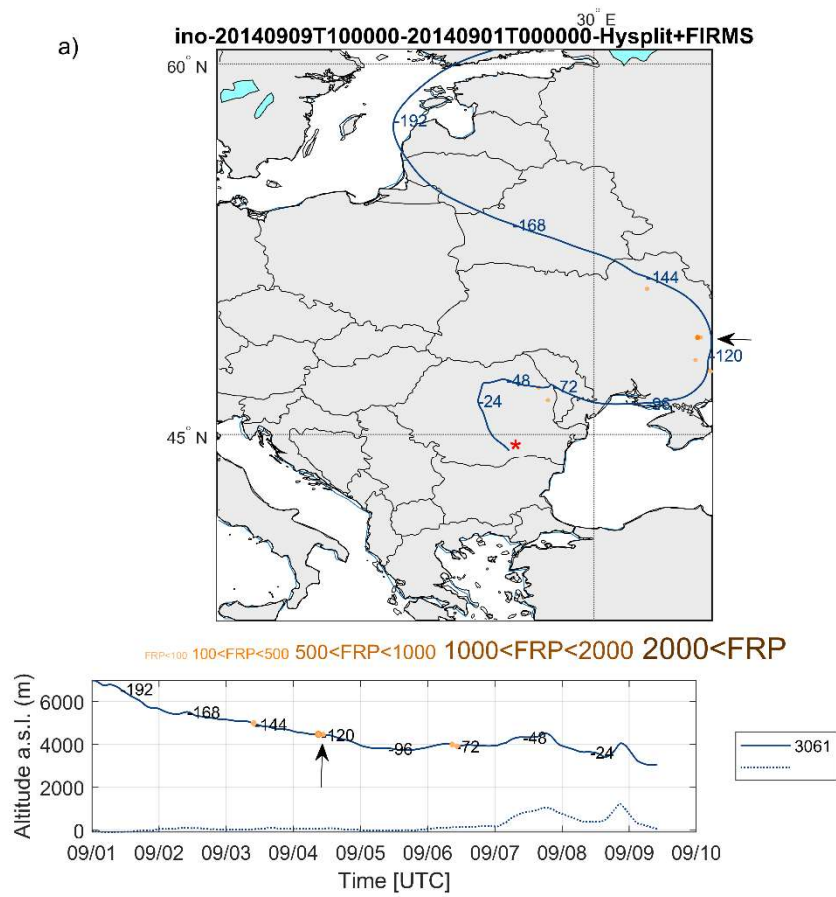
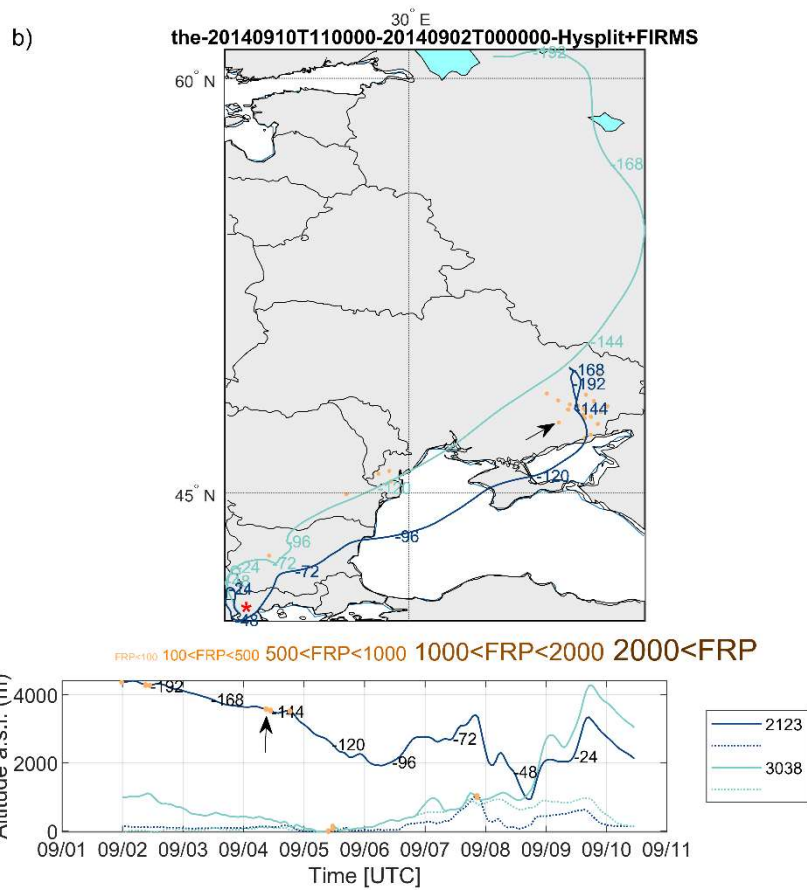



Figure 8. Backtrajectories and location of fires along backtrajectories within 100 km and +/- 1h for Bucharest ("ino") on 20140909 (a) and Thessaloniki ("the") on 20140910 (b). Lower plots show the altitude (a.s.l.) of the backtrajectories versus time. The common fire location is marked with an arrow. See text for more details.

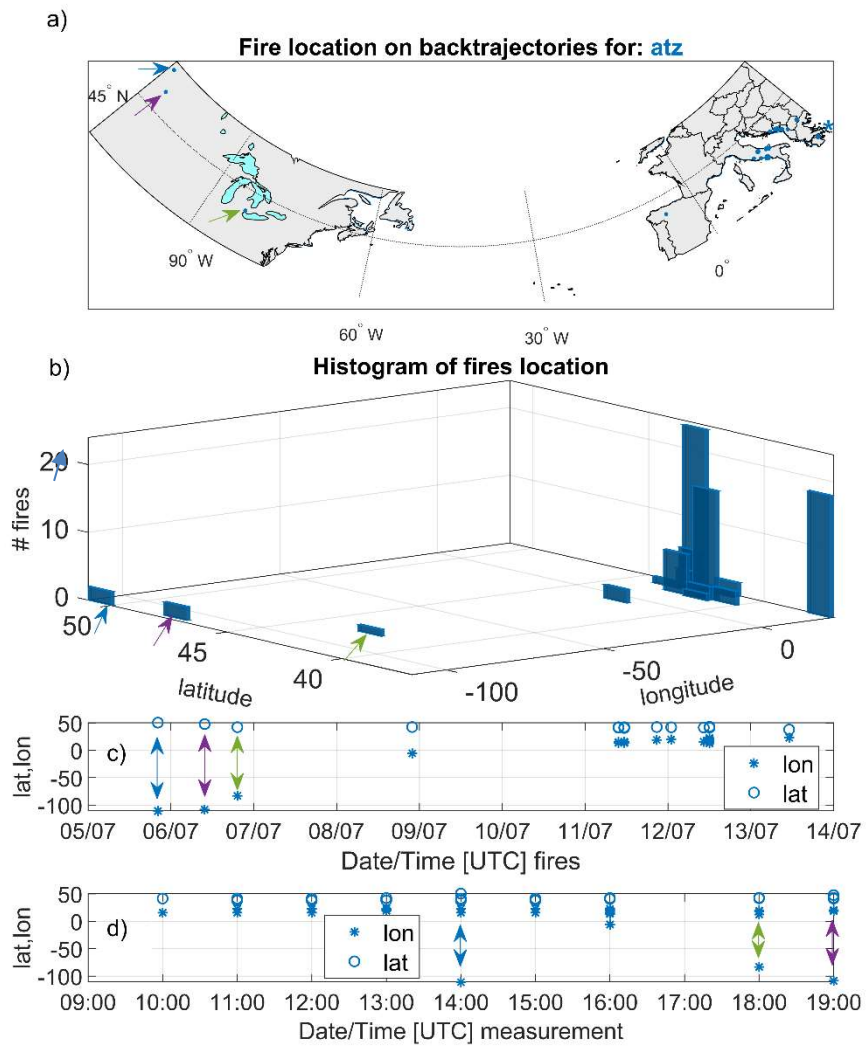
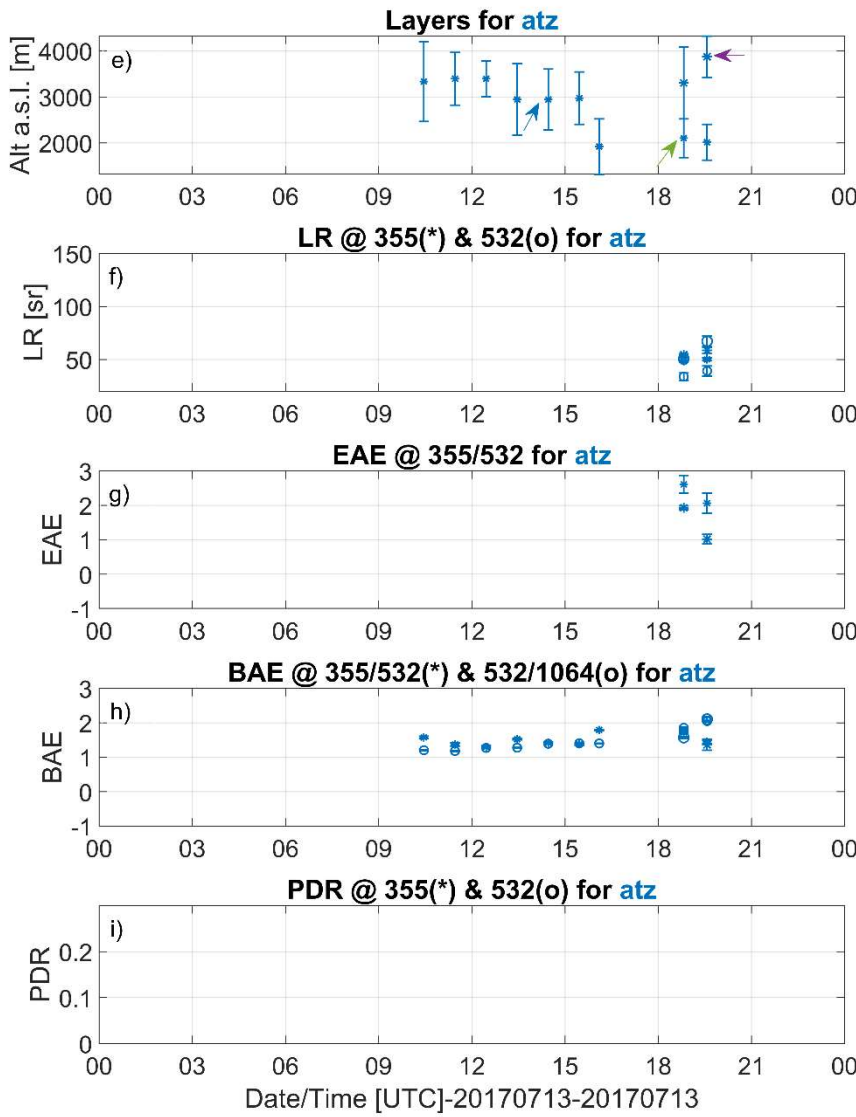



Figure 9. LRT as measured at Athens ("atz") on 20170713. a) Location of the fires; b) Histogram of the fires. The North America fires are marked by arrows; c) Fires' coordinates versus fires' occurrence time; d) Fires' coordinates versus measurements time; e) Location of the layers marked by arrows. First two BB layers are considered "mixed" while the third (magenta) "pure NA"; f) – i) Intensive parameters.

a)

atz-20170713T140000-20170705T000000-Hysplit+FIRMS

FRP<100 100<FRP<500 500<FRP<1000 1000<FRP<2000 2000<FRP

b)

atz-20170713T180000-20170705T000000-Hysplit+FIRMS

FRP<100 100<FRP<500 500<FRP<1000 1000<FRP<2000 2000<FRP

c)

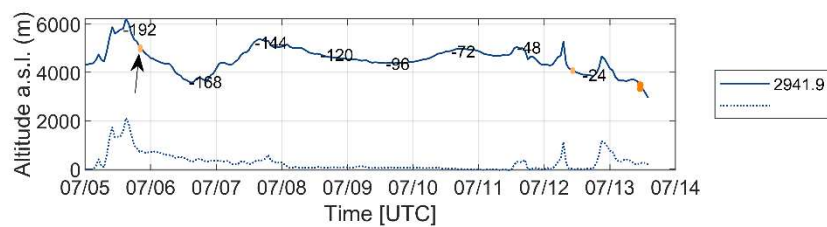
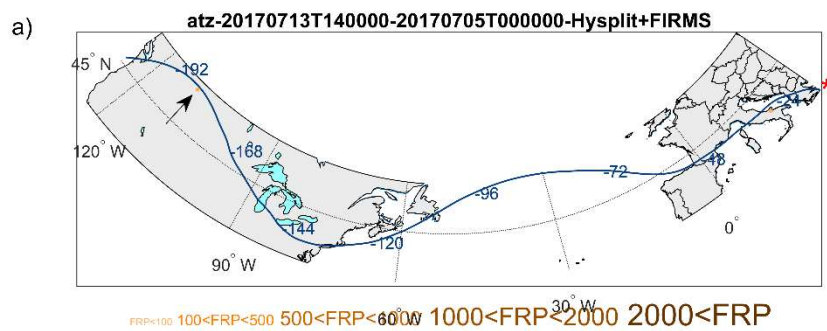
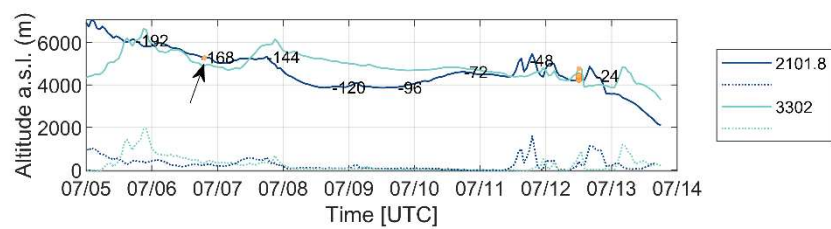
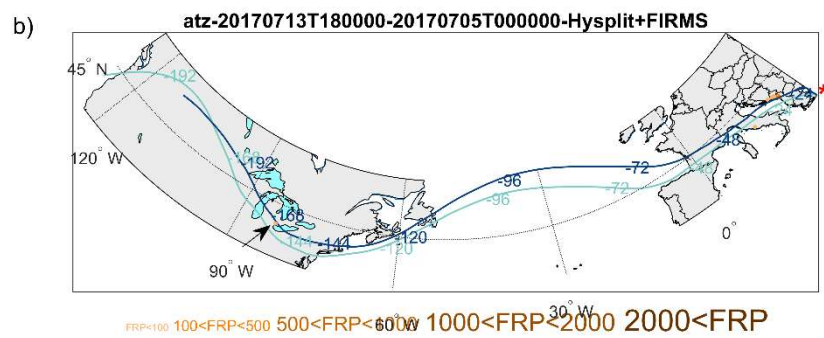
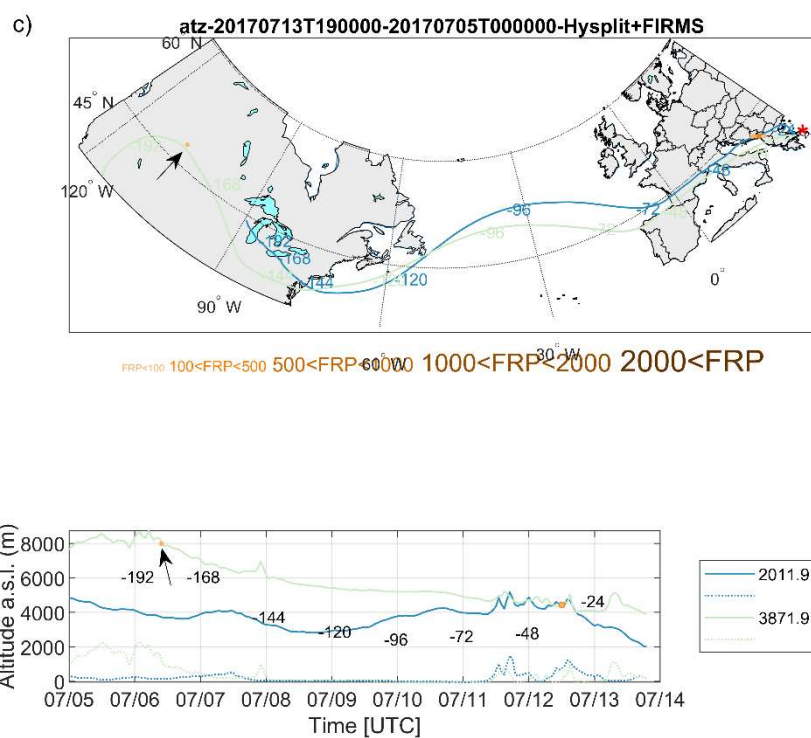



Figure 10. Backtrajectories for layers shown in Fig. 9 for Athens ("atz") station. Layers in a) and b) are considered "mixed" while the layer in c) is considered "pure N America". See text for more explanation. The fire location is marked with an arrow.

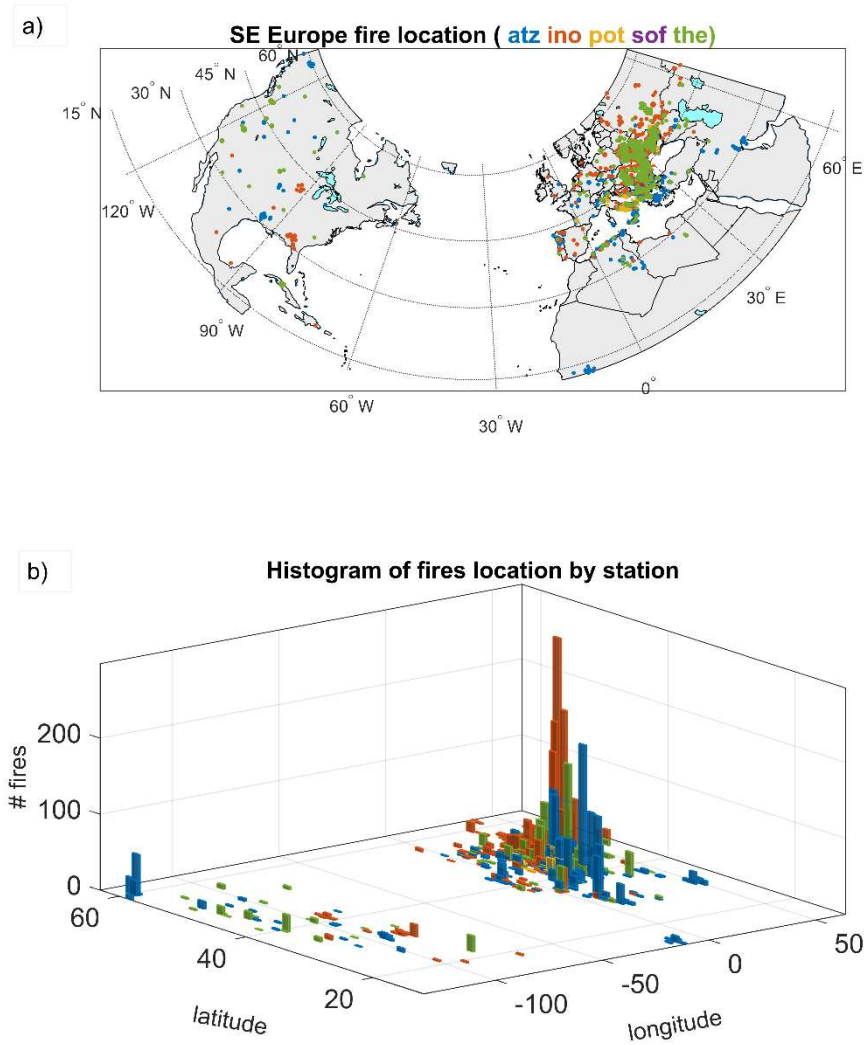



Figure 11. South-East Europe region formed by stations Athens ("atz"), Bucharest ("ino"), Potenza ("pot"), Sofia ("sof") and Thessaloniki ("the"). a) Location of the fires detected by each station. Note that due to overlap some are not seen. b) Histogram of the fires detected by each station.

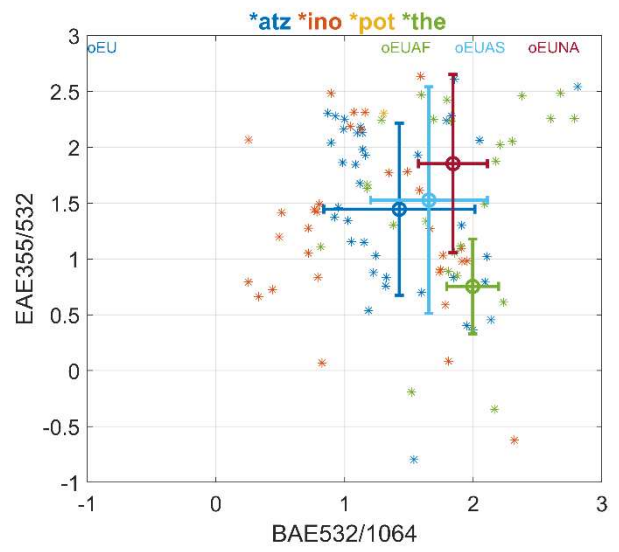
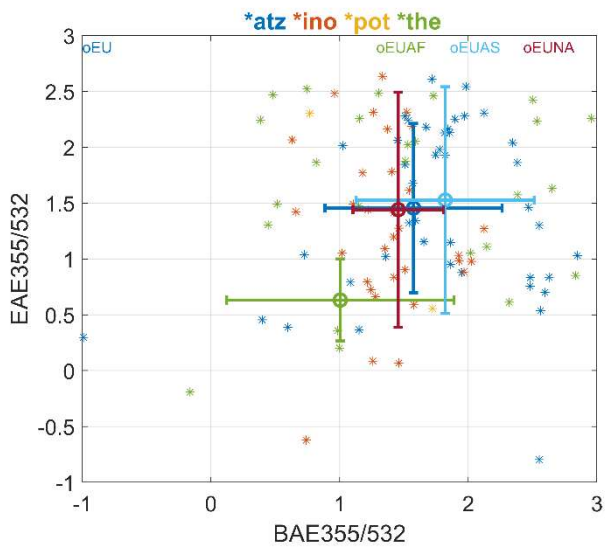
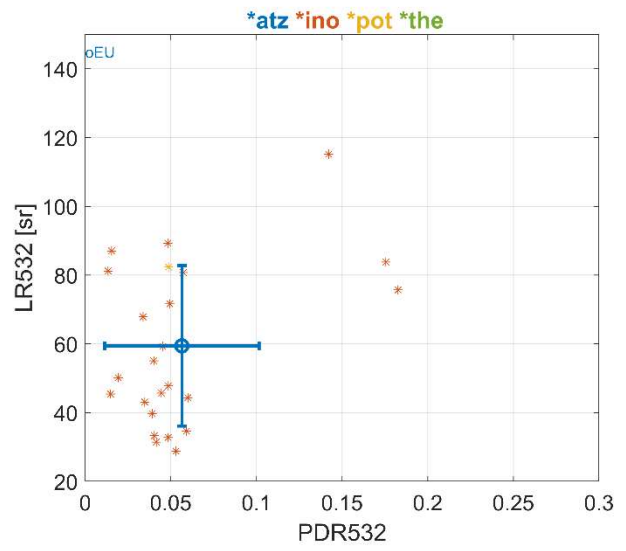
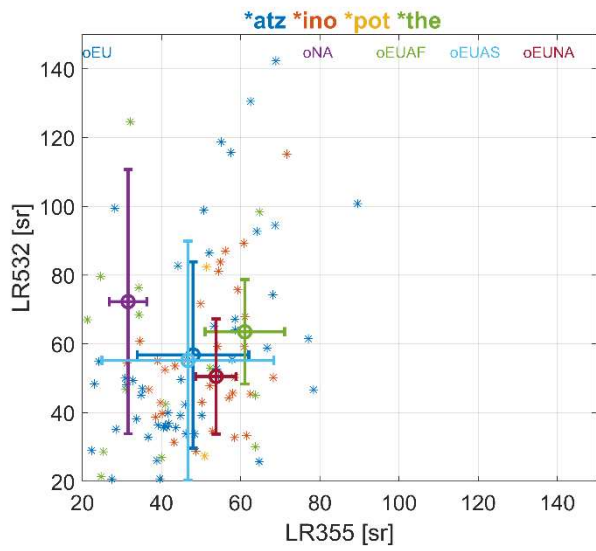
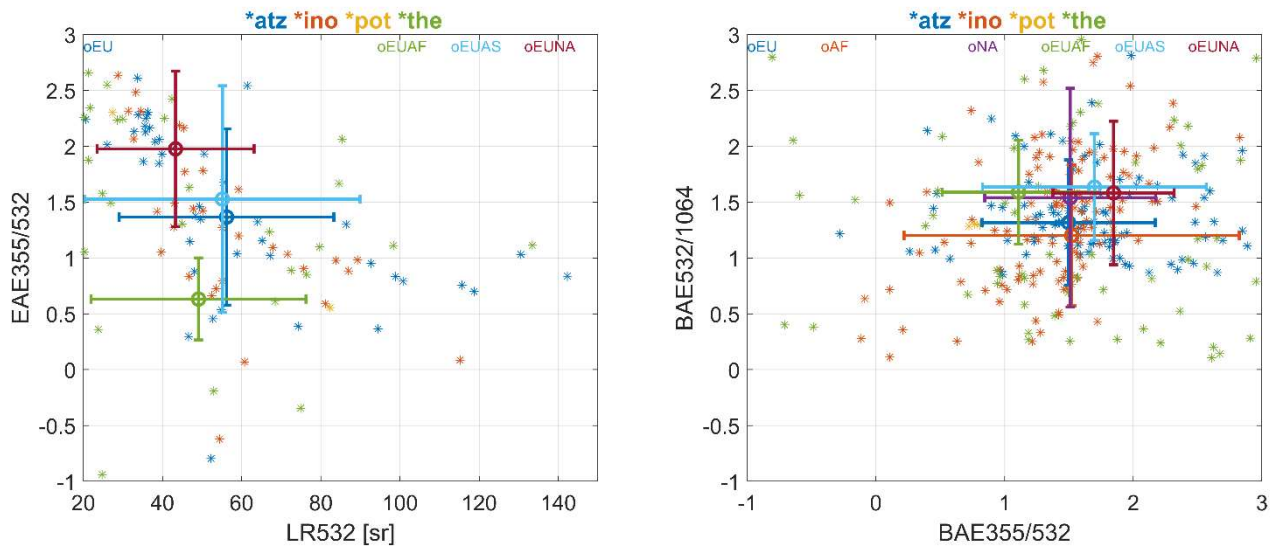



Figure 12. Scatter plots between various two intensive parameters for South-East region (LR@532 vs LR@355, LR@532 vs PDR@532, EAE355/532 vs BAE355/532, EAE355/532 vs BAE532/1064, EAE355/532 vs LR@532 and BAE532/1064 vs BAE355/532). The colour code of the points is station related (as labelled in the title). The colour code for the mean and STD values is related with the source origin (as stated on the plots).

Table 1. List of stations and their number of backscatter (b-files) and/or extinction (e-files) at 355nm, 532nm and 1064nm. The number of particles backscatter (βp) and extinction (κp) coefficients and particles linear depolarization ratio (PDR) are shown as well. Highlights systems: 3b+2e; 3b+2e+1d or 3b+2e+2d; 3b; 1b. b=backscatter, e=extinction, d=depolarization. The last column represents the number of time stamps with at least one profile of optical properties (total number being 1138).

	b355 files	b532 files	b1064 files	e355 files	e532 files	total b+e files	βp 355*	βp 532*	βp 1064	κp 355	κp 532	PDR 355	PDR 532	# times
"atz"	324	128	127	163	160	902	324	128	127	163	160			329
"brc"	3	10	10	0	0	23	3	10	10	0	0			13
"cog"	3	13	14	0	0	30	3	13	14	0	0			14
"ino"	169	150	169	71	65	624	169	153	169	71	65		144	173
"cbw"**	11	11	11	7	3	43	11	11	11	7	3			11
"evo"	19	42	43	22	22	148	19	42	43	22	22			43
"gra"	91	107	151	47	47	443	138	154	151	47	47			163
"lei"	1	15	10	0	0	26	1	15	10	0	0			17
"mas"	34	40	36	0	0	110	34	40	36	0	0			42
"hpb"	18	27	22	17	18	102	18	27	22	17	18		24	27
"pot"	7	8	6	7	7	35	8	8	6	7	7		8	8
"sof"	0	0	7	0	0	7	0	0	7	0	0			7
"the"	101	117	77	52	46	393	101	117	77	52	46			140
"waw"	138	145	147	138	135	703	140	146	147	138	135	131	144	151
#tot	919	813	830	524	503	3589	969	864	830	524	503	131	320	4141
#sta	13	13	14	9	9		12	12	13	8	8	1	4	

* backscatter coefficient profiles from both b-files and e-files; for the same time stamp, the profile from e-file is kept

#tot=total number of files or profiles; #sta=number of stations with particular file or profile

Table 2. (2) Number of time stamps with at least one profile of optical properties (# times); (3) number of time stamps with layers and at least one optical property in the layer; (4) total number of layers for time stamps in (3); (5) number of time stamps with optical properties of fire origin. (6) total number of layers for time stamps in (5); Columns 7 and 8 represent the number of times with good optical properties and the total number of layers with good optical properties. Columns 9 and 10 represent the final number of time stamps and corresponding number of layers within imposed acceptable range.

1	2	3	4	5	6	7	8	9	10
Station	# times	# times with layers with optical properties	total # layers for (3)	# times with layers with optical properties of fire origin	total # layers for (5)	# times with layers with optical properties of fire origin, SNR>=2	total # layers for (7)	# times with layers with IP of fire origin, SNR>=2, within accepted range	total # layers for (9)
"atz"	329	233	405	172	251	171	250	112	156
"brc"	13	13	31	12	23	12	23	6	14
"cog"	14	14	58	14	21	14	21	13	20
"ino"	173	163	289	109	168	109	167	86	130
"cbw"	11	11	26	8	14	8	14	8	14
"evo"	43	42	96	13	16	13	16	6	8
"gra"	163	146	296	80	120	80	120	68	97
"lei"	17	17	35	11	18	11	18	4	5
"mas"	42	32	86	25	52	25	52	21	43
"hpb"	27	21	50	12	18	12	17	11	15
"pot"	8	7	9	5	7	5	7	4	6
"sof"	7	7	11	4	4	4	4	0	0
"the"	140	103	216	88	151	88	151	66	107
"waw"	151	151	293	124	190	124	190	121	180
	1138	960	1901	677	1053	676	1050	526	795