EARLINET Level 3
Algorithm Theoretical Basis Document

Sergio Ciamprone, Lucia Mona, Giuseppe D'Amico

Consiglio Nazionale delle Ricerche

CNR-IMAA

Version 1.0 - 01/03/2019

Summary

1- Introduction

This document reports the details on the methods applied for calculating centrally at data center level the ACTRIS/EARLINET Level 3 products. In particular, the methods related to the first release of such Level 3 products is reported. First of all, methods on how integrated properties are evaluated starting from the aerosol optical property single profiles (Level 2 products) is reported. Then it is detailed how the information (center of mass and h63) related to the distribution in the vertical dimension of the aerosol is calculated. Explanation about the gridding of the profiles in a standardized grid are then reported. Finally, the statistical methods applied for calculating the climatological values is reported. Particular attention has been paid to this aspect to take into account the not continuous dataset of ACTRIS/EARLINET and then the need of avoiding biased climatological values.

2- Calculation of integrated quantities

For each extinction/backscatter vertical profile, an integrated quantity is calculated: aerosol optical depth for extinction profiles, and aerosol integrated backscatter for backscatter profiles.

The aerosol optical depth (AOD) and the integrated backscatter (IB) are the integrals over the altitude of the aerosol extinction and backscatter profiles, respectively. These integrated properties directly provide an information about the quantity of aerosol present in the considered portion of the atmospheric column. However, aerosol optical depth and integrated backscatter also depend on the type of the particles, because the extinction and backscatter efficiencies depend on the size, shape and refractive index of the particles.

Let $\alpha(p)$ be extinction values depending on altitude p, and $\beta(p)$ the same for backscatter values. We have:

- **Optical depth:**

$$\int_{h_0}^{h_1} \alpha(p)dp$$

- **Integrated backscatter:**

$$\int_{h_0}^{h_1} \beta(p)dp$$

Before calculating integrated quantities, extinction, backscatter and altitude are submitted to the following quality controls in order to avoid anomalies and missing values (we indicate with l the station altitude asl and with $\varepsilon(p)$ the errors associated to extinction values):

1. $p > l$
2. $-10 < \alpha(p) < 10$
3. $\varepsilon(p) \geq 0$
4. $\varepsilon(p) < 0.5 \cdot |\alpha(p)|$

Identical conditions are applied to backscatter values.

Integrations are calculated with trapezoidal rule [1], which is a common technique in numerical analysis for approximating definite integrals. We explain it in more details. Let $\alpha_1, \dots, \alpha_n$ be extinction (same for backscatter) values, and $p_1, \dots, p_n$ be the corresponding altitudes at which extinction values are retrieved. Then:

$$\int_{h_0}^{h_1} \alpha(p)dp \approx \sum_{j=1}^{n} \frac{\alpha_{j-1} + \alpha_j}{2} \cdot \frac{p_j - p_{j-1}}{2}$$

where $\alpha_0 = \alpha_1$ and $p_0 = l$.

this means that we are assuming that below the first data provided in altitude by the stations the aerosol is well mixed and the corresponding optical property is constant with the altitude down to the ground. This is a typical hypothesis made in such kind of study and of course is more accurate for stations equipped with lidar with a low overlap range.

The values of the integral bounds h_0 and h_1 are:

1. $h_0 = l$ and $h_1 = p_n$, if we calculate the integrated quantity on the entire vertical profile.
2. $h_0 = l$ and $h_1 = p_i$, if we calculate the integrated quantity on the aerosol boundary layer (where p_i is the highest altitude among $p_1, \dots, p_n$ lower than aerosol boundary layer upper bound).

The aerosol boundary layer is not evaluated centrally but at the stations and provided if any into the Level 2 products. Aerosol boundary layer (previously called dust layer height, but renamed in 2019 for avoiding misunderstanding) is defined as the lowest layer that generally contains most of the aerosol except special elevated layers like Saharan dust etc (Matthias et al., 2004) .

As for what concerns the integrated backscatter values, these are retrieved from both extinction and backscatter type files. Backscatter profiles contained in e-files is not considered in the calculation only when b-file correspondingly to the same date exists.

If aerosol boundary layer is not reported in the specific file, algorithm tries to retrieve it from another file referred to the same date.

3- Center of Mass and H63

Center of mass is a value associated to every backscatter vertical profile. The center of mass of the aerosol content in the portion of atmospheric column is estimated as the backscatter weighted altitude in the considered altitude range (Mona et al., 2006). This quantity is an approximation of the center of mass of the aerosol layer, that exactly coincides with the true center of mass if both composition and size distribution of the particles are constant with the altitude. This estimate of the center of mass gives us information about the altitude where the most relevant part of the aerosol load is located.

It is calculated in the following way:

$$\frac{\int_{h_0}^{h_1} p \cdot \beta(p) dp}{\int_{h_0}^{h_1} \beta(p) dp}$$

Altitudes, backscatter values and error backscatter values are submitted to the same quality controls of the previous section. Integrations are calculated using trapezoidal rule. The values of integral bounds can vary as shown in the previous section.

Backscatter values are retrieved from both extinction and backscatter type files. An extinction type file is discarded only when a backscatter type file with its same date already exists.

The calculation of the center of mass for the total column and inside the aerosol boundary layer is calculated. If aerosol boundary layer is not reported in the specific file, algorithm tries to retrieve it from another file referred to the same date.

H63 is the altitude below which stays 63% of total aerosol optical depth (or aerosol integrated backscatter). More precisely, let $p_1, \dots, p_n$ be altitude values at which are retrieved $\alpha(p_i)$ extinction values. Let l be the station altitude, considering $\alpha(l) = \alpha(p_1)$. H63 is the lowest altitude p among $p_1, \dots, p_n$ such that:

$$\int_{l}^{p} a(p)dp > 0.63 \cdot \int_{l}^{p_n} a(p)dp$$

Analogously, it can be calculated for integrated backscatter. H63 depends on wavelength at which integrated quantities are retrieved.

Both for the center of mass and H63, it is assumed a constant profile of the optical properties below the lowest in range available observation in the original profile, assuming well mixed condition at that altitude range.

4 - Profile climatological products

Profile products give information about where and how much aerosol particles are placed in vertical profile. We consider all values (extinction or backscatter) retrieved from 0 up to 12000 meters, divided in 60 layers, each one 200 meters wide. Climatological profile products are reported in a fixed altitude range allowing direct comparisons between the different stations. The bounds of layers are at fixed altitudes: 0 – 200m, 200 – 400m, and so on. Intervals are intended in this way: [0,200), [200,400), …, meaning that a measurement retrieved at a bound altitude (200, 400, 600, …) is always putted in the upper layer. When we calculate profile statistical values about backscatter, we retrieve values from both extinction and backscatter type files. An extinction type file is discarded in the backscatter calculations only when a backscatter type file with its same date already exists.

4 - Statistical values

ACTRIS/EARLINET is providing aerosol observations in a no-continuous way: since 2000 to now it is performing measurements 3 times per week plus during CALIPSO overpasses (additionally also during special events which are disregarded in order to avoid biases in this analysis). Additionally, the presence of low clouds, fog and precipitations inhibits the lidar measurements furthermore limiting the measurement. Because of this reason, particular attention has to be paid for avoiding biased climatological values. For taking into account the not uniformity of the temporal coverage in the observations suitable statistical methods are applied (Atkinson , Kendall 1989; Lange, Ed Springer).

- **Annual averages**: mean, median and standard deviation calculations are weighted. This is due to the unbalancing of the number of values in the different months. In the next lines procedures are described in full details.
  Let *n* be the number of values which are going to be averaged, retrieved during the year *y* (at a fixed wavelength laser pulse). Let *m* be the number of months with at least one value. Of course, *m*

is between 0 and 12. Let $k_1, \ldots, k_m$ the number of values referred to the *m* months with at least a one value. The weight associated to the values x_{ij} (i^{th} value in j^{th} month) is $w_j = \frac{1}{m \cdot k_j}$

As a consequence, values retrieved in the same month have the same weight. Moreover, the sum of all weights is 1.

Here we reported formulas to calculate weighted mean, weighted median and weighted standard deviation. About weighted mean, we can also say that this procedure is equivalent to calculate (non-weighted) mean within months, and then calculate (non-weighted) mean over months.

i. Weighted mean: $\mu = \sum_{j=1}^{m} \sum_{i=1}^{k_j} w_j x_{ij}$
ii. Weighted median is the mean of all x_k values such that $\sum_{i=1}^{k-1} w_i \leq \frac{1}{2}$ and $\sum_{i=k+1}^{n} w_i \leq \frac{1}{2}$
iii. Weighted standard deviation: $\sqrt{\frac{1}{1-\sum_{j=1}^{m} \sum_{i=1}^{k_j} w_j^2} \cdot \sum_{j=1}^{m} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$

- **Seasonal averages:** mean, median and standard deviation are not weighted for season averages, because it has assumed that no significant differences are expected between data collected in the 3 months belonging to each season, so we do not need to fix the unbalancing of number of values in different sub periods.

Besides annual ad seasonal averages, the climatological normal mean are also calculated. Following the WMO definition these are averages computed for a uniform and relatively long period [WMO 2017]. They are of big interest firstly, because they form a benchmark or reference against which conditions (especially current or recent conditions) can be assessed, and secondly, because they are widely used (implicitly or explicitly) as an indicator of the conditions likely to be experienced in a given location.

- **Normal month averages:** mean, median and standard deviation calculations are weighted. This is due to the unbalancing of the number of values in the different years. In the next lines procedures are described in full details.

  Let *n* be the number of values which are going to be averaged, retrieved during every month *m* in the selected range of years, in this case 2000 – 2015. Let *y* be the number of years with at least one value retrieved. Let $k_1, \ldots, k_y$ the number of values per years (retrieved during the month *m*). The weight associated to the value x_{ij} (i^{th} value in j^{th} year) is $w_j = \frac{1}{y \cdot k_j}$

  As a consequence, values retrieved in the same year have the same weight. Moreover, the sum of all weights is 1.

  Here we reported formulas to calculate weighted mean, weighted median and weighted standard deviation. About weighted mean, we can also say that this procedure is equivalent to calculate (non-weighted) mean within years, and then calculate (non-weighted) mean over years.

iv. Weighted mean: $\mu = \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j x_{ij}$
v. Weighted median is the mean of all x_k values such that $\sum_{i=1}^{k-1} w_i \leq \frac{1}{2}$ and $\sum_{i=k+1}^{n} w_i \leq \frac{1}{2}$
vi. Weighted standard deviation: $\sqrt{\frac{1}{1-\sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j^2} \cdot \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$

- **Normal season averages:** mean, median and standard deviation calculations are weighted. This is due to the unbalancing of the number of values in the different years. In the next lines procedures are described in full details.
  Let *n* be the number of values which are going to be averaged, retrieved during every season *s* in the selected range of years, in this case 2000 – 2015. Let *y* be the number of years with at least one value retrieved. Let $k_1, \ldots, k_y$ the number of values per years (retrieved during the season *s*). The weight associated to the value x_{ij} (i^{th} value in j^{th} year) is $w_j = \frac{1}{y \cdot k_j}$
  As a consequence, values retrieved in the same year have the same weight. Moreover, the sum of all weights is 1.
  Here we reported formulas to calculate weighted mean, weighted median and weighted standard deviation. About weighted mean, we can also say that this procedure is equivalent to calculate (non-weighted) mean within months, and then calculate (non-weighted) mean over months.
  i. Weighted mean: $\mu = \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j x_{ij}$
  ii. Weighted median is the mean of all x_k values such that $\sum_{i=1}^{k-1} w_i \leq \frac{1}{2}$ and $\sum_{i=k+1}^{n} w_i \leq \frac{1}{2}$
  iii. Weighted standard deviation: $\sqrt{\frac{1}{1-\sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j^2} \cdot \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$

4 - References

Atkinson, Kendall E. (1989), An Introduction to Numerical Analysis (2nd ed.), New York: John Wiley & Sons.

Lange, Kenneth. Numerical analysis for Statisticians (second ed.). Springer.

Matthias, V., Balis, D., Bosenberg, J., Eixmann, R., Iarlori, M.,Komguem, L., Mattis, I., Papayannis, A., Pappalardo, G., Perrone, M.R. and Wang, X. (2004) Vertical aerosol distribution over Europe: statistical analysis of Raman lidar data from 10 European Aerosol Research Lidar Network (EARLINET) stations. Journal of Geophysical Research, vol. 109. D18201, doi:10.1029/2004JD004638

Mona, L., Amodeo, A., Pandolfi, M., Pappalardo, G. (2006) Saharan dust intrusions in the Mediterranean area: Three years of Raman lidar measurements. J. Geophys. Res., vol. 111, D16203, doi:10.1029/2005JD006569

WMO (2017) WMO Guidelines on the Calculation of Climate Normals. library.wmo.int