Atmospheric
Measurement
Techniques

Open Access

EGU

EARLINET Single Calculus Chain – technical – Part 1: Pre-processing of raw lidar data

Giuseppe D'Amico[1], **Aldo Amodeo**[1], **Ina Mattis**[2,4], **Volker Freudenthaler**[3], **and Gelsomina Pappalardo**[1]

[1]Consiglio Nazionale delle Ricerche, Istituto di Metodologie per l'Analisi Ambientale (CNR-IMAA), Tito Scalo, Potenza, Italy
[2]Leibniz Institute for Tropospheric Research, Leipzig, Germany
[3]Ludwig-Maximilians-Universität, Meteorologisches Institut Experimentelle Meteorologie, Munich, Germany
[4]Deutscher Wetterdienst, Meteorologisches Observatorium Hohenpeißenberg, Germany

Correspondence to: Giuseppe D'Amico (giuseppe.damico@imaa.cnr.it)

Abstract. In this paper we describe an automatic tool for the pre-processing of aerosol lidar data called ELPP (EARLINET Lidar Pre-Processor). It is one of two calculus modules of the EARLINET Single Calculus Chain (SCC), the automatic tool for the analysis of EARLINET data. ELPP is an open source module that executes instrumental corrections and data handling of the raw lidar signals, making the lidar data ready to be processed by the optical retrieval algorithms. According to the specific lidar configuration, ELPP automatically performs dead-time correction, atmospheric and electronic background subtraction, gluing of lidar signals, and trigger-delay correction. Moreover, the signal-to-noise ratio of the pre-processed signals can be improved by means of configurable time integration of the raw signals and/or spatial smoothing. ELPP delivers the statistical uncertainties of the final products by means of error propagation or Monte Carlo simulations.

During the development of ELPP, particular attention has been payed to make the tool flexible enough to handle all lidar configurations currently used within the EARLINET community. Moreover, it has been designed in a modular way to allow an easy extension to lidar configurations not yet implemented.

The primary goal of ELPP is to enable the application of quality-assured procedures in the lidar data analysis starting from the raw lidar data. This provides the added value of full traceability of each delivered lidar product.

Several tests have been performed to check the proper functioning of ELPP. The whole SCC has been tested with the same synthetic data sets, which were used for the EARLINET algorithm inter-comparison exercise. ELPP has been successfully employed for the automatic near-real-time pre-processing of the raw lidar data measured during several EARLINET inter-comparison campaigns as well as during intense field campaigns.

1 Introduction

Lidar networks like EARLINET (European Aerosol Research LIdar NETwork) are powerful tools to investigate the role of the aerosols in a large number of important atmospheric processes (Pappalardo et al., 2014). They can perform coordinated measurements of the vertical profile of aerosol-related optical parameters with high vertical and temporal resolution. Coordinated lidar networks provide observations covering continental and global scales, which allow studies of the long-range transport of aerosol, the establishment of climatologies over large geographical scales, and large-scale monitoring of special events.

In this context, it is particularly important to develop common, automated data analysis tools for all network partners to improve the quality and the homogeneity of the network data. Furthermore, the automated data analysis promotes the near-real-time availability of aerosol related atmospheric parameters. EARLINET is particularly active in supporting such strategies, and several common tools have been implemented to harmonize the network activities (Pappalardo et al., 2014).

One of these tools is the Single Calculus Chain (SCC), a flexible chain of software modules for the automatic analysis of lidar data. A general overview of the SCC is provided by D'Amico et al. (2015). This paper describes ELPP (EARLINET Lidar Pre-Processor), which is the SCC module for the automatic pre-processing of the raw lidar data. The SCC module for the retrieval of aerosol optical properties from the pre-processed data is called ELDA (EARLINET Lidar Data Analyzer) and is described in detail by Mattis et al. (2016). The implementation of ELPP as a unified pre-processor module has been mainly triggered by the heterogeneity of the EARLINET lidar systems. Moreover, ELPP provides a way to standardize all of the instrumental corrections, and the data handling which must be applied to the raw lidar data before they can be used as input for the optical retrieval module. This is fundamental for the application of a rigorous quality assurance program on the lidar data analysis, in which all of the analysis steps starting from the raw lidar data up to the final lidar products (including pre-processing procedures) should be included.

The paper is structured in three main sections. Section 2 describes in detail ELPP. More technical aspects are covered in Sect. 2.1. The main features of the implemented procedures are summarized in Sect. 2.2. The algorithm for the automatic gluing of lidar signals is reported in Sect. 2.3. A description of the error propagation is provided in Sect. 2.4. Finally, the validation of ELPP and the conclusions are in Sects. 3 and 4, respectively.

2   EARLINET Lidar Pre-Processor (ELPP)

The typical SCC analysis scheme comprises two steps (D'Amico et al., 2015): the pre-processing of raw data with ELPP and the subsequent optical processing of the pre-processed lidar data with ELDA. ELPP is based on open source software, and it will be made available on-request to anyone interested to contribute in the development.

By "pre-processing" we mean the set of operations, which must be applied to the raw lidar data before they can be processed by ELDA. ELPP is designed to operate on the lidar data measured by all of the EARLINET lidar systems in a fully automatic way. This is made possible by registering all instrumental parameters needed for the pre-processing in a centralized SCC database (D'Amico et al., 2015), where this information is structured in terms of lidar configurations. Each single lidar system can be linked to several lidar configurations, which describe different lidar set-ups with specialized measurement capabilities (for example day-time or night-time conditions). When a raw measurement is submitted to the SCC, a corresponding entry is created in the SCC database linking the measurement session to the lidar configuration to be used for the analysis. According to that, the raw lidar data are handled and corrected for instrumental effects to provide pre-processed signals that, once saved on a local

storage, can be directly managed by the SCC optical processing module ELDA. The automatic procedures implemented in ELPP do not require the interaction of a human operator to run and to produce the final results. This is a fundamental point in the effort to minimize the manpower needed to perform lidar analysis and consequently to improve the near-real-time availability of the lidar data at network level.

ELPP has been developed as a very flexible and expandable tool: many different lidar configurations can be pre-processed using ELPP in different ways. This is made possible by introducing the concept of SCC usecases. In summary, a usecase represents a procedure to deliver a particular aerosol product like aerosol extinction or backscatter coefficient profiles. Usecases select specific retrieval schemes to calculate the corresponding optical products in both the pre-processing and the optical processing analysis modules. As it will be discussed in Sect. 2.2, each lidar configuration is connected to the retrieval of a specific set of aerosol products. The way in which each product is retrieved is determined by a specific usecase according to the lidar configuration characteristics. The aerosol backscatter coefficient, for example, must be retrieved in different ways depending on the number and the type of available lidar channels. That means, the raw elastic and corresponding nitrogen Raman signals need to be handled in a different way if they are split, for example, in near- and far-range channels, or not. From ELPP development point of view, the implementation of the SCC usecases required the identification of all pre-processing procedures and instrumental corrections adopted within the EARLINET community. All these procedures and corrections were critically evaluated and finally implemented in ELPP, which enables the usage of this tool by all EARLINET systems. More details about SCC usecase are discussed in D'Amico et al. (2015), where a list of all implemented usecases is reported in the Appendix.

The modular structure of ELPP permits an easy implementation of new usecases and thereby the use of ELPP and of the SCC for new EARLINET systems, non-EARLINET systems, and for other lidar networks independent of EARLINET. As a consequence, ELPP plays an important role in making the SCC extensible in more general frameworks like, for example, GALION (GAW Aerosol LIdar Observation Network) as well as in national lidar networks. Typically, in such networks the optical processing retrieval algorithms are the same (or very similar to the) ones used within EARLINET, but lidar experimental configurations may change significantly.

Another key feature of ELPP is the fully traceability of the whole data analysis process. As the input submitted to ELPP/SCC is the raw lidar data without any post-measurement handling, all operations performed on them in pre-processing or processing phases are traceable. All of the corrections, input parameters, and algorithms used to calculate a specific aerosol optical product are logged and provided together with the output data. In this way the end user
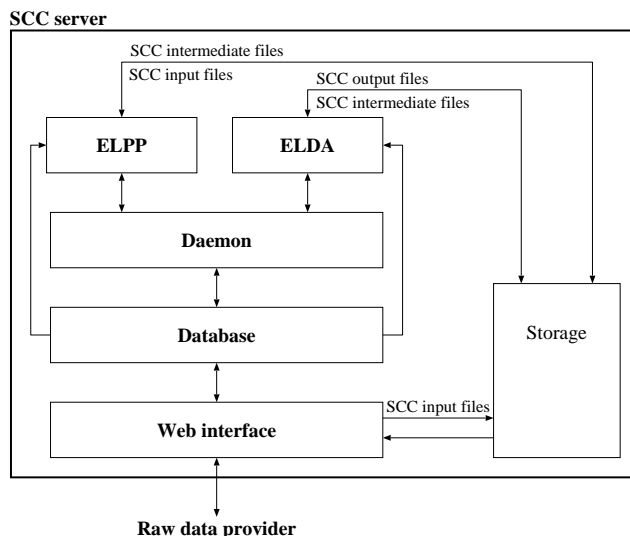
Figure 1. Block structure of the Single Calculus Chain.

has all of the information necessary to fully utilize and to evaluate the product, and it is easy to perform a consistent re-analysis of any data set keeping trace of the history of the analyses made.

Moreover, ELPP provides the possibility to check the quality of all corrections and data handling procedures. The implementation of quality-certified procedures on both pre-processing and processing levels (Böckmann et al., 2004; Pappalardo et al., 2004; Freudenthaler et al., 2016) allows the application of a rigorous and homogeneous quality assurance program for the data measured by lidars with different instrumental characteristics. This is particularly important for a network like EARLINET, where the standardization of aerosol optical products is a fundamental requirement.

ELPP is also an important tool for the long-term sustainability of the SCC. While lidar systems in the research community are often upgraded with new channels or new detection capabilities, ELPP, acting as the interface between the hardware level and the optical retrievals, delivers the pre-processed signals always in the same format. This allows the analysis of the data of a new lidar without modifying any of the other SCC modules.

In the next sub-section we describe the technical aspects of ELPP including its requirements and its different interfaces. After that, in Sect. 2.2, a description of the implemented algorithms will be provided specifying the procedures and the parameters that can be configured for the pre-processing of raw lidar data.

2.1 ELPP technical aspects

ELPP is a command line tool developed in ANSI C. It can be compiled using any C compiler, which is compatible with ANSI C, such as the freely available GNU Compiler Collec-

tion GCC (https://gcc.gnu.org/). The GCC can be used on both 32 and 64 bit environments for a quite large number of processor families and on many popular operating systems like Linux, Unix, Mac OS, and Windows. As ELPP is developed in C, its source code can be implemented on any platform supported by the GCC without any recoding. The main requirements of ELPP are a MySQL database (http://www.mysql.com) and the NetCDF C libraries (http://www.unidata.ucar.edu/software/netcdf/). All of the files used and generated by the SCC are in NetCDF format.

ELPP can be operated as a SCC module and also as a stand-alone module.

When it is used as a SCC module, it is automatically started by a further module (SCC daemon) whenever necessary. Figure 1 shows the general structure of the SCC and also the role played by ELPP in the automatic analysis of raw lidar data.

If used as a stand-alone module, the ELPP executable requires some mandatory command line parameters; i.e. the measurement ID of the lidar observation, that should be pre-processed, and the name of the MySQL database containing all of the instrumental parameters needed by the pre-processing phase. As an optional command line parameter the name of a configuration file can be provided, which contains, for example, the path of the input, output, and log files.

ELPP provides a user-configurable logging system, which produces a log file for each analyzed measurement.

The module ingests a NetCDF file containing a time series of raw lidar data to be analyzed. Raw time series corresponding to different lidar channels can be included in one single input file. The raw lidar data set is a three-dimensional array with the dimensions measurement time, channel, and range bin. It is fundamental that this array contains the raw lidar data as measured and without any modification. In particular, the photon-counting signals should be provided in counts (positive integer numbers) while the analog signals should be provided in mV (real numbers). If the lidar acquisition system provides photon-counting and/or analog raw signals in different units, they need to be converted. The conversion must be applied by the raw data provider before submitting the data to the SCC. As different data acquisition systems provide the raw lidar data in different units, which might even change with system versions, it is not possible to include this conversion step in the SCC. ELPP checks if the photon-counting raw profiles have been submitted using the required units, and if not, the corresponding raw data file is not accepted for the analysis. A further check on theoretical maximum count rate is performed in applying the dead-time corrections as reported in Sect. 2.2.1.

Together with the raw lidar data, more information can be included in the header of the NetCDF input file. In particular, all parameters, which are different for each measurement, can be provided using dedicated NetCDF variables or global attributes. These are, for example, the start and the stop time of each single lidar profile in the time series, the

number of laser shots accumulated for each signal profile, the laser pointing angles, the measurement ID, and other similar parameters. The parameters, which are not related to the individual measurement but to the lidar configuration, like the laser repetition rate, the emitted and detected wavelengths, the channels acquisition mode, and so on, are retrieved from the SCC database. To retrieve the full set of the SCC database parameters relevant for a specific raw data set, each single measurement (i.e. each single NetCDF input file) gets registered in the database, and it is associated to an alpha-numeric string (i.e. measurement ID) which is defined in NetCDF input file. To assure a one-to-one correspondence between each raw data set and the corresponding measurement ID string, it is not allowed to submit NetCDF input file with a measurement ID already present in the database. Using the measurement ID in appropriate database queries, it is possible to retrieve all information needed for the analysis of a specific measurement which is not included in the corresponding NetCDF input file.

The raw data provider should decide how many raw lidar profiles to be included in a single input file taking into account different aspects. The total size of a single NetCDF input file should be less than 200–300 MB to ensure a stable uploading on the SCC server. The maximum time length of a single NetCDF file also depends on the availability of ancillary information to be used in the analysis (i.e. radio sounding profiles, overlap correction functions). Each NetCDF input file can be linked to one specific set of ancillary data sets, which are used for the analysis of the whole time series. If, for example, there are four radio soundings per day available for a certain site, the maximum time length of a single NetCDF input file is 6 h.

The cloud screening is another important operation to be applied to the lidar data before the submission to the SCC. The quality of the SCC optical products can not be assured, if there are signatures of low-level clouds in the raw lidar time series. As a consequence, individual lidar profiles contaminated by low-level clouds should not be included in the NetCDF input file. A new module implementing a fully automatic cloud masking on high resolution lidar data is under development, and it will be included in the SCC in the framework of the ACTRIS-2 (Aerosol, Clouds and Trace gases Research InfraStructure Network) projects (http://www.actris.eu).

As already mentioned, it is possible to provide other kinds of input files to ELPP together with the raw data NetCDF input file, i.e. a file containing pressure and temperature profiles provided by a radio sounding to be used for the calculation of the signal backscattered by atmospheric molecules (as explained in Sect. 2.2.4), a file containing the overlap correction function, and a file consisting of the lidar-ratio profile to be used in the retrieval of the particle backscatter coefficient using elastic-only techniques. Even if these two last files typically are not needed in the pre-processing phase, ELPP interpolates them at the same vertical resolution of the

pre-processed data and saves the corresponding interpolated data in new files. In particular, ELDA is designed to use these files for the retrieval of the aerosol optical properties.

The output files of ELPP are in NetCDF format and contain the pre-processed, range-corrected signals, the so called intermediate files, which were handled according to all analysis steps reported in Sect. 2.2. Table 1 summarizes the description of all NetCDF variables used to identify different types of pre-processed signals in output files. For instance, the total elastic range-corrected signal is represented by the variable elT, the atmospheric nitrogen vibrational–rotational Raman range-corrected signal by vrRN2. The range-corrected signal for the near range and for the far range are represented by variables whose names contain the "nr" or "fr" string, respectively. Selecting appropriate usecases, it is possible to specify whether gluing procedures should be performed by ELPP gluing the raw signals, or by ELDA gluing the optical products. The products calculated from near-range and far-range pre-processed signals are the ones for which ELDA gluing has been selected by the raw data provider (Mattis et al., 2016). All of the NetCDF variables reported in Table 1 are bi-dimensional arrays with dimensions time and range bin. The time and vertical resolutions of these arrays are specified in the SCC database for each product as explained in the next section. Moreover, as the products are defined in the SCC database for a single emission wavelength (i.e. aerosol extinction coefficient at 355 nm or aerosol backscatter coefficient at 1064 nm), each intermediate file refers always to a single wavelength.

Other information included in the ELPP output files is the molecular extinction and the molecular atmospheric transmission profiles, the range resolution and the vertical resolution, the number of averaged laser shots, and so on. All parameters for the optical retrieval, which are provided by the user in the input NetCDF file, are directly transferred to ELDA within the header of the intermediate files.

The input parameters needed for the analysis of the lidar data are retrieved from the header of the NetCDF file or, if not provided in the file header, from a relational MySQL database (SCC database) with general values for a certain lidar system configuration. The structure of this database is described in D'Amico et al. (2015) in Sect. 3.1.

Once ELPP has been started, it is possible to monitor the status of the pre-processing using its return values. ELPP returns a null value if the pre-processing was successfully performed and positive integer values in case any error occurred. Each return value is associated to a specific type of error, such as a failure in gluing the lidar signals or an inconsistency in the definition of the variables in the raw input file, to provide detailed information about the problem occurred.

2.2 Description of implemented algorithms

All corrections and algorithms implemented in ELPP are schematically reported in Fig. 2. Most of them are well

Table 1. Description of different types of pre-processed, range-corrected signals delivered by ELPP and the corresponding NetCDF variable name in the ELPP output files. All of the variables refer to a single emission wavelength. As a consequence, pre-processed data corresponding to different wavelengths are saved in separate files.

| Pre-processed range-corrected signal	NetCDF variable name
Elastic total backscattered signal	elT
Perpendicular* polarization component of the total backscattered signal	elCP
Parallel* polarization component of the total backscattered signal	elPP
Vibrational–rotational Raman backscattered signal by nitrogen molecules	vrRN2
Near-range elastic total backscattered signal	elTnr
Near-range perpendicular* polarization component of the total backscattered signal	elCPnr
Near-range parallel* polarization component of the total backscattered signal	elPPnr
Near-range vibrational–rotational Raman backscattered signal by nitrogen molecules	vrRN2nr
Far-range elastic total backscattered signal	elTfr
Far-range perpendicular* polarization component of the total backscattered signal	elCPfr
Far-range parallel* polarization component of the total backscattered signal	elPPfr
Far-range vibrational–rotational Raman backscattered signal by nitrogen molecules	vrRN2fr

* With respect to the linear polarization state of the incident laser beam.

known and well described in the literature. For this reason the related relevant literature is mentioned in the following sub-sections without providing detailed descriptions of the implemented formulas. On the other hand, details of the implementation and user adjustable parameters are explained. The implementation of the automatic algorithm for the gluing of lidar signals is discussed in greater detail in Sect. 2.3.

As already mentioned in the previous section, ELPP requires the presence of a MySQL database where the characteristics of the analysis to be performed are specified. In particular, starting from a measurement ID passed to ELPP via the command line, it is possible to retrieve from the database all required information such as how many products should be calculated (N_p in Fig. 2), how many lidar channels are needed for the calculation of each product ($N_c(p)$ with $0 < p \leq N_p$), the full set of the input parameters needed for the analysis (dead-time value, trigger delay, etc.), and the name of the data file containing the raw input time series corresponding to all lidar channels linked to the measurement ID to analyze.

Once this information is obtained, ELPP starts to calculate pre-processed signals for all configured products. There are two main loops involved in the pre-processing chain: an external loop on the products to be calculated (index p in Fig. 2 with $p = 1, \dots, N_p$), and an internal loop in which all the product-related channels are pre-processed sequentially (index c with $c = 1, \dots, N_c(p)$). The pre-processing steps performed to calculate a specific set of optical products can be illustrated by means of a practical example. Let us suppose two products – ($N_p = 2$) the aerosol backscatter coefficient (product $p = 1$) and the aerosol extinction coefficient ($p = 2$), which should be calculated for a particular measurement ID by using two elastic channels at 355 nm (elTnr, elTfr) – and two vibrational–rotational N_2 Raman channels at 387 nm (vrRN2nr, vrRN2fr). To calculate the

aerosol backscatter coefficient the channels elTnr (channel $c = 1$), elTfr ($c = 2$), vrRN2nr ($c = 3$) and vrRN2fr ($c = 4$) are needed so $N_c(1) = 4$. Let us also suppose the two near-range channels are detected using analog mode and the two far-range ones are photon-counted. During the loop on index $c = 1, 2, 3, 4$, first, each channel is identified as analog or photon counting, querying the SCC database. Dead-time correction is only applied to photon-counting signals (see Sect. 2.2.1), and a different error propagation is used for analog and photon-counting signals as explained in Sect. 2.4. As a consequence, the elTnr and vrRN2nr are recognized as analog channels while elTfr and vrRN2fr are labelled as photon-counting signals and corrected for dead time. After this step, the following operations are made on the four signals: atmospheric and (optionally) electronic background subtraction as reported in Sect. 2.2.3, trigger-delay correction (see Sect. 2.2.2), and finally the signals are temporally integrated over a time window defined in the SCC database, which is larger than the raw data time resolution. The averaging time window should be selected by the user to ensure the optimal balance between the stability of atmospheric conditions and an adequately high signal-to-noise ratio (SNR). This is particularly important for the analog signals because in this case, as explained in Sect. 2.4, the statistical errors are estimated by the standard error of the mean calculated within the integration time interval. The way in which the error is propagated in the case of the time integration of photon-counting signals is described in Sect. 2.4. When all lidar channels needed for the calculation of the current product (aerosol backscatter coefficient) have been pre-processed, ELPP performs the gluing of near-range and far-range channels. If the gluing of one or more pairs of signals has been configured, the algorithm described in Sect. 2.3 is used for the corresponding signals. According to the example above, two signal gluings need to be performed: the gluing of elTnr with
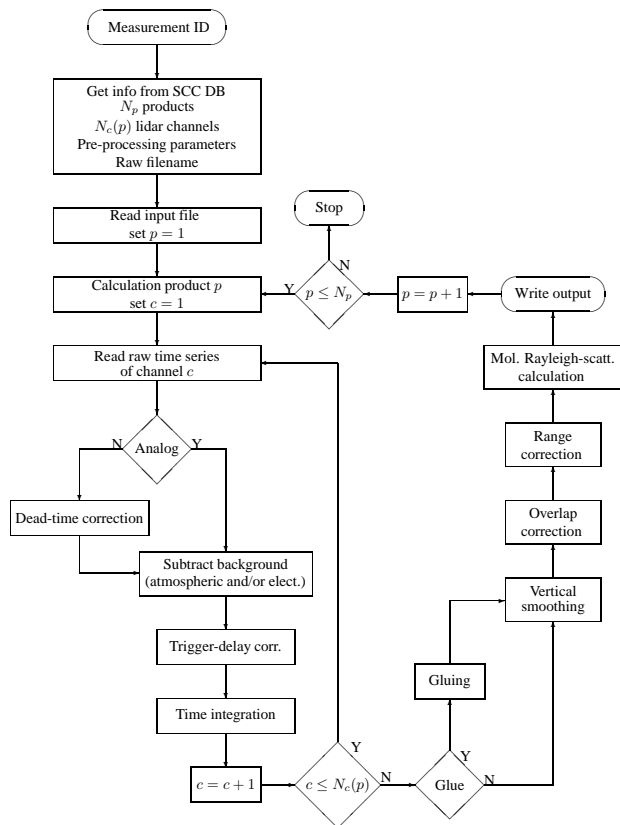
Figure 2. ELPP work flow. ELPP gets the full set of information and parameters needed for the pre-processing of a specific measurement ID performing suitable queries to the SCC database. This set includes how many products should be calculated (N_p), how many lidar channels are needed for the calculation of each product ($N_c(p)$) with $0 < p \le N_p$), all input parameters required for the analysis (dead-time value, trigger delay, etc.), and the name of the input NetCDF file corresponding to the selected measurement ID. There are two main loops involved in the pre-processing chain: an external loop on the products to be calculated (index p) and an internal loop in which all of the product-related channels are pre-processed sequentially (index c). The operations performed by the single blocks are described in the text. A single output file (intermediate NetCDF file) is generated for each product.

elTfr and the gluing of vrRN2nr with vrRN2fr. After this step, ELPP completes the calculation of the current product performing the operations reported on the right part of Fig. 2. Optionally a vertical smoothing of pre-processed lidar signals is performed. Typically, smoothing is done to increase the SNR of the pre-processed signals. Different smoothing options can be selected, like linear, polynomial, and natural cubic spline (Press et al., 2007). Moreover, the signals are range-corrected and optionally corrected for incomplete overlap. Finally, the molecular contribution to the atmospheric extinction and transmissivity are calculated at the same resolution of pre-processed lidar signals as described in Sect. 2.2.4. The pre-processed signals are then stored in

a specific intermediate NetCDF file. This file will be used as input by the ELDA module (D'Amico et al., 2015; Mattis et al., 2016) to retrieve the aerosol backscatter product. In the specific case of the example above, this file contains the time series of the elastic (N_2 Raman) glued pre-processed signals under the variable elT (vrRN2).

Once the pre-processing corresponding to the first product is ended, ELPP switches to the next scheduled product ($p = 2$) which is, according to the example above, the aerosol extinction coefficient. The procedure is very similar to the one already described for the aerosol backscatter coefficient. The only difference is that for this product there are only two signals to be pre-processed (vrRN2nr and vrRN2fr, $N_c(2) = 2$) and only one gluing needs to be performed. The results are stored in another intermediate NetCDF file which contains the time series of the N_2 Raman glued pre-processed signals under the variable vrRN2. ELDA will use this file to retrieve the aerosol extinction coefficient profile.

2.2.1 Dead-time correction

The dead-time correction of photon-counting signals is non-linear. A typical lidar photon-counting channel consists of a photo-multiplier, which ideally generates an electrical pulse for each photon impacting its photo-cathode (event), a pulse discriminator to reduce the noise counts, and finally a fast counter to count the number of events in a fixed interval of time, the time bin. As each electrical pulse has a certain width, two pulses closer to each other than about the pulse width can not be discriminated. The actual minimum time interval between two subsequently countable events, called dead-time τ (Evans, 1955), depends on the setting of the pulse discriminator and on the counting electronics. The dead-time corresponds to a maximum count rate. The dead-time causes a non-linearity between the actual intensity at the photo-multiplier photo-cathode and the counted events, which can be described theoretically by means of photon statistics. As the real processes are not ideal, the mathematical correction of the non-linearity works only in first approximation. Furthermore, there are two models to describe the counting characteristic of a photon-counter, i.e. the paralyzable and the non-paralyzable model. A paralyzable counting system is not able to provide a second output count if a time τ is not elapsed after the previous pulse. Moreover, if an additional pulse arrives within the dead-time τ, the actual dead-time of the system is further extended by τ. In this way, at high count rate, the unit is unable to respond, it is "paralyzed", and the count-rate output is 0. In contrast, a non-paralyzable counter outputs counts at maximum count rate as long as subsequent photon pulses are not discriminable. ELPP includes optionally both models for dead-time correction (in first approximation). The formulas used by the SCC to correct for dead-time are the following (Evans, 1955):

$$c_{\mathrm{m}} = c_{\mathrm{r}} \exp(-\tau c_{\mathrm{r}}) \tag{1}$$

$$c_{\mathrm{m}} = \frac{c_{\mathrm{r}}}{1 + \tau c_{\mathrm{r}}}, \qquad (2)$$

where c_{m} and c_{r} are the measured and the real count rate, respectively. The Eq. (1) refers to a paralyzable counter while the Eq. (2) is used if a non-paralyzable counter is assumed.

Once the dead-time value τ and the model to use for the correction are provided to ELPP, the corresponding photon-counting lidar signal will be automatically corrected by solving the Eqs. (1) or (2) for the unknown c_{r}. The Eq. (1) is solved numerically in the interval $[0, 1/\tau]$ using the well-known secant method (Press et al., 2007).

It is important to underline that the Eq. (1) can be solved only if c_{m} is less or at least equal to the absolute maximum of the exponential function on right-hand side. As a consequence, the following condition on the measured count rate has to be verified:

$$c_{\mathrm{m}} \le \frac{1}{e\tau}, \qquad (3)$$

where e is the Euler's number.

For the non-paralyzable model, the correction for dead time is made by inverting Eq. (2):

$$c_{\mathrm{r}} = \frac{c_{\mathrm{m}}}{1 - \tau c_{\mathrm{m}}}. \qquad (4)$$

As $c_{\mathrm{r}} \ge 0$ and $c_{\mathrm{m}} \ge 0$, the Eq. (4) can be solved only if the following condition on the measured count rate is valid:

$$c_{\mathrm{m}} < \frac{1}{\tau}. \qquad (5)$$

According to the selected model, the condition expressed by the Eqs. (3) or (5) is used as constraint on the actual values of the photon-counting signals rejecting all the cases in which it is not verified.

As the dead-time correction is non-linear, it is applied as the first stage of the pre-processing procedure as shown in Fig. 2.

Here it should be mentioned that in general the reliability of dead-time correction decreases with increasing count-rate: both correction models reported above usually fail in reproducing the correct behaviour of a real counting system at high count-rates. As a consequence, each photon-counting lidar channel should be carefully adjusted to not exceed a maximum count-rate (typically 10–30 MHz depending on the value of τ) in all the range bins for which the photon-counting signal is supposed to be used.

The dead time of a photon-counting system can be evaluated measuring the counting probability distribution generated by a Poissonian source (like a tungsten lamp) as described in Johnson et al. (1966); Whiteman et al. (1992).

2.2.2 Trigger delay

In general, the data acquisition unit of a lidar system gets a trigger from the laser to start the signal recording. Due to the electronic circuits in the laser and in the data acquisition unit, there is always a delay between the outgoing laser pulse and the time at which the acquisition system actually starts to record the lidar profile. If this trigger delay is not properly taken into account, a systematic error is made in associating each lidar range bin with the corresponding atmospheric range. A delay, for example, of 100 ns induces a systematic shift of the atmospheric ranges of 15 m. This shift causes a systematic error in the range-correction of the lidar signal, which propagates to the calculation of the final aerosol properties. The error is especially large for the aerosol extinction coefficient calculated with the Raman method in the near range. The exact trigger delay can be measured and provided to ELPP as input parameter for each lidar channel (Freudenthaler et al., 2016). If ΔT is the trigger delay of a particular lidar channel and $\mathrm{TS}_1 = (t_1, t_2, \ldots, t_{\mathrm{n}})$ is the time scale used by the acquisition system to sample the lidar profile, the actual lidar range scale is calculated from the delayed time scale $\mathrm{TS}_2 = (t_1 + \Delta T, t_2 + \Delta T, \ldots, t_{\mathrm{n}} + \Delta T)$.

If different lidar channels have different trigger delays, ELPP interpolates all recorded lidar signals from the time scale TS_2 (which may change from channel to channel) to the time scale TS_1 (which is the same for all channels). This operation enables the consistent calculation of the lidar products for which multiple channels are needed.

It is possible to choose a linear or a natural cubic spline interpolation (Press et al., 2007). The preferred option is the linear interpolation as usually the trigger-delay correction requires only a time shift of lidar signals. As first step, for each value t_k of the time scale TS_1 the closest higher and lower values of time scale TS_2 are selected. Let us suppose these values are $t_{l-1} + \Delta T$ and $t_l + \Delta T$ respectively. The value of the lidar signal S_{t_k} in t_k is then determined by the equation of the straight line passing through the points $(t_{l-1} - \Delta T, S_{t_{l-1} - \Delta T})$ and $(t_l - \Delta T, S_{t_l - \Delta T})$ as follows:

$$S_{t_k} = S_{t_{l-1} + \Delta T} + \frac{S_{t_l + \Delta T} - S_{t_{l-1} + \Delta T}}{\Delta t} (t_k - t_{l-1} - \Delta T), \quad (6)$$

with $t_{l-1} + \Delta T < t_k \le t_l + \Delta T$ and $\Delta t = t_l - t_{l-1}$ representing the lidar signal range bin width.

If the trigger delay is a multiple of the signal range bin width ($\Delta T = u \Delta t$), the Eq. (6) is equivalent to a re-binning of the signal ($S_{t_l} = S_{t_{l+u}}$). For all the cases in which the Eq. (6) is not equivalent to a re-binning, the implemented trigger-delay correction introduces correlations between neighbour range bins. ELPP takes into account for these correlations estimating the statistical errors of the signal corrected for trigger delay by using the Monte Carlo approach described in Sect. 2.4.

The natural cubic spline interpolation option should be used only if an additional smoothing on lidar signals is required.

2.2.3 Background subtraction

A raw lidar signal $S(z, \lambda)$ can be expressed by Eq. (7),

$$S(z, \lambda) = S_{par}(z, \lambda) + S_{mol}(z, \lambda) + S_{atm}(\lambda) + S_{el}, \tag{7}$$

where $S_{par}(z, \lambda)$ and $S_{mol}(z, \lambda)$ are the signal contributions backscattered by particles (par) and molecules (mol) at altitude z and at wavelength λ. $S_{atm}(\lambda)$ is the optical signal background from the atmosphere, i.e. the sky brightness, which is independent of range, and S_{el} represents the electronic signal background, which stems from electronic effects of the signal detection and data acquisition. S_{el} can have a temporally constant part and a temporally changing part, i.e. changing with lidar range.

It is fundamental to remove $S_{atm}(\lambda)$ and S_{el} from the measured lidar profiles before applying any optical retrieval algorithm.

The amount of the constant background components $S_{atm}(\lambda) + S_{el}$ can be determined either in the far range of the lidar signal, far enough that the expected contribution from atmospheric backscatter is negligible, or in the pre-trigger range before the laser pulse, where the signal must be free of electronic distortions, which could influence the determination of the constant background. In both cases the constant background value is calculated as mean value over signal ranges, which are large enough so that the residual standard error of the mean is negligible.

ELPP implements both options for the calculation of the range-independent contribution in Eq. (7), i.e.

1. the mean of the lidar signal in the far-range region;

2. the mean of lidar signal in the pre-trigger region.

The selection can be done in the SCC database or in the input file.

In the case of option 1, the minimum (z_{min}) and the maximum (z_{max}) ranges (expressed in m) for the background calculation have to be provided in the raw data input file. ELPP estimates the background value from the mean and the corresponding statistical uncertainty from the standard error of the mean of the lidar signal between z_{min} and z_{max}.

In the case of option 2, three parameters are needed: a minimum (i_{min}) and a maximum (i_{max}) range bin index in the pre-trigger region for the calculation of the background value and the uncertainty as above, and a first valid range bin index (i_0) with $i_0 \geq i_{max}$ explained in the following. After the background value and the corresponding statistical uncertainty have been calculated, all points up to i_0 are removed from the lidar signal, because they are not necessary for the further calculations. Then the background is subtracted from the lidar signal.

Table 2. Numerical values of the parameters involved in Eqs. (9) and (10) calculated for the most common lidar wavelengths according to Bucholtz (1995). The quantity δ_n represents the molecular depolarization factor for unpolarized (natural) incident light scattered at right angle, n_S is the refractive index of standard air, L_{mol} the molecular lidar-ratio, and σ_{mol} the total Rayleigh-scattering cross section per molecule given by Eq. (9) when $\rho_{mol} = 1$, and a value of $\rho_S = 2.54743 \times 10^{25}$ m^{-3} for the molecular number density for standard air in Eq. (9) is assumed.

λ [nm]	$\delta_n \times 10^{-2}$	$(n_S - 1) \times 10^4$	$\sigma_{mol} \times 10^{30}$ [m^2]	L_{mol} [sr]
355	3.010	2.9	2.7549	8.503
387	2.953	2.8	1.9188	8.501
532	2.841	2.8	0.5148	8.497
607	2.784	2.8	0.3010	8.494
1064	2.730	2.7	0.0312	8.492

Temporally changing and hence range-dependent contributions in S_{el} are typically due to electronic distortions, which mainly affect the analog lidar signals. They can have temporally random components and components which are synchronal with the repetition of the laser pulse. While the random components zero out in the average of many subsequent lidar signals, the synchronal components do not and can contribute a significant distortion to the lidar signal. The stationary synchronal components can be determined from so-called dark signals, which are measured, for example, with a fully obscured telescope so that no light from the atmosphere reaches the detectors and only the distortions are left. The dark signals have to be averaged over a long enough time period in order to decrease the random contributions sufficiently. ELPP automatically subtracts a dark measurement from the lidar signal if the former is included in the SCC input file as single dark signal or as dark time series. If a dark time series is provided, an average dark profile is calculated automatically and subtracted from the lidar signals.

Both dark signal and background subtraction can be applied together.

2.2.4 Molecular Rayleigh-scattering calculation

In both aerosol backscatter (Klett, 1981; Fernald, 1984; Di Girolamo et al., 1999; Ansmann et al., 1992a; Ferrare et al., 1998) and extinction (Ansmann et al., 1990, 1992b) retrievals the molecular contribution to the atmospheric extinction and transmissivity are required as input, which are calculated by ELPP at the emission and detection wavelengths in terms of vertical profiles at the same vertical resolution as the pre-processed lidar signals. These profiles are used by ELDA in the extinction and backscatter retrievals. The molecular number density profile (ρ_{mol}) is calculated by ELPP from vertical profiles of temperature $T(z)$ and pressure $P(z)$ using the ideal gas law and assuming as 1 the value of the air

compressibility factor (Penndorf, 1957):

$$\rho_{\text{mol}}(z) = \frac{P(z)}{RT(z)},$$ (8)

where R is the universal gas constant.

The temperature and pressure profiles are either calculated from a standard atmosphere model, or taken from the measurements of a close-by radiosounding that can be provided to the SCC as a separate input file. Once the molecular number density is obtained, the calculation of the molecular optical parameters, i.e. the backscatter and extinction coefficients, is done following the procedure reported in Bucholtz (1995) and Miles et al. (2001). In particular, the extinction coefficient (α_{mol}), the lidar ratio (L_{mol}), and the atmospheric transmission (T_{mol}) are calculated using the following formulas:

$$\alpha_{\text{mol}}(\lambda, z) = \frac{24\pi^3}{\lambda^4 \rho_S^2} \left(\frac{n_S^2 - 1}{n_S^2 + 2} \right)^2 \frac{6 + 3\delta_n}{6 - 7\delta_n} \rho_{\text{mol}}(z)$$ (9)

$$L_{\text{mol}}(\lambda) = \frac{8\pi}{3} \left(1 + \frac{\delta_n}{2} \right)$$ (10)

$$T_{\text{mol}}(\lambda, z) = \exp \left(- \int\limits_0^{z/\cos\theta} \alpha_{\text{mol}}(\lambda, \xi) \frac{\text{d}\xi}{\cos\theta} \right),$$ (11)

where λ is the wavelength (in cm), z is the altitude above the lidar station, and θ is the zenith angle of the lidar pointing. The other quantities, which are the molecular number density for standard air (ρ_S), the molecular depolarization ratio for unpolarized (natural) incident light scattered at right angle (δ_n), and the refractive index of standard air (n_S), are calculated according to Bucholtz (1995). The integral in the Eq. (11) is computed numerically using the trapezoidal rule (Press et al., 2007). The numerical values of the parameters involved in the Eqs. (9) and (10) calculated for the most common lidar wavelengths are reported in Table 2. ELPP writes in its output file the quantities given by the Eqs. (9) and (10), and the atmospheric transmission given by Eq. (11) at both emission and detection wavelengths.

2.3 Gluing

Lidar signals can cover a quite large dynamic range, because the intensity of the light backscattered from the aerosol-laden boundary layer in the near range (e.g. at 0.5 km altitude) is several orders of magnitudes higher than the intensity of the light backscattered from the rather clean troposphere (e.g. at 10 km altitude). As it is demanding to cover this large dynamic range with one data acquisition channel with linear response, several approaches are used to overcome this problem.

One option is to split the signal output from a single photomultiplier into two signals and to record one signal using analog detection mode and the other with the photon-counting technique (Whiteman et al., 2006; Newsom et al., 2009). The analog signal provides good performance for the strong backscatter from the near range but suffers from the high analog noise and distortions in the far range. In contrast, the photon-counting signal is saturated in the near range but provides a good performance in the far range. Therefore it is appropriate to use the analog for the near-range signal S_n and the photon-counting for the far-range signal S_f.

Another option is to split the lidar signal optically using a beam splitter and to detect the split components with two detectors and subsequent data acquisitions. Both signals are attenuated, if necessary, with neutral density filters to match the dynamic range of the data acquisitions for the stronger near-range and the weaker far-range signal. In general, the photon-counting technique is used for both signals due to its superior performance regarding detection linearity compared to analog detection.

A third option is to use two (or more) telescopes with separate detection electronics, i.e. one small telescope designed to detect the near-range signal and the other larger telescope optimized to measure the weak far-range signal.

In either case, the complementary signals need to be glued to get a single "extended" lidar signal for the signal analysis (Whiteman et al., 2006; Newsom et al., 2009; Walker et al., 2014).

Before gluing, the near-range and the far-range signals need to be screened for low-level clouds, corrected for instrumental effects like dead time, trigger delay, etc., and the backgrounds have to be subtracted as explained above.

For the first two options the signals are glued by ELPP and then analyzed by ELDA as one signal. Typically, if there are lidar configurations with multiple telescopes, the gluing is made by ELDA at product levels (Mattis et al., 2016) .

ELPP contains a fully automatic algorithm for the gluing of analog and photon-counting signals as well as for the gluing of two photon-counting signals. The algorithm is divided in three main parts. The procedure starts with the determination of a first guess of the gluing region as described in Sect. 2.3.1. After that, the algorithm optimizes the gluing region performing statistical tests as illustrated in Sect. 2.3.2. Finally, the signals are glued in the optimal gluing region as reported in Sect. 2.3.3.

2.3.1 First guess of the gluing region

The first guess of the gluing region uses empirical values. The lower range (z_0) of this region is determined from the far-range photon-counting signal by an upper threshold for the count-rate as long as the dead-time correction (see Sect. 2.2.1) is considered to work reliably. This upper threshold can be defined in the system configuration for each channel in the SCC database. Typical values used for that are 10–30 MHz (Whiteman et al., 2006; Newsom et al., 2009; Walker et al., 2014). The upper range (z_1) of the gluing region is determined from the near-range signal, which can be
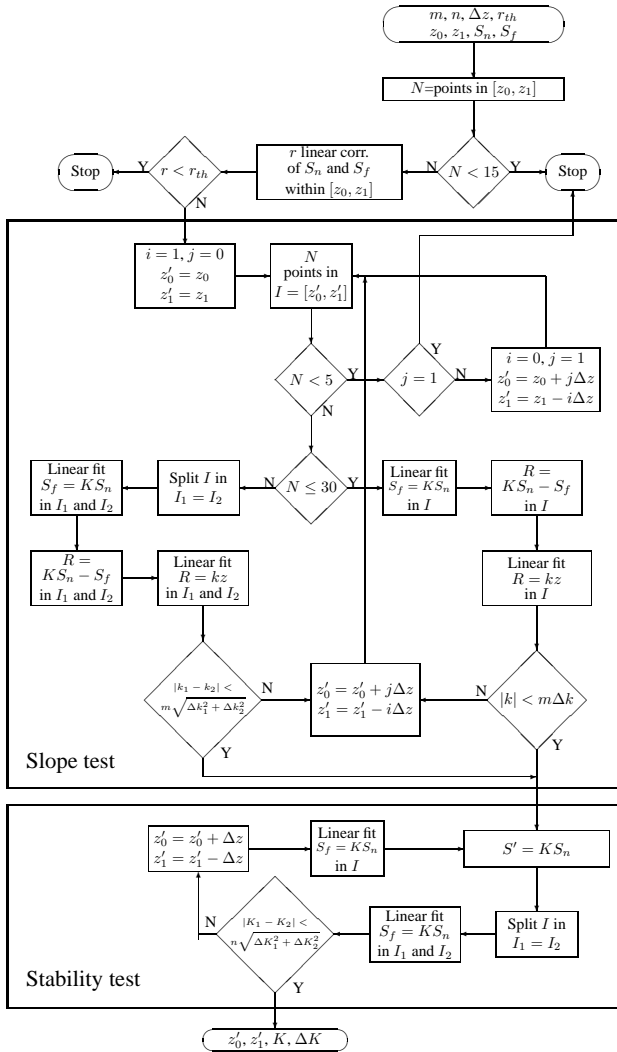
Figure 3. Work flow diagram of the automatic algorithm for the gluing of near-range and far-range lidar signals implemented in ELPP.

an analog or a photon-counting signal. Analog signals are in general measured using pre-amplifiers with several input ranges. Each input range is characterized by a minimum level below which signal distortions and/or the signal noise become significant. This minimum level, which is used to determine the upper altitude (z_1) of the gluing region, is expressed by the ratio S/F where S is the maximum detectable input signal level and F is a parameter characterizing the analog to digital converter (ADC). If we assume, for example, the ADC output is reliable only for values larger than N_{res} times its resolution we obtain

$$F = \frac{2^{n_b} - 1}{N_{res}}, \qquad (12)$$

where n_b is the number of the bits of the ADC. The values of the parameter F can be defined in the system configuration for each channel. If the near-range signal is detected in

photon-counting mode, the upper altitude z_1 is determined by setting a lower threshold for the SNR.

2.3.2　Optimal gluing region

Starting from the values of z_0 and z_1 determined in the previous section, ELPP tries to optimize the gluing region using the automatic algorithm shown in Fig. 3. Besides z_0 and z_1, the algorithm requires the following input data provided in the input file and in the SCC database, which are explained in detail later:

- the near-range and far-range signals S_n and S_f, respectively;

- a threshold r_{th} for the linear correlation of S_n and S_f;

- the step Δz with which the gluing region is decreased during the iterations;

- the statistical uncertainty limits to evaluate the slope test and the stability test given in number of standard deviations m and n, respectively.

First, the algorithm determines the number of range bins N between z_0 and z_1. If this number is less than 15, the gluing region is considered too small to perform a reliable gluing and consequently the gluing is not done. If N is larger than or equal to 15, the linear correlation r of the signals S_n and S_f is calculated between z_0 and z_1. As S_n and S_f should be highly linear correlated in the gluing region, only regions where r is larger than the threshold r_{th} (typically 0.9) are accepted; otherwise the gluing is not performed.

If $r \geq r_{th}$, a further investigation of the gluing region is done in order to exclude parts of the region with significant deviations between the two signals and to minimize the gluing error. This is made by changing iteratively the region $[z_0, z_1]$ until the signals S_n and S_f are consistent according to the additional tests described below. This procedure is illustrated by the block "Slope test" in Fig. 3.

In the optimal gluing region the signals S_n and S_f should coincide, even in the fine structure due to aerosol layers and photon noise, and only differ due to the different electronic noise sources with zero means and slopes. To investigate this the following steps are carried out:

- the signal S_n is normalized to the signal S_f in the gluing region. This is done performing the least square regression $S_f = KS_n$ in the gluing region, and using the obtained K to normalize the signal S_n;

- the residuals $R = KS_n - S_f$ are calculated in the gluing region;

- the slope of R over range z is evaluated making the linear least squares fit $R = kz$.

If the signals S_n and S_f are statistically equivalent in the gluing region, the values of the slope k should not be significantly different than 0, and the residuals R should be normally distributed around a null mean value. This condition is considered verified if the absolute value of k is smaller than m standard deviations (default 2) of the slope resulting from the least square fit.

If the gluing range is large (e.g. if the number of range bins in the gluing range is greater than 30), there could be a difference between the first and the second half of the gluing range. In this case we introduce a constraint on the absolute value of the curvature C of the residuals, which is estimated from the difference of the slopes of the residuals of the first and the second half of the gluing range ($C = |k_1 - k_2|$). The condition is met if

$$C < m \Delta C, \tag{13}$$

where $\Delta C = \sqrt{\Delta k_1^2 + \Delta k_2^2}$. The integer m represents the level of confidence of the Eq. (13) as exclusive condition. For a Gaussian distribution and for $m = 1$, there is about the 32 % of probability the two slopes (k_1 and k_2) agree (in statistical sense) even if the Eq. (13) is not verified (Taylor, 1997). For $m = 2$ the same probability is reduced to about 5 %.

Figure 3 (block "slope test") shows the work flow of the optimization of the gluing region. The starting gluing region $[z_0, z_1]$ is changed until the slope test described above is satisfied. First the algorithm tries to iteratively reduce z_1 in steps of Δz while keeping z_0 fixed. In Fig. 3 this phase starts with setting $i = 1$ and $j = 0$. In each iteration the slope test is evaluated: if the test is passed, the current region is used as optimal gluing region; if it is not passed, z_1 is further reduced by Δz. If there is no region in which the slope test is passed, the algorithm starts to increase iteratively z_0 in steps of Δz while keeping z_1 fixed at its starting value ($i = 0$ and $j = 1$ in Fig. 3). If no region can be found passing the slope test, the gluing is not done.

If a gluing region has passed the slope test, the stability test is further applied, which is shown by the block "stability test" of Fig. 3. The region, which has passed the slope test, is divided into two equal subregions, and in each of these subregions the signal S_n is normalized to the signal S_f, which results in two signals $S_1 = K_1 S_n$ and $S_2 = K_2 S_n$, where K_1 and K_2 are the two slopes obtained from the two least squares line fits. If the gluing region is chosen in a proper way, S_1 and S_2 are indistinguishable taking into account the corresponding signal uncertainties. To test this, the following condition (stability test) is evaluated:

$$|K_1 - K_2| < n \sqrt{\Delta K_1^2 + \Delta K_2^2}, \tag{14}$$

where ΔK_1 and ΔK_2 are the standard deviations on K_1 and K_2 obtained from the two least squares line fits, and n is a positive integer (default value is 1) having the same statistical meaning of the integer m in the Eq. (13). If the condition expressed by the Eq. (14) is met, we assume that the

selected interval is the optimal gluing region, otherwise the interval is progressively reduced increasing (decreasing) the lower (higher) border in step of Δz until the stability test is verified.

2.3.3 Signals combination

If the gluing algorithm described in the previous section ends successfully, the optimal gluing region is returned (z_0' and z_1') together with the normalization gluing factor K used to normalize the signal S_n and the corresponding error ΔK resulting from the least square line fit. Finally, the signals S_n and S_f are glued calculating first the quantity $S_n' = K S_n$ and then calculating the gluing point (z_g) as the range bin, within the optimal gluing region, that minimizes the square differences of the signal S_n' and S_f. The glued signal $S(z)$ and the corresponding statistical error $\Delta S(z)$ are the following:

$$S(z) = \begin{cases} K S_n(z), & \text{if } z < z_g; \\ S_f(z), & \text{otherwise;} \end{cases} \tag{15}$$

$$\Delta S(z) = \begin{cases} \sqrt{(K \Delta S_n)^2 + (S_n \Delta K)^2}, & \text{if } z < z_g; \\ \Delta S_f(z), & \text{otherwise.} \end{cases} \tag{16}$$

An example of the application of this algorithm to real lidar data is shown in Fig. 4. The algorithm is applied to the analog (near range) and photon-counting (far range) elastic cross-polarized signals measured by the EARLINET reference system MUSA (MUlti-wavelength System for Aerosol, Madonna et al., 2011). The blue curve (upper plot) is the photon-counting elastic cross signal at 532 nm summed-up over 1 h, which is used as far-range signal. The first-guess gluing region is indicated as region A in Fig. 4, i.e. between $z_0 = 2445$ m and $z_1 = 3917$ m, and the red curve represents the analog elastic cross signal at 532 nm normalized to the photon-counting signal in region A.

The region indicated with B (extending from 2445 up to 3097 m) is the region in which the slope test has passed, and region C ($z_0' = 2651$ m and $z_1' = 2891$ m) represents the optimal gluing region after the stability test. Region G is used to finally glue the signals. The green curve in Fig. 4 is the same as the red but normalized in region G.

The improvement in gluing the signals in region C instead of the first guess interval A is emphasized by the bottom panel of Fig. 4 in which the relative differences of the two normalized analog signals with respect to the photon-counting profile are shown. In particular, in the region between 2 and 3 km the red signal is clearly below the blue one, which is a clear indication of an unreliable gluing. On the other hand, above 2.5 km, the green signal overlaps the blue one better than the red signal. As a final step, the green and the blue signals are glued at altitude $z_g = 2775$ m.
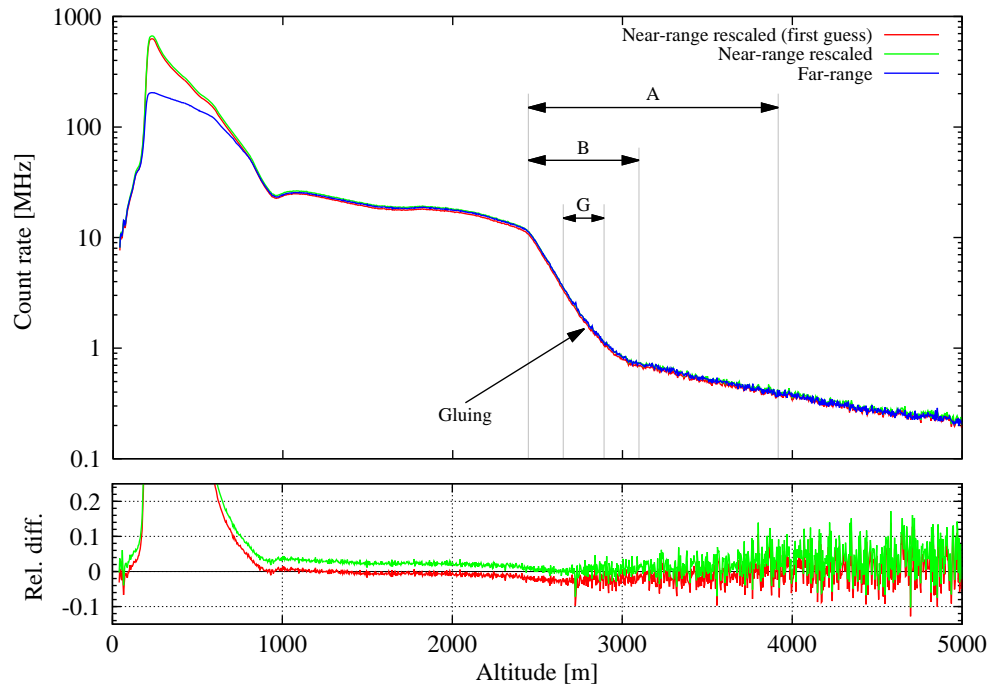
Figure 4. Example of the results of the automatic gluing algorithm shown in Fig. 3. The algorithm is applied to the analog (near range) and photon-counting (far range) elastic cross signals measured at 532 nm by the MUSA lidar of the Potenza station. In blue is shown the photon-counting signal, in red the near-range signal normalized to the photon-counting signal in region A, i.e. the first guess of the gluing region, and in green the near-range signal normalized in region G, i.e. the final optimal gluing region. Region B represents the gluing region obtained after the slope test shown in Fig. 3 and discussed in the text. "Gluing" marks the point at which the blue and green signals are glued. In the bottom plot the relative differences of the two rescaled analog signals with respect to the photon-counting profile are shown.

2.4 Error propagation

ELPP propagates the statistical errors in all steps shown in Fig. 2. Two different propagation methods are implemented: one based on the standard formula of statistical error propagation (Taylor, 1997), and another one based on Monte Carlo simulations (Robert and Casella, 2004), which is only used when the standard error propagation is not possible or too complex. This is the case, for example, if the interpolation or smoothing routines implemented in ELPP have been applied.

The details of the application of the Monte Carlo method to the error propagation are given in Amodeo et al. (2016). In this section only the basic concepts are briefly discussed. If s_i is either a raw or a processed lidar profile, Δs_i the corresponding error profile, and $\mathcal{F}$ a generic operator we want to apply to s_i (for example a smoothing procedure or a filter) to obtain $S_i = \mathcal{F}(s_i)$, the Monte Carlo method offers an efficient and general procedure to calculate ΔS_i, i.e. the uncertainty of S_i. The basic assumption is that each s_i is a mean value with an uncertainty width Δs_i according to a statistical distribution. The first step consists of randomly varying all values s_i considering their Δs_i as standard deviations. ELPP assumes that analog signals are governed by Gaussian statistic and photon-counting signals follow Poissonian statistic. In this way a new synthetic lidar signal s_i' can be generated

according to the assumed probability distribution and a corresponding transformed signal $S_i' = \mathcal{F}(s_i')$ can be calculated. Repeating this procedure a statistically meaningful number of times, the error profile ΔS_i can be estimated calculating the standard deviation of the S_i'. ELPP uses a default value of 30 variations of $S_i' = \mathcal{F}(s_i')$, which has been found to offer the best trade off between the calculation time needed and the accuracy of the retrieved errors. Optionally, the number of Monte Carlo variations can be also specified in the SCC database for each product.

The random extractor routine implemented in ELPP is based on a so-called Lehmer random number generator which returns a pseudo-random number uniformly distributed in the interval 0.0 and 1.0 (Park and Miller, 1988). This uniform distribution is then mapped in Poissonian or Gaussian one (Odeh and Evans, 1974).

ELPP deals with the error propagation of photon-counting and analog signals in different ways. As the photon-counting signals are assumed to obey the Poisson statistic, the statistical error can be evaluated for each photon-counting raw signal range bin as the square root of the corresponding count. As a consequence, the uncertainty of photon-counting signals can be propagated from the beginning to the end of the chain.

On the contrary, the evaluation of the statistical error corresponding to each single raw signal range bin in the case of analog signals is not so trivial. For Gaussian distributions the standard deviation can not be inferred from the mean value like for the Poissonian case. To overcome this difficulty two options are implemented in ELPP.

The first consists of the possibility to provide, along with the raw analog signal time series, the corresponding statistical error time series. This option is applicable only for systems which are able to measure such kind of values e.g. by storing not only the mean values but also the sum of the square values. In this case the error of analog time series is propagated in all the operational blocks shown in Fig. 2 using the standard propagation formula or the Monte Carlo method.

If the statistical error time series are not provided, ELPP calculates the statistical errors of analog signals only after the time averaging (block "time integration" in Fig. 2) as the standard error of the mean of each range-bin value. In all the operations made before the time integration (i.e. background subtraction and trigger-delay correction) the error of analog signals is not propagated due to the difficulty to estimate the statistical error of analog signals without other information. In this case, the analog signal time series (S_h^{a}) and the corresponding standard errors (ΔS_h^{a}) after the time integration are calculated according to the following equations:

$$S_h^{\mathrm{a}}(z) = \frac{1}{N} \sum_{j=Nh}^{N(h+1)-1} s_j^{\mathrm{a}}(z) \tag{17}$$

$$\Delta S_h^{\mathrm{a}}(z) = \sqrt{\frac{\sum_{j=Nh}^{N(h+1)-1}\left[s_j^{\mathrm{a}}(z) - S_h^{\mathrm{a}}(z)\right]^2}{N(N-1)}} \tag{18}$$

$$h \leq \frac{N_t - N}{N}, \tag{19}$$

where $s_j^{\mathrm{a}}(z)$ is the analog time series before the time integration with $j = 0, \ldots, N_t - 1$, and N is the number of the raw profiles belonging to the same time window (defined as the larger integer smaller than the ratio of the integration time window width and the raw time resolution of $s_j^{\mathrm{a}}(z)$ time series).

To summarize, the statistical error of analog signals, if not provided directly by the raw data submitter, are first estimated using Eq. (18) during the "time integration" stage and then propagated in all the subsequent blocks shown Fig. 2.

Finally, in the case of photon-counting detection mode, the signal time series (S_h^{p}) and the corresponding standard errors (ΔS_h^{p}) after the time integration are calculated using the following equations:

$$S_h^{\mathrm{p}}(z) = \sum_{j=Nh}^{N(h+1)-1} s_j^{\mathrm{p}}(z) \tag{20}$$

$$\Delta S_h^{\mathrm{p}}(z) = \sqrt{\sum_{j=Nh}^{N(h+1)-1}\left[\Delta s_j^{\mathrm{p}}(z)\right]^2}, \tag{21}$$

where $s_j^{\mathrm{p}}(z)$ and $\Delta s_j^{\mathrm{p}}(z)$ are the photon-counting time series and corresponding statistical error before the time integration ($j = 0, \ldots, N_t - 1$).

3 Applications and validation

ELPP has been intensively tested with both synthetic and real lidar data to evaluate its performance under different conditions.

The synthetic data set used for testing is the same as used for the algorithm inter-comparison exercise performed in EARLINET (Pappalardo et al., 2004). The data set contains a 30 min time series of synthetic raw lidar signals simulated under realistic experimental and atmospheric conditions. Both elastic and N_2 Raman raw lidar signals are taken into account to reproduce as much as possible a real measurement sample of a typical advanced multi-wavelength Raman lidar. The synthetic raw data were converted to SCC format and then submitted to and processed by the SCC. Finally, the performance of the whole SCC (ELPP and ELDA modules) was evaluated comparing the retrieved optical profiles with the original input profiles. The results of this comparison are discussed in details in Mattis et al. (2016). Here we just point out that all extinction and backscatter profiles retrieved by the SCC from the inter-comparison data set are in good agreement with the input profiles.

In the framework of the EARLINET quality assurance program (Freudenthaler et al., 2016) direct lidar inter-comparison campaigns are used to asses the overall performance of EARLINET lidar systems comparing them with reference lidar systems under different atmospheric conditions. Several inter-comparison campaigns have already been carried out starting from 2009 (Wandinger et al., 2015), during which the SCC format was used as the standard raw signal format and ELPP to provide the pre-processed signals from all participating lidar systems. In this way, all signals were pre-processed and corrected for known instrumental effects with the same procedures, and consequently differences between the signals could be only due to unknown lidar system effects. The use of ELPP during inter-comparison campaigns appeared to be an easy, efficient, and fast way to compare signals from different types of lidar systems. A good example of its flexibility is the EARLI09 inter-comparison campaign in Leipzig, Germany, in May 2009 (Wandinger et al., 2015), where 11 quite different lidar systems from 10 different EARLINET stations took co-located and coordinated measurements during 1 month. ELPP was used to pre-process the raw data and the results from all 11 systems could be made available for comparison just a few hours after the measurements.
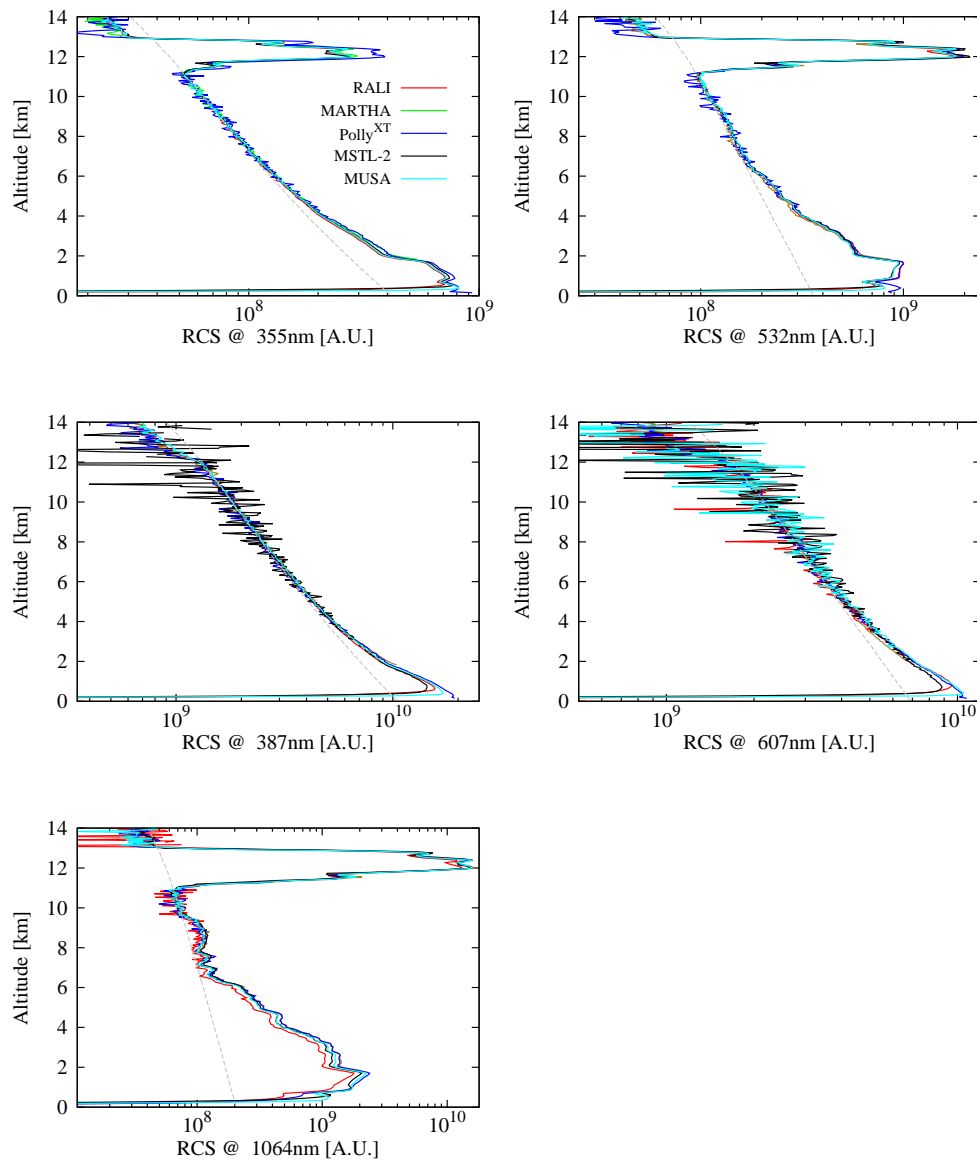
Figure 5. Range-corrected by ELPP for five lidar systems participating in the EARLI09 inter-comparison campaign (the same colour identifies the same lidar system in all the plots). All profiles were taken from 21:00 to 23:00 UT on 25 May 2009. From left to right, upper panel: elastic-backscattered signals at 355 and 532 nm; middle panel: N_2 Raman backscattered signals at 387 and 607 nm; bottom panel: elastic-backscattered signals at 1064 nm. The dotted grey curves represent the signals backscattered by atmospheric molecules computed using a close radiosounding. All signals are normalized in the atmospheric region between 9.5 and 10.5 km, which is assumed to be aerosol free.

To evaluate the SCC performance in analyzing raw data measured by different lidar systems, D'Amico et al. (2015) considered the EARLI09 session taken on 25 May 2009 from 21:00–23:00 UT for which a comparison among the SCC optical products (aerosol backscatter and extinction coefficient profiles) and the corresponding manually retrieved ones is reported. A subset of five EARLI09 participating systems has been selected on the basis of instrumental differences and representativeness within EARLINET comprising the following lidar systems: the Multi-wavelength Raman Lidar

– RALI of the Bucharest station (Nemuc et al., 2013), the MARTHA (Mattis et al., 2004) and the Polly^{XT} (Althausen et al., 2013) systems of the Leipzig station, the MSTL-2 system of the Minsk station (Chaikovsky et al., 2006), and the MUSA of the Potenza station (Madonna et al., 2011). In Fig. 5, we show the ELPP pre-processed signals that have been used as input for ELDA to deliver the EARLI09 optical products compared in D'Amico et al. (2015). The two plots in the upper panel represent the elastic backscattered range-corrected signals at 355 and 532 nm, while the two

plots in the middle panel show the nitrogen inelastic Raman range-corrected signals at 387 and 607 nm. In the bottom panel the elastic-backscattered range-corrected signals at 1064 nm are plotted. In all plots the molecular signals computed by ELPP from a correlative radiosounding are shown (grey dotted line). All range-corrected signals and the calculated molecular backscattered signals have been normalized in the atmospheric region below the cirrus (9.5–10.5 km), which is assumed to be aerosol free. Figure 5 shows the advantages in using ELPP in a lidar inter-comparison campaign: the discrepancies between the range-corrected signals generated by ELPP for different lidar systems can be easily estimated and evaluated. As a consequence, instrumental problems can be quickly detected and the causative misalignments or defects can be fixed. For instance, by using the profiles plotted in Fig. 5, it is possible to select valid and reliable altitude ranges for each channel of each participating instrument (Wandinger et al., 2015).

ELPP has been also successfully used to provide near-real-time pre-processed lidar signals ready to be assimilated in air-quality models. An example of this application is given by the intense period of coordinated measurements performed in July 2012 by 11 EARLINET systems in the Mediterranean area. During this campaign, 72 h of continuous lidar measurements were carried out by all participating systems, and the aerosol products were calculated automatically by the SCC in terms of both pre-processed data and backscatter and extinction profiles were made available in near-real time (Sicard et al., 2015). The pre-processed signals generated by ELPP were assimilated in the air-quality model Polyphemus developed by Centre d'Enseignement et de Recherche en Environnment Atmosphérique (CEREA) allowing a better quality of the PM_{10} and $PM_{2.5}$ forecast on the ground (Wang et al., 2014).

4   Conclusions

ELPP, a fully automatic tool for the pre-processing of lidar data, was developed and extensively tested with both synthetic and real lidar data. It is a fundamental part of the EARLINET SCC because this calculus module generates the input files for the SCC optical processing module (ELDA) starting from raw lidar data. ELPP requires the presence of a MySQL database and of NetCDF libraries both free available from Internet and can be also used as stand-alone module.

Depending on lidar configuration, ELPP applies different type of instrumental corrections and data handling procedures on raw lidar data. The ELPP outputs are NetCDF files containing range-corrected signals ready to be used to retrieve optical parameters like aerosol extinction and/or backscatter coefficient profiles. The output files contain also profiles of atmospheric molecular parameters calculated from the standard model or from the correlative measurement

of pressure and temperature profiles at the same resolution of the range-corrected signals. This information is used by ELDA to retrieve optical results.

The key features of ELPP are the flexibility (it is possible to handle many different kinds of lidar configurations choosing among a quite large number of pre-defined options called usecases), the expandability (it is developed in a modular way, and it is relatively easy to add new system configurations not already covered), and finally it allows the application of a quality assurance program on lidar analysis including also the pre-processing phase. Moreover all calculated products are fully traceable, and all metadata used to produce a specific product can be provided to allow its full evaluation.

ELPP passed the test of EARLINET algorithm inter-comparison exercise providing results in good agreement with the expected ones. It was also extensively tested with real lidar data: during several EARLINET inter-comparison campaigns ELPP was used to provide pre-processed range-corrected signals of all the participating lidar systems in near-real time. As all corrections are made with the same ELPP procedures, the comparison of such pre-processed signals can be used to discover problems or distortions of the individual lidar systems. Finally, the ability of ELPP to deliver pre-processed signals in near-real time during intense field campaigns was successfully tested during the EARLINET 72h operationally exercise performed by 11 Mediterranean EARLINET stations.

A new SCC module devoted to the automatic cloud masking on the raw lidar data is under development and will be implemented in the SCC in the framework of the ACTRIS-2 project (http://www.actris.eu). A big improvement in the automatism of both ELPP and the whole SCC is expected when this new module will be available.

We would like to point out that ELPP is open source, and that the procedures discussed in this paper are the first steps towards a fully automatic, robust, and flexible module for the pre-processing of lidar data. Improvements and enhancements from the lidar community are endorsed and promoted by the current developers.

References

Althausen, D., Engelmann, R., Baars, H., Heese, B., Kanitz, T., Komppula, M., Giannakaki, E., Pfüller, A., Silva, A. M., Preißler,

I., Wagner, F., Rascado, J. L., Pereira, S., Lim, J., Ahn, J. Y., Tesche, M., and Stachlewska, I. S.: PollyNET: a network of multiwavelength polarization Raman lidars, in: Proc. of SPIE, vol. 8894, Lidar Technologies, Techniques, and Measurements for Atmospheric Remote Sensing IX, International Society for Optical Engineering, PO Box 10, Bellingham, WA 98227-0010 USA, 88940I-1–88940I-10, 2013.

Amodeo, A., D'Amico, G., Mattis, I., Freudenthaler, V., and Pappalardo, G.: Error calculation for EARLINET products in the context of quality assurance and single calculus chain, Atmos. Meas. Tech. Discuss., in preparation, 2016.

Ansmann, A., Riebesell, M., and Weitcamp, C.: Measurement of atmospheric aerosol extinction profiles with a Raman lidar, Opt. Lett., 15, 746–748, 1990.

Ansmann, A., Riebesell, M., Wandinger, U., Weitcamp, 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, 1992a.

Ansmann, A., Wandinger, U., Riebesell, M., Weitcamp, C., and Michaelis, W.: Independent measurement of extinction and backscatter profile in cirrus clouds by using a combined Raman elastic-backscatter lidar, Appl. Opt., 31, 7113–7131, 1992b.

Böckmann, C., Wandinger, U., Ansmann, A., Bösenberg, J., Amiridis, V., Boselli, A., Delaval, A., De Tomasi, F., Frioud, M., Grigorov, I. V., Hågård, A., Horvat, M., Iarlori, M., Komguem, L., Kreipl, S., Larchevêque, G., Matthias, V., Papayannis, A., Pappalardo, G., Rocadenbosch, F., Rodrigues, J. A., Schneider, J., Shcherbakov, V., and Wiegner, M.: Aerosol lidar intercomparison in the framework of the EARLINET project. 2. Aerosol backscatter algorithms, Appl. Opt., 43, 977–989, 2004.

Bucholtz, A.: Rayleigh-scattering calculations for the terrestrial atmosphere, Appl. Opt., 34, 2765–2773, 1995.

Chaikovsky, A., Ivanov, A., Balin, Y., Elnikov, A., Tulinov, G., Plusnin, I., Bukin, O., and Chen, B.: Lidar network CIS-LiNet for monitoring aerosol and ozone in CIS regions, in: Proc. of SPIE, vol. 6160, Twelfth Joint International Symposium on Atmospheric and Ocean Optics/Atmospheric Physics, International Society for Optical Engineering, P.O. Box 10, Bellingham, WA 98227-0010 USA, 616035–1–616035–9, 2006.

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.

Di Girolamo, P., Ambrico, P. F., Amodeo, A., Boselli, A., Pappalardo, G., and Spinelli, N.: Aerosol Observations by Lidar in the Nocturnal Boundary Layer, Appl. Opt., 38, 4585–4595, 1999.

Evans, R. D.: The Atomic Nucleus, McGrow-Hill, New York, chapter 28, 785–794, 1955.

Fernald, F. G.: Analysis of atmospheric lidar observations: some comments, Appl. Optics, 23, 652–653, 1984.

Ferrare, R. A., Melfi, S. H., Whitemann, D. N., Evans, K. D., and Leifer, R.: Raman lidar measurements or aerosol extinction and backscattering: 1. Methods and comparisons, J. Geophys. Res., 103, 19663–19672, 1998.

Freudenthaler, V., Linné, H., Chaikovsky, A., Groß, S., and Rabus, D.: Internal quality assurance tools, Atmos. Meas. Tech. Discuss., in preparation, 2016.

Johnson, F. A., Jones, R., McLean, T. P., and Pike, E. R.: Dead-Time Corrections to Photon Counting Distributions, Phys. Rev. Lett., 16, 589–592, 1966.

Klett, J. D.: Stable analytical inversion solution for processing lidar returns, Appl. Opt., 20, 211–220, 1981.

Madonna, F., Amodeo, A., Boselli, A., Cornacchia, C., Cuomo, V., D'Amico, G., Giunta, A., Mona, L., and Pappalardo, G.: CIAO: the CNR-IMAA advanced observatory for atmospheric research, Atmos. Meas. Tech., 4, 1191–1208, doi:10.5194/amt-4-1191-2011, 2011.

Mattis, I., Ansmann, A., Müller, D., Wandinger, U., and Althausen, D.: Multiyear aerosol observations with dual-wavelength Raman lidar in the framework of EARLINET, J. Geophys. Res., 109, D13203, doi:10.1029/2004JD004600, 2004.

Mattis, I., D'Amico, G., Madonna, F., Amodeo, A., and Baars, H.: EARLINET Single Calculus Chain – technical Part 2: Calculation of optical products, Atmos. Meas. Tech. Discuss., in preparation, 2016.

Miles, R. B., Lempert, W. R., and Forkey, J. N.: Laser Rayleigh scattering, Meas. Sci. Technol., 12, R33–R51, 2001.

Nemuc, A., Vasilescu, J., Talianu, C., Belegante, L., and Nicolae, D.: Assessment of aerosol's mass concentrations from measured linear particle depolarization ratio (vertically resolved) and simulations, Atmos. Meas. Tech., 6, 3243–3255, doi:10.5194/amt-6-3243-2013, 2013.

Newsom, R. K., Turner, D. D., Mielke, B., Clayton, M., Ferrare, R., and Sivaraman, C.: Simultaneous analog and photon counting detection for Raman lidar, Appl. Opt., 48, 3903–3914, 2009.

Odeh, R. E. and Evans, J. O.: Algorithm AS 70: The Percentage Points of the Normal Distribution, J. Roy. Stat. Soc. C-App., 23, 96–97, 1974.

Pappalardo, G., Amodeo, A., Pandolfi, M., Wandinger, U., Ansmann, A., Bösenberg, J., Matthias, V., Amiridis, V., De Tomasi, F., Frioud, M., Iarlori, M., Komguem, L., Papayannis, A., Rocadenbosch, F., and Wang, X.: Aerosol lidar intercomparison in the framework of the EARLINET project. 3. Raman lidar algorithm for aerosol extinction, backscatter, and lidar ratio, Appl. Opt., 43, 5370–5385, 2004.

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.

Park, S. K. and Miller, K. W.: Random Number Generators: Good Ones Are Hard To Find, Commun. ACM, 31, 1192–1201, doi:10.1145/63039.63042, 1988.

Penndorf, R.: Tables of the refractive index for standard air and the Rayleigh scattering coefficient for the spectral region between 0.2 and 20.0 μ and their application to atmospheric optics, J. Opt. Soc. Am., 47, 176–182, 1957.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P.: Numerical Recipes, The Art of Scientific Computing, Cambridge University Press, New York, NY, USA, 3 Edn., 2007.

Robert, C. and Casella, G.: Monte Carlo Statistical Methods, 2nd Edn., Springer-Verlag, New York, NY, USA, 2004.

Sicard, M., D'Amico, G., Comerón, A., Mona, L., Alados-Arboledas, L., Amodeo, A., Baars, H., Baldasano, J. M., Bele-

gante, L., Binietoglou, I., Bravo-Aranda, J. A., Fernández, A. J., Fréville, P., García-Vizcaíno, D., Giunta, A., Granados-Muñoz, M. J., Guerrero-Rascado, J. L., Hadjimitsis, D., Haefele, A., Hervo, M., Iarlori, M., Kokkalis, P., Lange, D., Mamouri, R. E., Mattis, I., Molero, F., Montoux, N., Muñoz, A., Muñoz Porcar, C., Navas-Guzmán, F., Nicolae, D., Nisantzi, A., Papagiannopoulos, N., Papayannis, A., Pereira, S., Preißler, J., Pujadas, M., Rizi, V., Rocadenbosch, F., Sellegri, K., Simeonov, V., Tsaknakis, G., Wagner, F., and Pappalardo, G.: EARLINET: potential operationality of a research network, Atmos. Meas. Tech., 8, 4587–4613, doi:10.5194/amt-8-4587-2015, 2015.

Taylor, J. R.: Introduction To Error Analysis, University Science Books, 55D Gate Five Road, Sausalito, CA 94965, USA, 2nd Edn., 1997.

Walker, M., Venable, D., and Whiteman, D. N.: Gluing for Raman lidar systems using the lamp mapping technique, Appl. Opt., 53, 8535–8543, 2014.

Wandinger, U., Freudenthaler, V., Baars, H., Amodeo, A., Engelmann, R., Mattis, I., Groß, S., Pappalardo, G., Giunta, A., D'Amico, G., Chaikovsky, A., Osipenko, F., Slesar, A., Nicolae, D., Belegante, L., Talianu, C., Serikov, I., Linné, H., Jansen, F., Apituley, A., Wilson, K. M., de Graaf, M., Trickl, T., Giehl, H., Adam, M., Comerón, A., Muñoz, C., Rocadenbosch, F., Sicard, M., Tomás, S., Lange, D., Kumar, D., Pujadas, M., Molero, F., Fernández, A. J., Alados-Arboledas, L., Bravo-Aranda, J. A., Navas-Guzmán, F., Guerrero-Rascado, J. L., Granados-Muñoz, M. J., Preißler, J., Wagner, F., Gausa, M., Grigorov, I., Stoyanov, D., Iarlori, M., Rizi, V., Spinelli, N., Boselli, A., Wang, X., Lo Feudo, T., Perrone, M. R., De Tomasi, F., and Burlizzi, P.: EARLINET instrument intercomparison campaigns: overview on strategy and results, Atmos. Meas. Tech. Discuss., 8, 10473–10522, doi:10.5194/amtd-8-10473-2015, 2015.

Wang, Y., Sartelet, K. N., Bocquet, M., Chazette, P., Sicard, M., D'Amico, G., Léon, J. F., Alados-Arboledas, L., Amodeo, A., Augustin, P., Bach, J., Belegante, L., Binietoglou, I., Bush, X., Comerón, A., Delbarre, H., García-Vízcaino, D., Guerrero-Rascado, J. L., Hervo, M., Iarlori, M., Kokkalis, P., Lange, D., Molero, F., Montoux, N., Muñoz, A., Muñoz, C., Nicolae, D., Papayannis, A., Pappalardo, G., Preissler, J., Rizi, V., Rocadenbosch, F., Sellegri, K., Wagner, F., and Dulac, F.: Assimilation of lidar signals: application to aerosol forecasting in the western Mediterranean basin, Atmos. Chem. Phys., 14, 12031–12053, doi:10.5194/acp-14-12031-2014, 2014.

Whiteman, D. N., Melfi, S. H., and Ferrare, R. A.: Raman lidar system for the measurement of water vapor and aerosols in the Earth's atmosphere, Appl. Opt., 31, 3068–3082, 1992.

Whiteman, D. N., Demoz, B., Rush, K., Schwemmer, G., Gentry, B., Di Girolamo, P., Comer, J., Veselovskii, I., Evans, K., Melfi, S. H., Wang, Z., Cadirola, M., Mielke, B., Venable, D., and Van Hove, T.: Raman Lidar Measurements during the International H_2O Project. Part I: Instrumentation and Analysis Techniques, J. Atmos. Oceanic Tech., 23, 157–169, 2006.