○

EARLINET Level3 Data Product Catalogue ALGORITHM THEORETICAL BASIS DOCUMENT

Sergio Ciamprone, Lucia Mona, Giuseppe D'Amico

Consiglio Nazionale delle Ricerche

Italy

Version 2.0 -23.11.2022

4. APPENDIX A – ALGORITHM THEORETICAL BASIS DOCUMENT

This appendix reports the details on the methods applied for calculating, centrally at the data center level, the ACTRIS/EARLINET Level 3 products. In particular, the methods related to the second release of such Level 3 products are reported. Statistical corrections have been applied in order to avoid biased climatological values since ACTRIS/EARLINET Level 2 dataset is not continuous.

a. Calculation of integrated quantities

For each extinction/backscatter vertical profile, an integrated quantity is calculated: aerosol optical depth (AOD) for extinction profiles, and aerosol integrated backscatter (IB) for backscatter profiles. AOD and IB are the integrals over the altitude of the aerosol extinction and backscatter profiles, respectively. These integrated products represent a proxy for the quantity of aerosol present in the considered portion of the atmospheric column. However, optical depth and integrated backscatter also depend on the type of particles because the extinction and the backscatter efficiencies depend on the size, shape, and refractive index of the particles.

Another integrated quantity is the center of mass, which is a value associated to every backscatter vertical profile. The center of mass of the aerosol content in a certain portion of the atmospheric column is estimated as the backscatter weighted average altitude in the considered altitude range [Mona et al., 2006]. This quantity approximates the center of mass of the aerosol layer, and 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 provides a proxy for the altitude where the most relevant part of the aerosol load is located.

H63 related to AOD is the altitude below which stays 63% of the aerosol optical depth calculated on the entire vertical profile. H63 related to IB can be defined in a similar way.

Lidar ratio, Angstrom exponent and particle depolarization cannot be strictly considered integrated quantities, since they are simply averaged in a given portion of the atmospheric column. Anyway, they are part of the integrated product files for sake of simplicity and completeness.

i. Aerosol optical depth and Angstrom coefficient

Let $\alpha(z)$ be the extinction values depending on the altitude z. The AOD is given by:

$$\int_{h_0}^{h_1} \alpha(z)dz$$

Integration is calculated with trapezoidal rule [Atkinson, 1989], which is a common technique in numerical analysis for approximating definite integrals. In more detail, let $\alpha_1, \dots, \alpha_n$ be the extinction values retrieved at the altitude $z_1, \dots, z_n$. Then:

$$\int_{h_0}^{h_1} \alpha(z)dz \approx \sum_{j=1}^{n} \frac{\alpha_{j-1} + \alpha_j}{2} \cdot \frac{z_j - z_{j-1}}{2}$$

where $\alpha_0 = \alpha_1$ and z_0 is the station altitude above sea level. 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.

Before calculating the integrated quantity, extinction values are submitted to the following quality controls in order to avoid anomalies (let ε_α be the error associated with the extinction):

a) $-0.01\ m^{-1} \leq \alpha(z) \leq 0.01 m^{-1}$
b) $\alpha(z) + \varepsilon_\alpha(z) \geq 0 m^{-1}$

Only profiles satisfying these quality controls are used for the climatological computations.

The values of the integral bounds h_0 and h_1 are:

a) $h_0 = z_0$ and $h_1 = z_n$, if we calculate the AOD on the entire vertical profile;
b) $h_0 = z_0$ and $h_1 = z_i$, if we calculate the AOD on the aerosol boundary layer (where z_i is the highest altitude among $z_1, \dots, z_n$ lower than the aerosol boundary layer upper bound)

The aerosol boundary layer 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].

Angstrom coefficient A is obtained by optical depth at 355 and 532 nm when available as

$$\left. \frac{AOD_{355}}{AOD_{532}} \middle/ \frac{532}{355} \right.$$

Where AOD at each wavelength is calculated as integrated quantity with methods, bounds and quality controls as reported above,

ii. Integrated backscatter

Let $\beta(z)$ be the backscatter values depending on the altitude z. The IB is given by:

$$\int_{h_0}^{h_1} \beta(z)dz$$

Everything said about the AOD can be identically repeated for IB, replacing the extinction with backscatter. Just quality controls are different:

a) $-0.0001m^{-1}sr^{-1} \leq \beta(z) \leq 0.0001m^{-1}sr^{-1}$
b) $\beta(z) + \varepsilon_\beta(z) \geq 0m^{-1}sr^{-1}$

where $\varepsilon_\beta(z)$ is the error associated to the backscatter value $\beta(z)$.

Only profiles satisfying these quality controls are used for the climatological computations.

iii. Center of mass

Let $\beta(z)$ be the backscatter values depending on the altitude z. The center of mass is given by:

$$\frac{\int_{h_0}^{h_1} z \cdot \beta(z)dz}{\int_{h_0}^{h_1} \beta(z)dz}$$

Backscatter 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 for AOD and IB.

iv. H63 of AOD and H63 of IB

Let $z_1, \dots, z_n$ be the altitudes at which the extinction values $\alpha_1, \dots, \alpha_n$ are retrieved. Let z_0 be the station altitude, and $\alpha_0 = \alpha_1$. H63 of AOD is the lowest altitude z among $z_1, \dots, z_n$ such that:

$$\int_{z_0}^{z} \alpha(z)dz > 0.63 \cdot \int_{z_0}^{z_n} \alpha(z)dz$$

Analogously, it can be calculated for integrated backscatter.

v. Lidar ratio and particle depolarization

Let $z_1, \dots, z_n$ be the altitudes at which the extinction values $\alpha_1, \dots, \alpha_n$ and the backscatter values $\beta_1, \dots,$ β_n are retrieved. Lidar ratio values are $s_i = \frac{\alpha_i}{\beta_i}$. Two average lidar ratio values are calculated. The first one is the mean of all the s_i values, on the entire vertical profile. The second one is the mean of the s_i values corresponding to the z_i lower than the altitude of the aerosol boundary layer upper bound. Linear particle depolarization average values are retrieved in the same way. Before calculating the mean, lidar ratio values and particle depolarization values are submitted to the following quality controls. Values that do not pass quality controls are not taken into account in the calculations:

Lidar ratio:

a) $s(z) \geq -100sr$
b) $s(z) \leq 200$ sr
c) $s(z) + \varepsilon_s(z) \geq 0$sr

Particle depolarization:

a) $p(z) + \varepsilon_p(z) \geq 0$
b) $p(z) - \varepsilon_p(z) \leq 1$

where ε_s and ε_p are the errors associated to lidar ratio and particle depolarization, respectively.

b. Calculation of layer products:

For each Level 2 file, geometrical information about layers is retrieved using an algorithm based on the one shown in [Siomos, 2018], applied to the backscatter vertical profile. These geometrical products (mainly, the altitude of the lower and upper bound of the layers) are then used to calculate the following optical products:

i. Extinction, backscatter, lidar ratio, particle depolarization

Let h_0 and h_1 be the altitude of the base and of the top of the layer, respectively. Let z_1, … , z_m the altitudes where extinction, backscatter, particle depolarization and lidar ratio are retrieved, such that $h_0 \leq z_i \leq h_1$ for all $i \in \{1, \dots, m\}$. The extinction value associated to the layer is the mean of the extinction values $\alpha_1, \dots, \alpha_m$ related to $z_1, \dots, z_m$. Backscatter, lidar ratio and particle depolarization are calculated in the same way. The values of level 2 optical products are submitted to the following quality controls, more restrictive than the ones seen in the previous section.

Extinction:

a) $\alpha(z) + \varepsilon_\alpha(z) \geq 0m^{-1}$
b) $\alpha(z) > -1 \cdot 10^{-5}m^{-1}$
c) $\alpha(z) < 1 \cdot 10^{-3}m^{-1}$

Backscatter:

a) $\beta(z) + \varepsilon_\beta(z) \geq 0 m^{-1} sr^{-1}$
b) $\beta(z) > -1 \cdot 10^{-8} m^{-1} sr^{-1}$
c) $\beta(z) < 1 \cdot 10^{-5} m^{-1} sr^{-1}$

Lidar ratio:

a) $s(z) + \varepsilon_s(z) \geq 0 sr$
b) $s(z) > -40 sr$
c) $s(z) < 140\ sr$

Particle depolarization:

a) $p(z) + \varepsilon_p(z) \geq 0$
b) $p(z) - \varepsilon_p(z) \leq 1$
c) $-1.1 < p(z) < 1.1$

Only profiles satisfying these quality controls are used for the climatological computations.

ii. Aerosol optical depth, integrated backscatter, center of mass

AOD, IB and center of mass are calculated as shown for the calculation of the integrated quantities. In this case, the integral bounds h_0 and h_1 are the altitude of the lower bound and of the upper bound of the layer, respectively. Extinction and backscatter values are submitted to the same quality controls of the previous subsection. Integrations are calculated using trapezoidal rule.

c. Calculation of profile values

Profile products give information about where and how much aerosol particles are placed in vertical profile. Extinction, backscatter, and volume linear depolarization ratio are the variables involved in this kind of products. All the values considered are retrieved from 100 up to 12100 meters. At this stage no calculations are performed. Data are just divided in 60 layers, each one 200 meters wide. Climatological profile products are reported in a fixed altitude range allowing direct comparisons between different stations. The bounds of layers are at fixed altitudes: 100 – 300 m, 300 – 500 m, and so on. Intervals are intended in this way: [100,300), [300,500), … , meaning that a measurement retrieved at a bound altitude is always putted in the upper layer.

d. Aggregated statistical values

Level 3 files report time-aggregated statistical values of all the variables shown in the previous section, retrieved from Level 2 files. Four different temporal aggregations are performed: seasonal, annual, normal monthly and normal seasonal. *Normal* means that the statistical products are computed for a uniform and relatively long period, following the WMO definition [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 and in a given time period.

ACTRIS/EARLINET is providing aerosol observations in a non-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 level 3 products). Moreover, the presence of low clouds, fog, and precipitations inhibits the lidar measurements furthermore limiting the measurement continuity. 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, 1989; Lange, 1999].

i. Integrated quantities and profile products

In the next lines there is a detailed description of the four temporal aggregations performed for the integrated and profile products :

- *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. 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 1 and 12 (if $m = 0$ then it means that no values are retrieved during the year y, so no computations are performed). Let $k_1, \dots, k_m$ the number of values referred to the different months. The weight associated to the value 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 the weights is 1. Here formulas to calculate weighted mean, weighted median and weighted standard deviation are reported. About weighted mean, it is easy to see that the following 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: first of all, all the x_{ij} are sorted from the lowest value to the highest one. Therefore, these values can be re-indexed using a single index, and indicated with x_q. The weight associated to x_q is indicated with w_q. The weighted median is the mean of all the x_q values such that $\sum_{i=1}^{q-1} w_i \leq \frac{1}{2}$ and $\sum_{i=q+1}^{n} w_i \leq \frac{1}{2}$.

  iii) Weighted standard deviation: $\sigma = \sqrt{\sum_{j=1}^{m} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$

- *Seasonal averages*: mean, median and standard deviation are not weighted for season averages, since no significant differences are expected among data collected in a certain season. So, there is no need to fix the unbalancing of number of values in different sub-periods
- *Normal month averages*: mean, median and standard deviation calculations are weighted. This is due to the unbalancing of the number of values retrieved in a given month through the different years. Let n be the number of values which are going to be averaged, retrieved during a given month m in the selected range of years (in this case 2000 – 2019). Let y be the number of years with at least one value retrieved during m. Of course, y is between 1 and 20 (if $y = 0$, then it means that no values are retrieved during the month m in the period 2000 – 2019, so no computations are performed). Let $k_1, \dots, k_y$ the number of values referred to the different years. 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[1]. Moreover, the sum of all the weights is 1. Here formulas to calculate weighted mean, weighted median and weighted standard deviation are reported. About weighted mean, it is easy to see that the following procedure is equivalent to calculate (non-weighted) mean within years, and then calculate (non-weighted) mean over years.

  i) Weighted mean: $\mu = \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j x_{ij}$
  ii) Weighted median: first of all, all the x_{ij} are sorted from the lowest value to the highest one. Therefore, these values can be re-indexed using a single index, and indicated with x_q. The weight associated to x_q is indicated with w_q. The weighted median is the mean of all the x_q values such that $\sum_{i=1}^{q-1} w_i \leq \frac{1}{2}$ and $\sum_{i=q+1}^{n} w_i \leq \frac{1}{2}$.
  iii) Weighted standard deviation: $\sigma = \sqrt{\sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$
- *Normal season averages*: mean, median and standard deviation calculations are weighted. This is due to the unbalancing of the number of values retrieved in a given season through the different years. Let n be the number of values which are going to be averaged, retrieved during a given season s in the selected range of years (in this case 2000 – 2019). Let y be the number of years with at least one value retrieved during s. Of course, y is between 1 and 20 (if $y = 0$, then it means that no values are retrieved during the season s in the period 2000 – 2019, so no computations are performed). Let $k_1, \dots, k_y$ the number of values referred to the different years. 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 the weights is 1. Here formulas to calculate weighted mean, weighted median and weighted standard deviation are reported. About weighted mean, it is easy to see that the following procedure is equivalent to calculate (non-weighted) mean within years, and then calculate (non-weighted) mean over years.
  i) Weighted mean: $\mu = \sum_{j=1}^{y} \sum_{i=1}^{k_j} w_j x_{ij}$

[1] Here an example is provided for clarification. If over a year data are available just for 3 months m=3. If for the 3 different months the following number of measurements are available: 4, 5,7, then the weight for each measurement collected in the 3 months are 1/12 (4 measurements) , 1/15 (5 measurements) and 1/21 (7 measurements), respectively. So that the sum of all weights is 1/3 +1/3+1/3 =1.

ii) Weighted median: first of all, all the x_{ij} are sorted from the lowest value to the highest one. Therefore, these values can be re-indexed using a single index, and indicated with x_q. The weight associated to x_q is indicated with w_q. The weighted median is the mean of all the x_q values such that $\sum_{i=1}^{q-1} w_i \leq \frac{1}{2}$ and $\sum_{i=q+1}^{n} w_i \leq \frac{1}{2}$.

iii) Weighted standard deviation: $\sigma = \sqrt{\sum_{j=1}^{m} \sum_{i=1}^{k_j} w_j \left(x_{ij} - \mu\right)^2}$

ii. Layer products

Four temporal aggregations are performed also for layer products: annual, seasonal, normal monthly, normal seasonal. In comparison with integrated and profile products, statistical tools used here are different. Data are collected generating a frequency histogram, establishing for each variable the histogram intervals. Histogram interval bounds are calculated in the following way:

- Base layer altitude, top layer altitude, center of mass (m): [0,1000); [1000,2000); [2000,3000); [3000,4000); [4000,5000); [5000,6000); [6000,7000); [7000,8000); [8000,9000); [9000,10000); [10000,11000); [11000,12000); [12000,13000); [13000,14000); [14000,15000); [15000,16000); [16000,17000); [17000,18000); [18000,19000); [19000,20000)
- All the other variables: let *v* be the set of all the retrieved values referred to the given variable. Let *m* be the minimum value of *v*, let *M* be the maximum value, *d1* the first decile and *d9* the ninth decile. Moreover: $p = \frac{d9-d1}{18}$. Then, the n-th interval (where *n*=2, … , 19) is given by [*d1*+(*n-2*)**p*,*d1*+(*n-1*)*p), while the first interval is [*m*,*d1*) and the last interval is [*d9*,*M*].

e. References

1. K. E. Atkinson: An introduction to Numerical Analysis (2nd ed.), New York: John Wiley & Sons (1989).
2. K. Lange: Numerical Analysis for Statisticians (2nd ed.). Springer (1999)
3. V. Matthias, D. Balis, J. Bosenberg, R. Eixmann, M. Iarlori, L. Komguem, I. Mattis, A. Papayannis, G. Pappalardo, M. R. Perrone, and X. Wang (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
4. L. Mona, A. Amodeo, M. Pandolfi, G. Pappalardo (2006): Saharan dust intrusions in the Mediterranean area: Three years of Raman lidar measurements. J. Geophys. Res., vol. 111, D16203, doi:10.1029/2005JD006569
5. N. Siomos, D. S. Balis, K. A. Voudouri, E. Giannakaki, M. Filioglou, V. Amiridis, A. Papayannis, and K. Fragkos.: Are EARLINET and AERONET climatologies consistent? The case of Thessaloniki, Greece, Atmos. Chem. Phys., 18, 11885–11903, https://doi.org/10.5194/acp-18-11885-2018, 2018
6. WMO (2017): WMO Guidelines on the Calculation of Climate Normals. library.wmo.int