https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

# Collaborative development of the Lidar Processing Pipeline (LPP)

Juan Pallotta[1], Silvânia A. Carvalho[2], Fabio Lopes[3], Alexandre Cacheffo[4], Eduardo Landulfo[3], and Henrique M. J. Barbosa[5,6]

[1]Centro de Investigaciones en Láseres y Aplicaciones, UNIDEF (CITEDEF-CONICET), Buenos Aires, Argentina.
[2]Universidade Federal Fluminense, Volta Redonda, Brazil
[3]Instituto de Pesquisas Energéticas e Nucleares, São Paulo, Brazil
[4]Universidade Federal de Uberlândia, Ituiutaba, Brazil
[5]University of Maryland Baltimore County, Baltimore, United States
[6]Universidade de São Paulo, São Paulo, Brazil

**Correspondence:** Juan Pallotta (juanpallotta@gmail.com)

**Abstract.** Lidars can simultaneously measure clouds and aerosols with high temporal and spatial resolution and hence help understand their interactions, which are the source of the largest uncertainties in current climate projections. However, lidars are typically custom-built, so there are significant differences between them. In this sense, lidar networks play a crucial role as they coordinate the efforts of different groups, providing the guidelines for quality-assured routine measurements aiming to homogenize the physical retrievals. With that in mind, this work describes an ongoing effort to develop a lidar processing pipeline (LPP) collaboratively. The LPP is a collection of tools developed in C/C++, python, and Linux script that handle all the steps of a typical lidar analysis. The first publicly released version of LPP produces data files at levels 0 (raw and metadata), 1 (averaging and layer-mask), and 2 (aerosol optical properties). We discussed the application of LPP for two case studies for Sao Paulo and Amazon, which shows the capabilities of the current release but also highlights the need for new features. From this exercise, we developed and presented a roadmap to guide future development, accommodating the needs of our community.

## 1 Introduction

Aerosols, clouds, and their interactions are the source of the largest uncertainties in current climate change estimates (IPCC, 2013, 2021). More frequent and higher quality measurements of aerosol, clouds, and the physical processes governing their link with climate are needed to reduce these uncertainties (Mather, 2021; National Academies of Sciences, Engineering, and Medicine, 2018), and lidars are a powerful instrument to accomplish this task. However, lidars are generally developed by individual groups for particular applications; hence the characteristics of their hardware differ in essential aspects, such as focal length, emitted and detected wavelengths, polarization capability, and signal-to-noise ratio, to name a few. Even in the realm of single-wavelength elastic lidars, typical differences between custom-built lidar systems are large enough to require a careful, dedicated analysis of their datasets. In this sense, lidar networks play a crucial role as they coordinate the efforts of different groups, providing the guidelines for quality-assured routine measurements on a regional scale (Antuña-Marrero et al., 2017). Moreover, a coordinated effort is of utmost importance to homogenize the physical retrievals from the highly non-uniform instruments in lidar networks. This homogenization is only possible by developing a unified processing pipeline

# Summary of Comments on gi-2022-19-commented.pdf

## Page: 1

Number: 1 Author: Date: 31.01.2023 13:35:55
? authors?

Number: 2 Author: Subject: [redacted] [redacted]: 31.01.2023 11:05:41
t

Number: 3 Author: Subject: KDate: 31.01.2023 13:36:18
reference for this statement? E.g.: doi:10.5194/amt-9-1001-2016

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions
Open Access

that accounts for the hardware heterogeneity in the pool of instruments, as it has been accomplished recently in the context of EARLINET (D'Amico et al., 2015) and AdNet (Sugimoto and Uno, 2009). Homogeneous networks also rely on uniform calibration and data processing routines, like the NASA Micro Pulse Lidar NETwork (MPLNET) (Welton et al., 2001) or the Italian Automated LIdar-CEilometer network (ALICEnet) (Dionisi et al., 2018).

The Latin America Lidar Network (hereafter Lalinet) is a Latin American coordinated lidar network to obtain extensive and intensive aerosol optical properties profiled in the atmosphere. This federative lidar network aims to establish a consistent and statistically sound database to enhance the understanding of aerosol distribution over Latin America and its direct and indirect influence on climate. There are currently 12 stations in 6 countries, most of which are equipped with tropospheric aerosol lidars measuring one more elastic channels; only a few systems can measure inelastic Raman channels for $N_2$ and $H_2O$. More details about the network can be found in Landulfo et al. (2020).

In recent years, the Lalinet network has worked towards establishing routine quality-assurance tests and intercomparing the retrieval algorithms used by the different groups (Guerrero-Rascado et al., 2016; Barbosa et al., 2014). Here, we describe the Lidar Processing Pipeline (LPP), an ongoing coordinated effort to homogenize the physical retrievals from different instruments in the Lalinet network, only possible with the development of a processing pipeline that accounts for the hardware heterogeneity in the pool of instruments. This manuscript describes the tools developed in C/C++, Python, and Linux scripts to handle all the steps of a typical lidar analysis. The modular framework is generic enough to be applicable to any lidar instrument or network. The code is open source and is available on GitHub. The first public release of LPP includes a layer mask (clouds or aerosol plumes without discrimination), molecular calibration, and aerosol backscatter retrievals for elastic lidar signals. This manuscript is organized as follows. Section 2 presents the overall concept of LPP, the algorithms, and the structure of the different levels of output files. Sample results from the application of LPP to actual lidar data from two Lalinet stations are presented in section 3. Future perspectives and conclusions are given in section 4.

## 2 PROCESSING PIPELINE

The Lidar Processing Pipeline (LPP) is being developed in a partnership between the lidar groups of the Latin American Lidar Network. The LPP reads a series of raw data files in the standard Licel format and produces three NetCDF files corresponding to data levels 0, 1, and 2. The processing pipeline has three main modules responsible for data processing at each level. These modules are independent, and the whole pipeline can be automated with a Linux shell script, or each module can be run directly in a Linux terminal.

In the current version, each module is a C/C++ executable that uses a shared library of lidar analysis routines. Future releases of LPP will include python modules, which will facilitate the customization of the pipeline to fulfill the needs of the different groups, and new developments of novel analysis techniques that could be later ported to C/C++, which is faster for operational applications. Each executable receives three mandatory arguments: (1) an input, (2) an output, and (3) a configuration file. For the level 0 module, the input is the path to the directory with lidar files to be processed, and the output is the path for the level

# Page: 2

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 11:10:24
phrasing. One cannot measure inelastic channels

Number: 2 Author: Date: 31.01.2023 11:09:58
?

Number: 3 Author: Subject: [redacted] Date: 31.01.2023 11:10:53

Number: 4 Author: Subject: [redacted] Date: 31.01.2023 11:11:42
what does it mean?

Number: 5 Author: Subject: [redacted] Date: 31.01.2023 11:13:31
what is this? please add reference.

Number: 6 Author: Subject: [redacted] Date: 31.01.2023 11:14:35
very technical

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions

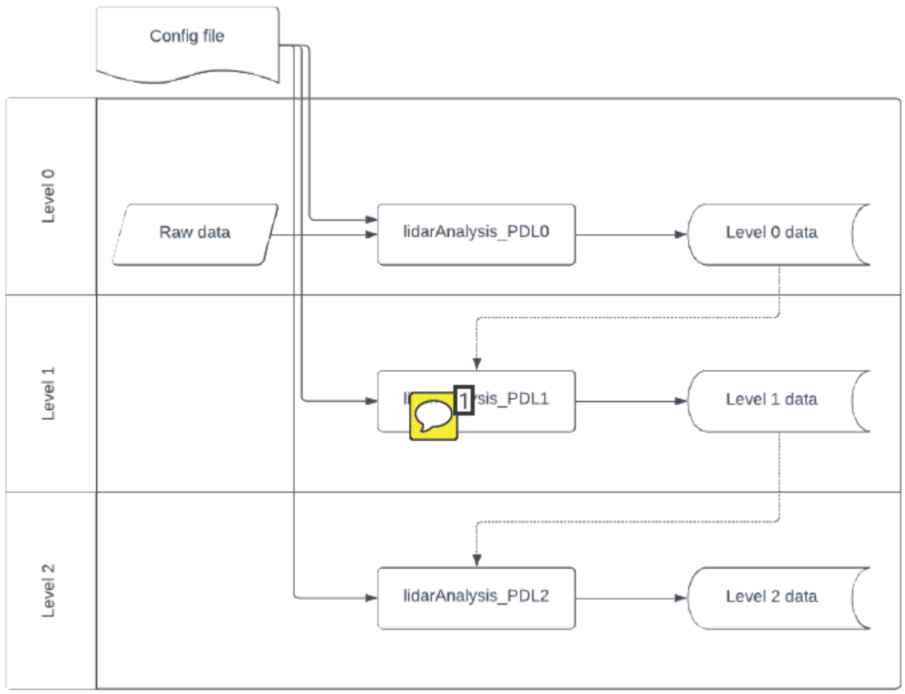


**Figure 1.** Flow diagram showing the structure of LPP version 1.0.0.

NetCDF file. The input for the level 1 module is the path for the level 0 NetCDF file produced in the previous stage, and the output is the path for the level 1 NetCDF file. Similarly for the level 2 module, as indicated in Figure 1.

The output file of a given level (e.g., 2) contains all the content of the previous level (e.g., 1) in addition to the new information generated in that level of data processing. In other words, all the information used to process the data to a given level is available in the corresponding file, thus allowing its reprocessing if needed. Figure 2 shows the content of a level 2 data file. Dimensions and variables are explained in detail in the LPP's documentation on the GitHub repository (see section 2.3 Open Science Development). The following sections explain each of the data levels in detail.

## 2.1 Data level 0 (L0)

The main goal of the level 0 data processing is to convert data from the standard Licel format to NetCDF format. The level 0 module dumps all information from a series of raw lidar files into a single output file with additional metadata. The single values in the header of the raw files become arrays indexed in time. All the information from the file headers is saved, including but not limited to: filename, site, start and stop time, altitude, latitude and longitude, zenith and azimuth, accumulated pulses, repetition rate, and the number of channels. For each channel, the following information is saved: channel ID, polarization state, type (elastic/Raman), number of height bins, PMT voltage, and wavelength. The number of bits and range of ADC for analog channels are also recorded, while for photon counting channels, the discriminator level is recorded. For each channel,

# Page: 3

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 11:16:21
not understandable with current description. nor in caption or text. E.g. what is PDL2?

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 11:15:05
very, very technical

Number: 3 Author: Subject: [redacted] Date: 31.01.2023 11:17:28
as stated below, need to be published so that it is available for many years. Best with a DOI. otherwise the whole publication is worthless.

Number: 4 Author: Subject: [redacted] Date: 31.01.2023 11:18:43
I know what you mean, but it might not be obvious for each user. Maybe add a table describing all these variables/attributes

Number: 5 Author: Subject: [redacted] Date: 31.01.2023 11:19:02
not explained

Number: 6 Author: Subject: [redacted] Date: 31.01.2023 11:19:25
what is this?

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022




| Name | Long Name | Type |
|---|---|---|
| lidar_signals_L0_L1_L2.nc | lidar_signals_L0_L1_L2.nc | Local File |
| Accumulated_Pulses | Accumulated Pulses | 1D |
| ADC_Bits | ADC Bits | 1D |
| Azimuth | Azimuth | 1D |
| Bkg_Noise | Bkg Noise | 2D |
| DAQ_Range | DAQ Range | 1D |
| DAQ_type | DAQ type | 1D |
| L1_Data | L1_Data | — |
| Azimuth_AVG_L1 | Azimuth AVG L1 | 1D |
| Cloud_Mask | Cloud Mask | 2D |
| Laser_Zero_Bin_Offset | Laser Zero Bin Offset | 1D |
| MaxRangeAnalysis | MaxRangeAnalysis | 1D |
| Molecular_Density | Molecular Density | 1D |
| Pressure_Ground_Level | Pressure Ground Level | 1D |
| Raw_Lidar_Data_L1 | Raw Lidar Data L1 | 2D |
| Start_Time_AVG_L1 | Start Time AVG L1 | 1D |
| Stop_Time_AVG_L1 | Stop Time AVG L1 | 1D |
| Temperature_Ground_Level | Temperature Ground Level | 1D |
| Zenith_AVG_L1 | Zenith AVG L1 | 1D |
| L2_Data | L2_Data | — |
| Aerosol_Backscattering | Aerosol Backscattering | 2D |
| Aerosol_Extinction | Aerosol Extinction | 2D |
| AOD_LR | AOD LR | 2D |
| Azimuth_AVG_L2 | Azimuth AVG L2 | 1D |
| LRs | LRs | 1D |
| MaxRangeAnalysis | MaxRangeAnalysis | 1D |
| Range_Corrected_Lidar_Signal_L2 | Range Corrected Lidar Signal L2 | 2D |
| Start_Time_AVG_L2 | Start Time AVG L2 | 1D |
| Stop_Time_AVG_L2 | Stop Time AVG L2 | 1D |
| Zenith_AVG_L2 | Zenith AVG L2 | 1D |
| Laser_Source | Laser Source | 1D |
| Number_Of_Bins | Number Of Bins | 1D |
| PMT_Voltage | PMT Voltage | 1D |
| Polarization | Polarization | — |
| Raw_Data_Start_Time | Raw Data Start Time | 1D |
| Raw_Data_Stop_Time | Raw Data Stop Time | 1D |
| Raw_Lidar_Data | Raw Lidar Data | 2D |
| Wavelengths | Wavelengths | 1D |
| Zenith | Zenith | 1D |

**Figure 2.** Output NetCDF file inspected with Panoply software (Panoply). Product data levels 1 and 2 are stored in subgroups named ***LX Data***. Data level 0 information is stored in the root tree of the file.

the raw ADC readings are saved as 2-dimensional arrays indexed in height and time. In the current version, the values are not converted from integer to mV (analog) or counts (photon counting). All this information is saved in the root tree of the netCDF file (see Figure 2). In future versions, global information about the lidar system, such as the person in charge, contact, and hardware features, will be saved as global attributes. All this data will be kept in the files generated in the following data levels.

# Page: 4

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 12:41:34
what does this mean? what is stored in the raw data?

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions

### 2.2 Data level 1 (L1)

The main goal of level 1 data processing is to apply the necessary corrections on the lidar profiles and compute a layer mask, which usually requires accumulating multiple profiles to increase the averaging and signal-to-noise ratio. The level 1 module receives a single level 0 data file as input and produces a single level 1 NetCDF file as output, with all new L1 information added to the subgroup L1 Data (see Figure 2). L1 data includes updated start and stop times for each averaged profile and time-averaged values for zenith, azimuth, and lidar signal.

In the current version, the following corrections are implemented, based one Guerrero-Rascado et al. (2016) and Freudenthaler et al. (2018). The first is the trigger delay correction. There is typically a delay between the emission of the laser pulse and the start of the data acquisition, resulting in a vertical displacement of the measured signals, which affects both analog and photon-counting channels. If the zero-bin (analog) and bin-shift (photon-counting) tests have been performed, the number of bins that need to be shifted upward or downward can be informed in the configuration file, and the correction will be applied. Second, there is the dark current correction. This accounts for signal distortions that are not due to the atmosphere but the lidar system itself. Typical examples are transient peaks from firing the laser flashlamp and time-dependent electronic noise in analog channels. If a dark current test has been performed, a file with this information can be provided, and the dark current will be subtracted from the measurements. Third, there is the background correction to remove the scattered solar radiation entering the telescope, which is unrelated to the lidar signal. The background correction can be performed by averaging the signal in the far range or by a molecular fit.

Besides the corrections and time averaging, L1 data also includes a molecular density profile calculated from the input pressure and temperature profiles. Finally, the L1 data processing creates a layer-mask product based on dynamic thresholds calculated at the time-average resolution of level 1. Figure 3 gives an example of the range-corrected lidar signal (RCLS) and the corresponding layer mask for measurements at the lidar station of Sao Paulo on 30/07/2021.

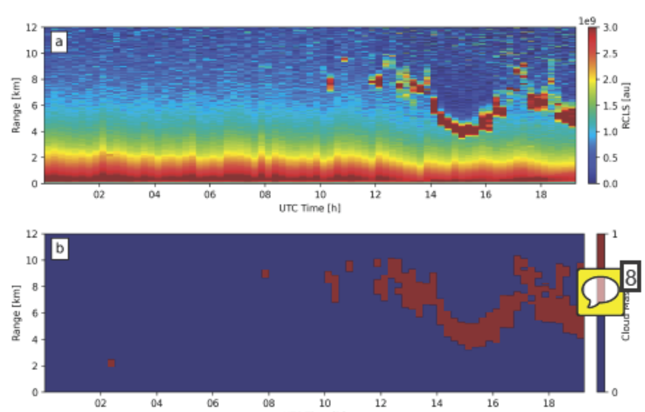


**Figure 3.** Color map of the range-corrected lidar signal (a) and its cloud-mask (b) from the L1 data processing for the Sao Paulo lidar station, on 30 July 2021.

# Page: 5

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 12:42:18

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 12:43:04
you never explained what this is.

Number: 3 Author: Subject: [redacted] Date: 31.01.2023 12:43:40
phrasing

Number: 4 Author: Subject: [redacted] Date: 31.01.2023 12:44:44
that's what I do not understand. Any reference?

Number: 5 Author: Subject: [redacted] Date: 31.01.2023 12:45:01
where do they come from?

Number: 6 Author: Subject: [redacted] Date: 31.01.2023 13:41:38
what does this mean? any reference? Especially, what do you mean with dynamic thresholds?

Number: 7 Author: Subject: [redacted] Date: 31.01.2023 12:46:04
ever explained?

Number: 8 Author: Subject: [redacted] Date: 31.01.2023 12:46:37
what does it mean? 0 and 1? I assume what it is but it is not written.

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions

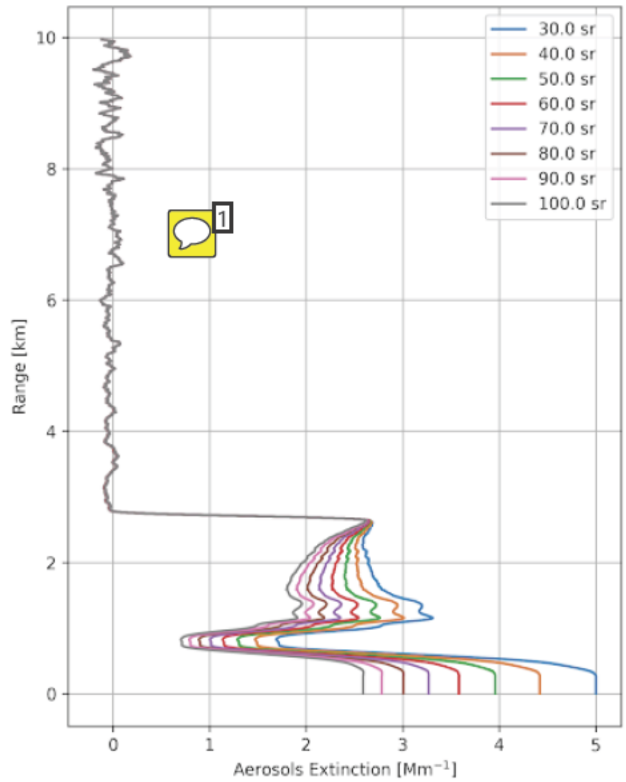


**Figure 4.** Extinction coefficient obtained using a set of pre-defined lidar ratios by the L2 data processing for the Sao Paulo lidar station, at 12:53 UTC on 14 September 2021. Extinction coefficient was assumed constant below 300 m.

### 2.3 Data level 2 (L2)

The main goal of level 2 data processing is to obtain the optical profiles of aerosol particles' optical properties. The level 2 module receives a single level 1 data file as input and produces a single level 2 NetCDF file as output, with all new L2 information added to the subgroup L2 Data (see Figure 2). LPP allows the time averaging at L2 data to be different from that of L1 data. Therefore, L2 data includes updated start and stop times for each average profile and time-averaged values for zenith, azimuth, and lidar signals.

In the current version of LPP, the optical properties can be obtained from one of the analog elastic channels at a time and include the aerosol extinction and backscatter profiles. The configuration file sets the lidar ratio (LR) used for the inversion with the Fernald method (Fernald, 1984). When multiple LR values are given, the inversion is performed for each LR value, producing a set of profiles of optical properties. The aerosol optical depth (AOD) is calculated assuming the extinction to be constant below a specific range (defined in the setting file), where the incomplete overlap precludes its calculation. Figure 4 shows an example of multi-LR optical retrieval for a lidar signal from Sao Paulo lidar.

# Page: 6

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 13:41:59

how is the reference height determined? In this example, it looks like there is a negative bias in the free troposphere. Please comment!

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 12:47:16

phrasing

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions

In future releases, an option to automatically download Aeronet data will be implemented. This will allow using the column AOD to constrain the retrieval and obtain an estimation of the LR in the L2 datasets.

### 2.4 Open Science Development

The development of LPP follows an open-science approach. It is based on cooperative work and uses new digital technologies to diffuse knowledge and allow others to collaborate and contribute. This first released version reported here is ready to use. It can be obtained from GitHub (see *Code availability*) under an MIT license that enables the reuse, redistribution, and reproduction of all methods. Besides the three LPP main modules, this repository also includes sample configuration files, shell scripts for automating the operation, sample lidar data files, and detailed instructions on using LPP.

## 3 RESULTS AND DISCUSSION

We show LPP’s first results for two lidar systems from the LALINET network: a station in São Paulo, the largest metropolitan area in South America, and a station in Manaus, in the central Amazon rainforest. The case studies chosen for each site are based on each lidar characteristic, exploiting and highlighting the features of the current version of LPP.

### 3.1 SPU Lidar - Brazil - 14 September 2020: Boundary Layer Aerosols

The lidar deployed at Sao Paulo (23[1]56’ S, 46[2]74’ W, 740 m above sea level) is a multiwavelength Raman LIDAR operated by the Lasers Environmental Applications Research Group at the Center for Lasers and Applications (CLA), Nuclear and Energy Research Institute (IPEN) (Landulfo et al., 2020). It is a monostatic coaxial system, vertically pointed to the zenith and using a commercial Nd:YAG laser by Quantel, model Brilliant B at a repetition rate of 10 Hz. The output energy per pulse is 850 mJ for 1064 nm, 400 mJ for 532 nm, and 230 mJ for 355 nm. A 300 mm diameter telescope with a 1.5 m and 0.1 field-of-view (FOV) is used as a collection system, reaching a full overlap at 300 m above ground level. The detection box collects six different wavelengths: elastic 355 and 532 nm with the corresponding shifted Raman signals from nitrogen: 387 and 530 nm respectively. Also, the water vapor line at 408 nm and the elastic from 1064 nm. The electronic acquisition system is a Licel transient recorder, registering the lidar signals in its native lidar data files format, which is compatible with the LPP software.

The day chosen for the case analysis is 14 September 2020. It was cloudless for most of the day, with some high clouds at 7 PM UTC. For data level 1, the wavelength used for the layer mask was 532 nm, and no time averaging was applied for data level 1 products (each raw lidar data has a 1 min acquisition time). Figure 5 shows the colormaps of the range-corrected lidar signal (RCLS) and its cloud layer mask from the L1 output file of LPP:

Data level 2 analysis was performed for the cloudless profiles between 15 h to 19 h (UTC) when AERONET data was available for comparison. LPP was configured to use the 532 nm elastic channel, with an average of 10 minutes (10 files) and LR values ranging from 30 to 100 sr, with steps of 5 sr (15 values). LPP calculated the AOD values integrating each of the extinction profiles, assuming it to be constant below 300 m.

Page: 7

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 12:55:49

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 12:55:53

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022




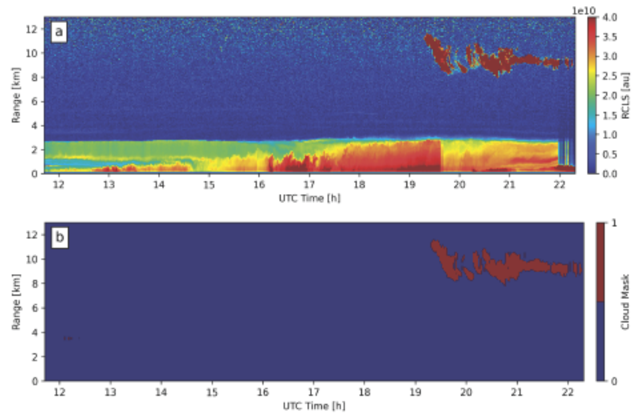


**Figure 5.** Level 1 data products showing the (a) RCLS and (b) cloud layer-mask from measurements at the Sao Paulo lidar station on 14 September 2020.

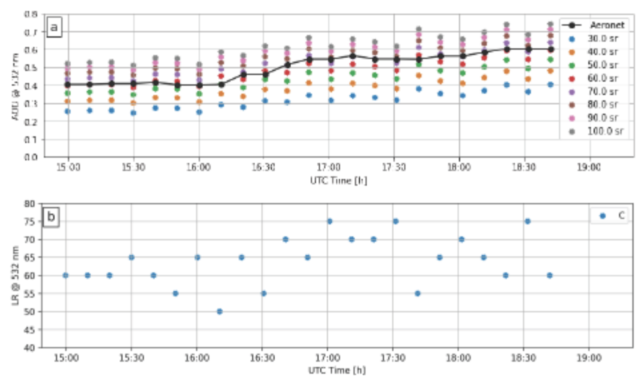


**Figure 6.** (a) AOD's multi-LR lidar product and with the closest Aeronet value and (b) LRs derived from each inversion.

Figure 6a shows the AODs obtained from the multi-LR inversion for each averaged profile that can be compared with the AERONET retrievals shown in black. Using 15 LRs allows searching for which LR produces the closest AOD value measured by AERONET. The time series of the optimal LR values is shown in Figure 5b. Most of the values are between 50 and 75 sr, with a standard deviation of 6.9 sr, well within the range of values expected for urban aerosols (Ansmann and Müller, 2005).

The current version of LPP produces level 2 data of the elastic multi-LR inversion, with the extinction, backscattering profiles, and the AODs for each LR. Future versions of LPP will include an optimization routine to find the LR value that gives precisely the reference AOD from AERONET. This could either be downloaded automatically from the AERONET website or informed by the user on the configuration file.

# Page: 8

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 13:00:37

ok, that sounds reasonable, but not only the LR but also the reference height and value are important. How are these determined?

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 13:01:36

this is indeed very good, but I know at least two publications where such retrievals have been done and presented. So it is not new.

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems
Discussions

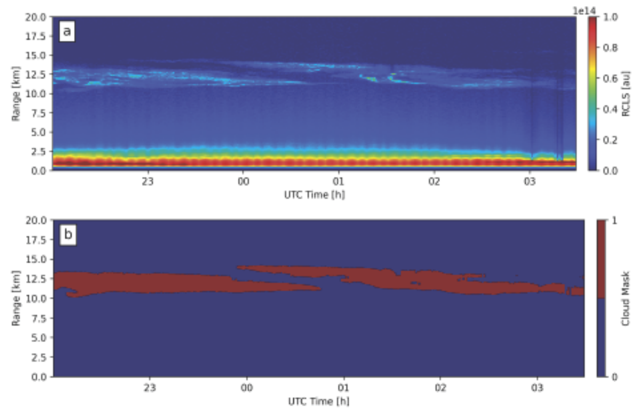


**Figure 7.** (a) RCLS (a) and layer-mask (b) from Manaus lidar for 21 September 2011.

### 3.2 Manaus Lidar - Brazil - 14 August 2011: Cirrus Clouds

The Manaus lidar site (Barbosa et. al., 2014b) is located inside the campus of Embrapa Amazônia (2.89$^{\circ}$ S, 59.97$^{\circ}$ W, 100 m altitude above sea level), approximately 30 km away from the city of Manaus-AM, Brazil. The instrument is a commercial lidar system model LR-102-U-400/HP by Raymetrics. The emitter is a Quantel CFR-400 Nd-YAG laser at 355 nm with 95 mJ per pulse and a 10 Hz repetition rate, with a beam expander producing a final divergence of 0.36 mrad. The reception optics consists of a Cassegrain telescope with a focal length of 4000 mm and a 400 mm primary mirror, resulting in an f/10 system. The field stop at the focal plane can be changed, allowing the telescope field of view to be adjusted between 0.25 and 3 mrad. The setup is bi-axial, with a 300 mm separation between the laser and the telescope axes. With a 7 mm field stop, the complete overlap of the system starts at 1.5 km. The detection box permits the acquisition of elastics and inelastic lidar signals in the UV using photon-current (ADC) and photon-counting (PC) methods. The lines acquired are 355 nm (ADC and PC), $N_2$'s Raman line at 387 (ADC and PC), and the water vapor line, 408 nm (PC). Data acquisition uses a Licel transient recorder model TR-20-160, with a raw resolution of 7.5 m.

For this lidar station, we focus here on measurements of cirrus clouds. Previous studies in the Amazon region showed that their frequency of occurrence is higher than in other tropical regions and that they can have an important radiative effect(Gouveia et al., 2017). Figure 7 shows a color map of the RCLS and the corresponding cloud layer mask from the level 1 data, where a thin cirrus between 10 and 15 km is clearly seen. The acquisition time of each lidar profile was 1 minute, and no time average was used for this layer mask retrieval, which clearly captured the whole cirrus cloud. For the retrieval of the optical properties, we used a lidar ratio of 23.3 sr for the cirrus clouds following the measurements reported by Gouveia et al. (2017). Figure 8 shows the particle extinction profile, corresponding to a cloud optical depth of 0.77. These results might be affected by multiple scattering, which is important for cirrus clouds and has not been accounted for. Nonetheless, values are reasonable and in the typical range of previous work (Gouveia et al., 2014), and show what can be obtained with the current version of LPP.

## Page: 9

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 13:03:40
here is much more detailed information compared to the Sao Paulo system - please homogenize.

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 13:02:48
which lines? please phrase correctly.

Number: 3 Author: Subject: [redacted] Date: 31.01.2023 13:47:13
the choice of the threshold and the overall methodology would be of particular interest. up to now, it is a "black box"

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022

Geoscientific Instrumentation Methods and Data Systems Discussions

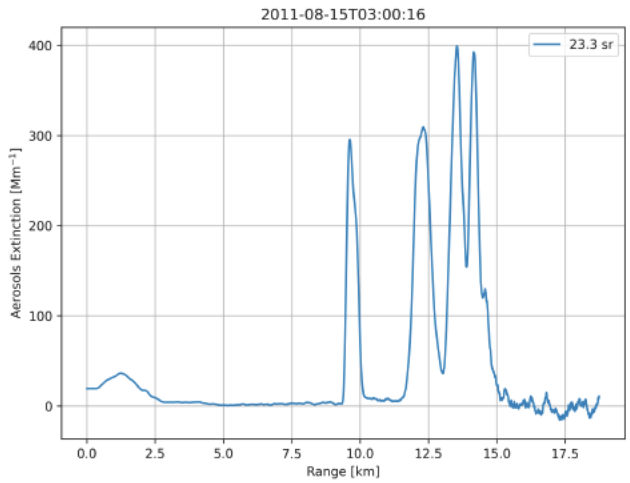


Figure 8. Extinction profile taken at 00:15 UTC for 21 September 2011.

The next release of LPP will introduce the use of a profile of LR instead of assuming a fixed value. This will allow processing cases, as depicted in this example, with cirrus clouds above biomass burning. Further studies and developments are necessary for implementing an automatic algorithm for removing the multiple scattering effects in semi-transparent clouds.

### 3.3 Future roadmap

With the first release of LPP and its use by the different groups in LALINET, we have identified the necessary improvements and built a roadmap to guide future development. The benefit of LPP's highly modular concept is the possibility of different groups modifying and testing different modules without interfering with the rest of the pipeline.

An initial consideration is that LPP processed data files need to be FAIR (Findable, Accessible, Interoperable, and Reusable) to be compatible with Open Science. In this sense, more information about the site, hardware, operation, files processed, and even the version of the LPP used needs to be added as metadata in the NetCDF structure in future releases. The second important point is that not all lidar stations in Latin America have a nearby radio-sounding site, and even when it is available, it is only twice daily. To facilitate the processing of level 1 and level 2 data, an automatic "thermodynamic profile downloader" will be needed. This should allow for downloading a co-located thermodynamic profile, either from a radio-sounding, a forecast model, or a reanalysis. Alternative methods of rescaling standard atmosphere profiles based on ground-based temperature and pressure measurements, which could also be retrieved automatically from meteorological databases, could be beneficial. Finally, a well-known problem with the inversion of elastic lidar data is the need to assume an a priori lidar ratio. The typical solution for day-time measurements is to choose a lidar ratio that brings the estimated AOD value closer to the reference value measured by AERONET. This analysis can be made a posteriori, but implementing an optimization routine would allow LPP to automatically identify the best LR for each profile, given that the user provides the reference AOD value. Alternatively, the reference values could be obtained by an "AOD data downloader" tool as part of the LPP framework.

## Page: 10

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 13:46:58

location missing. Why is here the height axis suddenly horizontally? Please homogenize in the manuscript.

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022




## 4 Conclusions

The development of LPP is a joint effort leveraging the expertise and manpower of different Latin-American groups. The goal is to have a set of open-source tools for each step of the typical lidar data analysis routine. We described the first working version that is now released for the scientific community on our GitHub repository for open access to everyone who wants to be involved or use the software. The two case studies for Sao Paulo and Amazon showed the capabilities of the current release but also highlighted the need for new features. From this exercise, we have built a roadmap to guide future development, which includes adding additional metadata, developing a thermodynamic profile and ERONET data downloader tools, and LR optimization methods. Future releases will bring these and other new features, accommodating the needs of our community.

*Code availability.* Version 1.0.0 reported here can be obtained from the GitHub repository https://github.com/juanpallotta/LPP under a MIT license that enables reuse, redistribution, and reproduction of all methods.

*Author contributions.* Conceptualisation of this on-going effort is shared by all authors. JP lead the software development, with contributions from FL and HB. FL and HB prepared the example datasets analyzed in this study. SC performed the first test of the code across different machines. JP and HB were responsible for visualization and writing the original draft. All authors reviewed and edited the published version of the paper. HB supervised the research.

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

*Acknowledgements.* We would like to thank the financial support from CITEDEF used to acquire the server-PC in which this project is developed. The authors also thank the financial support from Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) / (ARC: E-26/010/002464/2019 Ref.: 211.599/2019). We are grateful to Eng. Federico Verstraeten for sharing their knowledge in software development.

# Page: 11

Number: 1 Author: Subject: [redacted] Date: 31.01.2023 13:46:26
the differences to the SCC and other processing chains should be highlighted, e.g. more flexibility by using config files.

Number: 2 Author: Subject: [redacted] Date: 31.01.2023 13:10:13
also this is already used in some other processing chains

Number: 3 Author: Date: 31.01.2023 11:00:54
should be published with a DOI. Otherwise it might happen that the code is not longer available if, e.g., the user account is deactivated.

https://doi.org/10.5194/gi-2022-19
Preprint. Discussion started: 19 December 2022


Geoscientific Instrumentation Methods and Data Systems Discussions
Open Access EGU

## References

Ansmann, A. and Müller, D.: Lidar: Range-Resolved Optical Remote Sensing of the Atmosphere, chap. Lidar and Atmospheric Aerosol Particles, pp. 105–141, Springer New York, New York, NY, https://doi.org/10.1007/0-387-25101-4_4, 2005.

Antuña-Marrero, J. C., Landulfo, E., Estevan, R., Barja, B., Robock, A., Wolfram, E., Ristori, P., Clemesha, B., Zaratti, F., Forno, R., Armandillo, E., Álvaro E. Bastidas, Ángel M. de Frutos Baraja, Whiteman, D. N., Quel, E., Barbosa, H. M. J., Lopes, F., Montilla-Rosero, E., and Guerrero-Rascado, J. L.: LALINET: The First Latin American–Born Regional Atmospheric Observational Network, Bulletin of the American Meteorological Society, 98, 1255 – 1275, https://doi.org/10.1175/BAMS-D-15-00228.1, 2017.

Barbosa, H. M. J., Barja, B., Pauliquevis, T., Gouveia, D. A., Artaxo, P., Cirino, G. G., Santos, R. M. N., and Oliveira, A. B.: A permanent Raman lidar station in the Amazon: description, characterization, and first results, Atmospheric Measurement Techniques, 7, 1745–1762, https://doi.org/10.5194/amt-7-1745-2014, 2014.

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, Atmospheric Measurement Techniques, 8, 4891–4916, https://doi.org/10.5194/amt-8-4891-2015, 2015.

Dionisi, D., Barnaba, F., Diémoz, H., Di Liberto, L., and Gobbi, G. P.: A multiwavelength numerical model in support of quantitative retrievals of aerosol properties from automated lidar ceilometers and test applications for AOT and $\mathrm{PM}_{10}$ estimation, Atmospheric Measurement Techniques, 11, 6013–6042, https://doi.org/10.5194/amt-11-6013-2018, 2018.

Fernald, F. G.: Analysis of atmospheric lidar observations: some comments, Appl. Opt., 23, 652–653, https://doi.org/10.1364/AO.23.000652, 1984.

Freudenthaler, V., Linné, H., Chaikovski, A., Rabus, D., and Groß, S.: EARLINET lidar quality assurance tools, Atmospheric Measurement Techniques Discussions, 2018, 1–35, https://doi.org/10.5194/amt-2017-395, 2018.

Gouveia, D. A., Barbosa, H. M. J., and Barja, B.: Characterization of cirrus clouds in central Amazon (2.89ºS, 59.97ºW): Firsts results from observations in 2011, Special Section: VII Workshop on Lidar Measurement in Latin-America, 47, 109–114, https://doi.org/http://dx.doi.org/10.7149/OPA.47.2.109, 2014.

Gouveia, D. A., Barja, B., Barbosa, H. M. J., Seifert, P., Baars, H., Pauliquevis, T., and Artaxo, P.: Optical and geometrical properties of cirrus clouds in Amazonia derived from 1 year of ground-based lidar measurements, Atmospheric Chemistry and Physics, 17, 3619–3636, https://doi.org/10.5194/acp-17-3619-2017, 2017.

Guerrero-Rascado, J. L., Landulfo, E., Antuña, J. C., de Melo Jorge Barbosa, H., Barja, B., Álvaro Efrain Bastidas, Bedoya, A. E., da Costa, R. F., Estevan, R., Forno, R., Gouveia, D. A., Jiménez, C., Larroza, E. G., da Silva Lopes, F. J., Montilla-Rosero, E., de Arruda Moreira, G., Nakaema, W. M., Nisperuza, D., Alegria, D., Múnera, M., Otero, L., Papandrea, S., Pallota, J. V., Pawelko, E., Quel, E. J., Ristori, P., Rodrigues, P. F., Salvador, J., Sánchez, M. F., and Silva, A.: Latin American Lidar Network (LALINET) for aerosol research: Diagnosis on network instrumentation, Journal of Atmospheric and Solar-Terrestrial Physics, 138-139, 112–120, https://doi.org/https://doi.org/10.1016/j.jastp.2016.01.001, 2016.

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, vol. 1535 pp, [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/doi:10.1017/CBO9781107415324, 2013.

This page contains no comments

IPCC: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, vol. In Press, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, https://doi.org/10.1017/9781009157896, 2021.

Landulfo, E., Cacheffo, A., Yoshida, A. C., Gomes, A. A., da Silva Lopes, F. J., de Arruda Moreira, G., da Silva, J. J., Andrioli, V., Pimenta, A., Wang, C., Xu, J., Martins, M. P. P., Batista, P., de Melo Jorge Barbosa, H., Gouveia, D. A., González, B. B., Zamorano, F., Quel, E., Pereira, C., Wolfram, E., Casasola, F. I., Orte, F., Salvador, J. O., Pallotta, J. V., Otero, L. A., Prieto, M., Ristori, P. R., Brusca, S., Estupiñan, J. H. R., Barrera, E. S., Antuña-Marrero, J. C., Forno, R., Andrade, M., Hoelzemann, J. J., Guedes, A. G., Sousa, C. T., dos Santos Oliveira, D. C. F., de Souza Fernandes Duarte, E., da Silva, M. P. A., and da Silva Santos, R. S.: Lidar Observations in South America. Part II - Troposphere, in: Remote Sensing, edited by Hammond, A. and Keleher, P., chap. 2, IntechOpen, Rijeka, https://doi.org/10.5772/intechopen.95451, 2020.

Mather, J. M.: Atmospheric Radiation Measurement (ARM) User Facility 2020 Decadal Vision, Tech. Rep. DOE/SC-ARM-20-014, 1782812, https://doi.org/10.2172/1782812, 2021.

National Academies of Sciences, Engineering, and Medicine: Thriving on Our Changing Planet: A Decadal Strategy for Earth Observation from Space, National Academies Press, Washington, D.C., https://doi.org/10.17226/24938, pages: 24938, 2018.

Sugimoto, N. and Uno, I.: Observation of Asian dust and air-pollution aerosols using a network of ground-based lidars (ADNet): Realtime data processing for validation/assimilation of chemical transport models, IOP Conference Series: Earth and Environmental Science, 7, 012 003, https://doi.org/10.1088/1755-1307/7/1/012003, 2009.

Welton, E. J., Campbell, J. R., Spinhirne, J. D., and III, V. S. S.: Global monitoring of clouds and aerosols using a network of micropulse lidar systems, in: Lidar Remote Sensing for Industry and Environment Monitoring, edited by Singh, U. N., Asai, K., Ogawa, T., Singh, U. N., Itabe, T., and Sugimoto, N., vol. 4153, pp. 151 – 158, International Society for Optics and Photonics, SPIE, https://doi.org/10.1117/12.417040, 2001.

This page contains no comments