EARLINET Level3 Data Product Catalogue

Sergio Ciamprone, Lucia Mona, Giuseppe D’Amico

Consiglio Nazionale delle Ricerche

Italy

Version 2.0 -23.11.2022

Contents

1. About this document

In this document the climatological products of aerosol optical properties derived from the EARLINET measurements are presented. From now on, these products will also be named as Level 3 products, obtained as aggregated products from the EARLINET fully quality-controlled (Level 2) aerosol optical products.

This document consists of a synthetic description of the Level 3 data and of two appendixes: Appendix A, the ATBD (Algorithm Theoretical Basis Document) of such datasets, and a further appendix B with the data format description.

Level 3 products are essential for climatological studies for the different kinds of potential users (related to air quality, meteorological application, climate study, modelling, and so on). In particular, these products provide high-quality information about the vertical distribution and the optical properties of the aerosol over the European continent on a long-term scale, and this could be precious also for the European climate and air quality policymakers. Additionally, these products could also be a term of reference for the stations performing measurements in order to investigate specific processes observed at those stations.

This is the second release of the data, following the first one implemented on March 2019 [EARLINET Level 3 Data Products, Mona et al., D2.13]. The main differences between the two releases are the following:

- A new version of quality controls for Level 2 data is used (namely quality controls version 3.0) [EARLINET Data Quality Check Procedure – v3.0; EARLINET Data Quality Check – Action Report];
- The reference period for the climatological products is enlarged (first release: 2000-2015, second release: 2000-2019);
- New variables are introduced (mainly, optical products related to aerosol layer data).

The Level 3 products are centrally and automatically produced by the CNR, guaranteeing harmonized procedures on the datasets. This is compliant with the general structure of the EARLINET database, allowing for versioning, traceability, and harmonization of the processes [Lund Myhre, D10.1].

2. Level 3 climatological products

The Level 3 standard products contain climatological datasets obtained as aggregated products from the fully quality controlled (QC) aerosol optical products (i.e. Level 2 products). For this second release, the dataset for 2000-2019 period is considered. In particular, this release considers only data fully compliant to the QC procedure v3.0 [EARLINET Data Quality Check]. In order to avoid biases due to measurements made on purpose specifically for capturing special events, it is considered only subset of data corresponding to regular schedule and measurements done for satellite validation purposes (i.e. *climatological* and *calipso* category files following the current EARLINET database organization [Pappalardo et al., 2014]).

Data are organized for stations in order to allow future updates of the climatological products even at station level. For each station, three types of data are released: profile values, integrated quantities, layer statistics. For each type of data, four different temporal aggregations are provided: seasonal, annual, normal seasonal, normal monthly. 33 stations have been considered for the Level 3 products:

Station Code	Name	Station Code	Name	Station Code	Name
aby	Aberystwyth	**hpb**	Hohenpeissenberg	**nap**	Naples
arr	Andoya	**ino**	Bucharest	**pot**	Potenza
atz	Athens	**ipr**	Ispra	**puy**	Clermont-Ferr
brc	Barcelona	**kuh**	Kuehlungsborn	**rme**	Roma
cbw	Cabauw	**lei**	Leipzig	**sal**	Lecce
cog	Belsk	**lkp**	Linkoping	**sir**	Palaiseau
dus	Dushanbe	**lle**	Lille	**sof**	Sofia
evo	Evora	**mas**	Minsk	**spl**	St. Petersburg
gar	Garmisch	**mdr**	Madrid	**the**	Thessaloniki
gra	Granada	**mel**	Melpitz	**ucc**	Cork
hbu	Hamburg	**muc**	Maisach	**waw**	Warsaw

Therefore, Level 3 dataset consists of 4158 files, 126 for each station: 42 contain profile values, 42 contain integrated quantities, 42 contain layer statistics. For each category, 20 files report annual statistics (one file per year, from 2000 to 2019); 20 files report seasonal statistics (one file per year, from 2000 to 2019); 1 file reports normal seasonal statistics; 1 file reports normal monthly statistics.

The specific data format (reported in Appendix A) has been constructed following the Level 2 data nomenclature as much as possible and trying to comply with international harmonization standards.

Traceability is also a key aspect for this release: the list of Level 2 data used for the calculation of aggregated products is reported, together with the PI information [Eaton, NetCDF Conventions].

Level 2 files can be divided in two different categories: type b files, which contain only backscatter profiles, and type e files, which contain both extinction and backscatter values. Level 3 products about backscatter are calculated using backscatter values from both type of Level 2 files. An extinction type file is discarded in the backscatter calculations only when a backscatter type file with the same date already exists.

3. References

1. EARLINET Data Quality Check – Action Report – available at www.earlinet.org
2. EARLINET Data Quality Check Procedure – v3.0 – available at www.earlinet.org
3. EARLINET Level 3 Data Products – First Release 2000-2015 – available at www.earlinet.org
4. B. Eaton, J. Gregory, B. Drach, K. Taylor, S. Hankin, J. Blower, J. Caron, R. Signell, P. Bentley, G. Rappa, H. Hock, A. Pamment, M. Juckes, M. Raspaud: NetCDF Climate and Forecast (CF) Metadata

Conventions – Version 1.7 - http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html

5. C. Lund Myhre, L. Mona, M. Fiebig, G. D'Amico, F. Amato, E. O'Connor, A. M. Fjaeraa, T. Hamburger, A. Hirsikko, P. Laj: Deliverable D10.1: Definition of the ACTRIS Data Management Plan – ACTRIS-2 Project
6. L. Mona, S. Ciamprone, G. D'Amico: Deliverable D2.13: Implementation of EARLINET Level 3 products – ACTRIS-2 Project
7. G. Pappalardo, A. Amodeo, A. Apituley, A. Cameron, V. Freudenthaler, H. Linné, A. Ansmann, J. Bosenberg, G. D'Amico, I. Mattis, L. Mona, U. Wandinger, V. Amiridis, L. Alados-Arboledas, D. Nicolae, and M. Wiegner: EARLINET: towards an advanced sustainable European aerosol lidar network, Atmos. Meas. Tech, 7, 2389-2409, https://doi.org/10.5194/amt-7-2389-2014, 2014.

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

$$\frac{AOD_{355}}{AOD_{532}} \Big/ \frac{532}{355}$$

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 $\sum_{i=1}^{q-1} w_i \le \frac{1}{2}$ and $\sum_{i=q+1}^{n} w_i \le \frac{1}{2}$.

  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

5. APPENDIX B – DATA PRODUCT CATALOGUE

This appendix describes the content of Level 3 data files, including data format and organization. Following the standards adopted within EARLINET DC and in agreement with the whole ACTRIS database, Level 3 data are included in NetCDF files. Variables nomenclature and file format are aligned with those of pre-processed and processed ACTRIS aerosol remote sensing data.

The Level 3 data are centrally obtained by the ACTRIS aerosol remote sensing Data Center node of the CNR in Potenza. This allows the harmonization and reproducibility of the products.

a. File naming convention

File names for the Level 3 aerosol data follow this convention:

ACTRIS_AerRemSen_sss_Lev03_mmmmmm_ppp_ttt_vxx_qcyyy.nc

Where:

- *sss* is the station ID code;
- *mmmmmm* is the temporal aggregation (Annual=annual averages, Season=season averages, NorMon=normal month averages, NorSea=normal season averages);
- *pppp* is the period of the calculated averages. For seasonal and annual averages pppp is a year (i.e. pppp=2000), for normal monthly and normal seasonal averages pppp=0019;
- *ttt* is the product type (Int=integrated values, Pro=profile data, Lay=layer statistics);
- *vxx* version of the data product. Since this is the second release, it is v02;
- *qcyyy* is the version of the qc used for the evaluation of Level 3 data products (in this case, qc030).

b. Definition and meaning of the parameters

In this section, all the parameters reported in the Level 3 files are described. Some variables can be found only in a specific type of product, some others are present in all the Level 3 files.

i. Dimensions

In the next lines, the list of all dimensions in level 3 files is reported, together with the type of file where they are contained:

- *wavelength* (Int,Lay,Pro) provides the wavelength at which products are retrieved. It assumes one among the following three values: 355 nm, 532 nm, 1064 nm;
- *nv* (Int,Lay,Pro) assumes two integer values (1 and 2), and it is used following the cf convention whenever a bound is needed for some evaluation;
- *time* (Int,Lay,Pro) provides the reference period where aggregations are performed. More precisely it reports the middle time of the reference period (i.e., 23:59:59 of June 30th for annual files) It is expressed in seconds since 1970/1/1 00:00, and it assumes one value for annual files, four values for seasonal and normal seasonal files, and twelve values for normal monthly files;
- *n_char* (Int,Lay,Pro) allows the recording of the names of source files from which the Level 3 products are evaluated. It is an integer number which represents the number of characters used for the list of the source files;
- *stats* (Int,Pro) provides the type of statistics reported in the file (mean, statistical error mean (i.e. the mean of the statistical error of Level 2 data used in the aggregation for obtaining Level3 data),, median, standard deviation, number of values aggregated);
- *breaks* (Lay) provides the number of break points of the histogram related to a certain variable.

ii. Variables

In next sections, variables are reported, separately for each kind of files. More information about how these variables are computed can be found in the Appendix A. For each variable, the associated dimensions are listed inside the square brackets:

Integrated optical products:

In this section, variables related to the integrated optical products are reported:

- *integral_bounds* [nv] provides the portion of the vertical profiles where the variables are integrated: the entire profile or the aerosol boundary layer;
- *aerosol_optical_depth* [time,nv,wavelength,stats] is the integral over the altitude of the aerosol extinction profile;
- *integrated_backscatter* [time,nv,wavelength,stats] is the integral over the altitude of the aerosol backscatter profile;
- *lidar_ratio* [time,nv,wavelength,stats] is the ratio of the extinction coefficient to the backscatter coefficient;
- *aerosol_boundary_layer* [time,stats] is the altitude of the top of the lowest layer that typically contains the most of the aerosol, except special elevated layers like Saharan dust etc.;
- *h63_of_aerosol_optical_depth* [time,wavelength,stats] is the altitude below which 63% of the total integrated extinction profile is confined;
- *h63_of_integrated_backscsatter* [time,wavelength,stats] is the altitude below which 63% of the total integrated backscatter profile is confined;

- *center_of_mass* [time,nv,wavelength,stats] is the altitude of the punctual particle equivalent of a given object for application of Newton's laws of motion. It is estimated as the altitude weighted by the backscatter coefficient;
- *particle_depolarization* [time,nv,wavelength,stats] is the ratio of the perpendicular polarization component to the parallel component of the aerosol scattering;
- *angstrom_coefficient* [time,nv,stats] is a parameter that describes how the optical thickness of an aerosol depends on the wavelength;
- *time_bounds* [time,nv] provides the bound of the temporal interval where variables are aggregated;
- *latitude* [ ] gives the latitude coordinate of the station;
- *longitude* [ ] gives the longitude coordinate of the station;
- *source* [n_char] provides the list of the Level 2 files from which the current Level 3 file originates;
- *station_altitude* [ ] provides the altitude of the station (in meters, above sea level).

Layer products:

In this section, variables contained in the layer files are reported:

- *altitude_intervals* [breaks] provides the break point values of the frequency histograms related to the following variables: center of mass, base layer altitude, top layer altitude;
- *lidar_ratio_intervals* [breaks] provides the break point values of the lidar ratio frequency histogram;
- *extinction_intervals* [breaks] provides the break point values of the extinction frequency histogram;
- *backscatter_intervals* [breaks] provides the break point values of the backscatter frequency histogram;
- *particle_depolarization_intervals* [breaks] provides the break point values of the particle depolarization frequency histogram;
- *aerosol_optical_depth_intervals* [breaks] provides the break point values of the AOD frequency histogram;
- *integrated_backscatter_intervals* [breaks] provides the break point values of the integrated backscatter frequency histogram;
- *base_layer_altitude_frequency* [time,breaks] provides the frequencies of the base layer altitude values referred to the intervals given by the altitude_intervals variable;
- *top_layer_altitude_frequency* [time,breaks] provides the frequencies of the top layer altitude values referred to the intervals given by the altitude_intervals variable;
- *center_of_mass_altitude_frequency* [time,wavelength,breaks] provides the frequencies of the center of mass altitude values referred to the intervals given by the altitude_intervals variable;
- *lidar_ratio_frequency* [time,wavelength,breaks] provides the frequencies of the lidar ratio values referred to the intervals given by the lidar_ratio_intervals variable;
- *extinction_frequency* [time,breaks] provides the frequencies of the extinction mean values referred to the intervals given by the extinction_intervals variable;

- *backscatter_frequency* [time,breaks] provides the frequencies of the backscatter mean values referred to the intervals given by the backscatter_intervals variable;
- *particle_depolarization_frequency* [time,breaks] provides the frequencies of the particle depolarization values referred to the intervals given by the particle_depolarization_intervals variable;
- *aerosol_optical_depth_frequency* [time,breaks] provides the frequencies of the AOD values referred to the intervals given by the aerosol_optical_depth_intervals variable;
- *integrated_backscatter_frequency* [time,breaks] provides the frequencies of the integrated backscatter values referred to the intervals given by the integrated_backscatter_intervals variable;
- *time_bounds* [time,nv] provides the bound of the temporal interval where variables are aggregated;
- *latitude* [ ] gives the latitude coordinate of the station;
- *longitude* [ ] gives the longitude coordinate of the station;
- *source* [n_char] provides the list of the Level 2 files from which the current Level 3 file originates;
- *station_altitude* [ ] provides the altitude of the station (in meters, above sea level).

Profile products:

In this section, variables contained in the profile files are reported:

- *extinction* [altitude,time,wavelength,stats] provides the statistics about aerosol extinction values retrieved at a specific wavelength, within specific temporal and vertical ranges;
- *backscatter* [altitude,time,wavelength,stats] provides the statistics about aerosol backscatter values retrieved at a specific wavelength, within specific temporal and vertical ranges;
- *volume_depolarization* [altitude,time,wavelength,stats] provides the statistics about the volume depolarization ratio retrieved at a specific wavelength, within specific temporal and vertical ranges;
- *time_bounds* [time,nv] provides the bound of the temporal interval where variables are aggregated;
- *latitude* [ ] gives the latitude coordinate of the station;
- *longitude* [ ] gives the longitude coordinate of the station;
- *source* [n_char] provides the list of the Level 2 files from which the current Level 3 file originates;
- *station_altitude* [ ] provides the altitude of the station (in meters, above sea level).

iii. Attribute description

In this section, attributes are briefly described. The attributes of the variables are structured in such a way as to provide all needed information about units, corresponding names (if existing) in CF convention, FillValue, and in some cases a short description of the variable.

Global attributes are mainly information related to traceability of the data where is reported information about: software used to generate Level 3 products, the station to which Level 3 data refer, the data originator and data provider, and info about potential modification and versions of the Level 3 data. In particular, in agreement with the GEOMS definition (https://evdc.esa.int/documents/1/geoms-1.0.pdf), the Data Originator (DO) is the person that generated and quality controlled the data. Where no single DO exist, the DO_NAME and DO_AFFILIATION will hold the name of the entity responsible for the instrument, while the DO_ADDRESS and DO_EMAIL will contain the appropriate contact information. The DO may or may not be the same person as the PI.

c. File Examples

In the following sections, the detailed structure of the Level 3 annual files will be shown, as an example. The other Level 3 files which originate from a different temporal aggregation have a very similar structure.

i. Level 3 average integrated values file - NetCDF structure:

Dimensions:

- nv
- time
  - i) units: seconds since 1970-01-01T00:00Z
  - ii) long_name: Time
  - iii) calendar: gregorian
  - iv) axis: T
  - v) standard_name: time
  - vi) bounds: time_bounds
- wavelength
  - i) units: nm
  - ii) long_name: Wavelength of the transmitted laser pulse
- stats
  - i) long_name: statistics
  - ii) flag_value: 0,1,2,3,4
  - iii) flag_meaning: 0:mean, 1:statistical error mean, 2:median, 3:standard deviation, 4:number of values aggregated
- n_char

Variables:

- integral_bounds [nv] (type: bytes)

i) long_name: integral bounds of integrated values
ii) flag_value: 0,1
iii) flag_meaning: 0:total, 1:aerosol boundary layer

- aerosol_optical_depth [time,nv,wavelength,stats] (type: double)
  i) units: 1
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: aerosol optical depth
  iv) standard_name: atmosphere_optical_thickness_due_to_ambient_aerosol_particles
- integrated_backscatter [time,nv,wavelength,stats] (type: double)
  i) units: 1/sr
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: integrated_backscatter
- lidar_ratio [time,nv,wavelength,stats] (type: double)
  i) units: sr
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: aerosol extinction-to-backscatter ratio
- aerosol_boundary_layer [time,stats] (type: double)
  i) units: m
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: altitude of the upper bound of the aerosol planet boundary layer
- h63_of_aerosol_optical_depth [time,wavelength,stats] (type: double)
  i) units: m
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: altitude below which stays the 63% of the total aerosol optical depth
- h63_of_integrated_backscatter [time,wavelength,stats] (type: double)
  i) units: m
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: altitude below which stays the 63% of the total aerosol optical depth
- center_of_mass [time,nv,wavelength,stats] (type: double)
  i) units: m
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: center of mass
- particle_depolarization [time,nv,wavelength,stats] (type: double)
  i) units: 1
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: aerosol linear particle depolarization ratio
- time_bounds [time,nv] (type: double)
  i) units: seconds since 1970-01-01T00:00Z
- latitude [ ] (type: float)
  i) units: degrees_north
  ii) long_name: latitude of the station
  iii) standard_name: latitude
- longitude [ ] (type: float)
  i) units: degrees_east
  ii) long_name: longitude of the station

    iii) standard_name: longitude
- source [n_char] (type: char)
    i) long_name: source files
    ii) description: List of level 2 files from which are retrieved values averaged in this file
- station_altitude [ ] (type: float)
    i) units: m
    ii) long_name: station altitude above sea level

Global attributes:

- processor_name: EAR_clim_v1.exe
- processor_version:
- processor_institution: CNR – IMAA
- system:
- location:
- institution:
- PI:
- PI_affiliation:
- PI_affiliation_acronym:
- PI_address:
- PI_phone:
- PI_email:
- data_originator: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_originator_affiliation_acronym: CNR – IMAA
- data_originator_address:
- data_originator_phone:
- data_originator_email: earlinetdb@actris.imaa.cnr.it
- data_provider: ACTRIS ARES
- data_provider_affiliation: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_provider_affiliation_acronym: CNR – IMAA
- data_provider_address:
- data_provider_phone:
- data_provider_email: earlinetdb@actris.imaa.cnr.it
- conventions: C.F. – 1.8
- references: link doc earlinet.org
- station_ID:
- __file_format_version:
- history: YYYY-MM-DD hh:mm:ss Generated by free software R, using package ncdf4
- title: Annual average integrated measurements – year YYYY

ii. Level 3 average layer values file – NetCDF structure:

Dimensions:

- nv
- time
  - i) units: seconds since 1970-01-01T00:00Z
  - ii) long_name: Time
  - iii) calendar: gregorian
  - iv) axis: T
  - v) standard_name: time
  - vi) bounds: time_bounds
- breaks
  - i) long_names: Histogram breaks
- wavelength
  - i) units: nm
  - ii) long_name: Wavelength of the transmitted laser pulse
- n_char

Variables:

- base_layer_altitude_frequency [time,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the base layer altitude values
  - iii) histogram_intervals: altitude_intervals
- top_layer_altitude_frequency [time,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the top layer altitude values
  - iii) histogram_intervals: altitude_intervals
- center_of_mass_altitude_frequency [time,wavelength,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the center of mass altitude values
  - iii) histogram_intervals: altitude_intervals
- lidar_ratio_frequency [time,wavelength,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the lidar ratio values
  - iii) histogram_intervals: lidar_ratio_intervals
- extinction_frequency [time,wavelength,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the extinction values
  - iii) histogram_intervals: extinction_intervals
- particle_depolarization_frequency [time,wavelength,breaks] (type: double)
  - i) _FillValue: 9.96920996838687e+36
  - ii) long_name: Frequency distribution of the linear particle depolarization ratio
  - iii) histogram_intervals: particle_depolarization_intervals

- aerosol_optical_depth_frequency [time,wavelength,breaks] (type: double)
  i) _FillValue: 9.96920996838687e+36
  ii) long_name: Frequency distribution of the aerosol optical depth values
  iii) histogram_intervals: aerosol_optical_depth_intervals
- integrated_backscatter_frequency [time,wavelength,breaks] (type: double)
  i) _FillValue: 9.96920996838687e+36
  ii) long_name: Frequency distribution of the integrated backscatter values
  iii) histogram_intervals: integrated_backscatter_intervals
- backscatter_frequency [time,wavelength,breaks] (type: double)
  i) _FillValue: 9.96920996838687e+36
  ii) long_name: Frequency distribution of the backscatter values
  iii) histogram_intervals: backscatter_intervals
- altitude_intervals [breaks] (type: double)
  i) units: m
  ii) description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- lidar_ratio_intervals [breaks] (type: double)
  i) units: sr
  ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- extinction_intervals [breaks] (type: double)
  i) units: 1/km
  ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- particle_depolarization_intervals [breaks] (type: double)
  i) units: 1
  ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- aerosol_optical_depth_intervals [breaks] (type: double)
  i) units: 1
  ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- integrated_backscatter_intervals [breaks] (type: double)
  i) units: $(10^3 sr)^{-1}$

ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.

- backscatter_intervals [breaks] (type: double)
  - i) units: 1/Mm*sr
  - ii) description: description: Histogram interval bounds are reported. The n-th value represents the lower bound of the n-th interval, while the higher bound is the (n+1)-th value, since intervals are adiacent. The last interval (n=20) has no higher bound, since it is right-open.
- time_bounds [time,nv] (type: double)
  - i) units: seconds since 1970-01-01T00:00Z
- latitude [ ] (type: float)
  - i) units: degrees_north
  - ii) long_name: latitude of the station
  - iii) standard_name: latitude
- longitude [ ] (type: float)
  - i) units: degrees_east
  - ii) long_name: longitude of the station
  - iii) standard_name: longitude
- source [n_char] (type: char)
  - i) long_name: source files
  - ii) description: List of Level 2 files from which are retrieved values averaged in this file
- station_altitude [ ] (type: float)
  - i) units: m
  - ii) long_name: station altitude above sea level

Global attributes:

- processor_name: EAR_clim_v1.exe
- processor_version:
- processor_institution: CNR – IMAA
- system:
- location:
- institution:
- PI:
- PI_affiliation:
- PI_affiliation_acronym:
- PI_address:
- PI_phone:
- PI_email:
- data_originator: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_originator_affiliation_acronym: CNR – IMAA

- data_originator_address:
- data_originator_phone:
- data_originator_email: earlinetdb@actris.imaa.cnr.it
- data_provider: ACTRIS ARES
- data_provider_affiliation: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_provider_affiliation_acronym: CNR – IMAA
- data_provider_address:
- data_provider_phone:
- data_provider_email: earlinetdb@actris.imaa.cnr.it
- conventions: C.F. – 1.8
- references: link doc earlinet.org
- station_ID:
- __file_format_version:
- history: YYYY-MM-DD hh:mm:ss Generated by free software R, using package ncdf4
- title: Annual distribution of layer optical values – year YYYY

iii. Level 3 average profile values file – NetCDF structure:

Dimensions:

- altitude
  i) units: m
  ii) long_name: Altitude
  iii) axis: Z
  iv) positive: up
  v) standard_name: altitude
- time
  i) units: seconds since 1970-01-01T00:00Z
  ii) long_name: Time
  iii) calendar: gregorian
  iv) axis: T
  v) standard_name: time
  vi) bounds: time_bounds
- nv
- n_char
- wavelength
  i) units: nm
  ii) long_name: Wavelength of the transmitted laser pulse
- stats
  i) long_name: statistics

ii) flag_value: 0,1,2,3,4
iii) flag_meaning: 0:mean, 1:statistical error mean, 2:median, 3:standard deviation, 4:number of profiles aggregated

Variables:

- extinction [altitude,time,wavelength,stats] (type: double)
  i) units: 1/m
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: aerosol particle extinction coefficient
  iv) standard_name: volume_extinction_coefficient_in_air_due_to_ambient_aerosol_particles
- backscatter [altitude,time,wavelength,stats] (type: double)
  i) units: 1/m*sr
  ii) _FillValue: 9.96920996838687e+36
  iii) long_name: aerosol particle backscatter coefficient
- volume_depolarization [altitude,time,wavelength,stats] (type: double)
  i) _FillValue: 9.96920996838687e+36
  ii) long_name: aerosol volume depolarization coefficient
- time_bounds [nv,time] (type: double)
  i) units: seconds since 1970-01-01T00:00Z
- source [n_char] (type: char)
  i) long_name: source files
  ii) description: List of Level 2 files from which are retrieved values averaged in this file
- latitude [ ] (type: float)
  i) units: degrees_north
  ii) long_name: latitude of the station
  iii) standard_name: latitude
- longitude [ ] (type: float)
  i) units: degrees_east
  ii) long_name: longitude of the station
  iii) standard_name: longitude
- station_altitude [ ] (type: float)
  i) units: m
  ii) long_name: station altitude above sea level

Global attributes:

- processor_name: EAR_clim_v1.exe
- processor_version:
- processor_institution: CNR – IMAA
- system:
- location:
- institution:

- PI:
- PI_affiliation:
- PI_affiliation_acronym:
- PI_address:
- PI_phone:
- PI_email:
- data_originator: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_originator_affiliation_acronym: CNR – IMAA
- data_originator_address:
- data_originator_phone:
- data_originator_email: earlinetdb@actris.imaa.cnr.it
- data_provider: ACTRIS ARES
- data_provider_affiliation: Consiglio Nazionale delle Ricerche – Istituto di Metodologie per l'Analisi Ambientale
- data_provider_affiliation_acronym: CNR – IMAA
- data_provider_address:
- data_provider_phone:
- data_provider_email: earlinetdb@actris.imaa.cnr.it
- conventions: C.F. – 1.8
- references: link doc earlinet.org
- station_ID:
- __file_format_version:
- history: YYYY-MM-DD hh:mm:ss Generated by free software R, using package ncdf4
- title: Annual average profile measurements – year YYYY

d. References

1. B. Eaton, J. Gregory, B. Drach, K. Taylor, S. Hankin, J. Blower, J. Caron, R. Signell, P. Bentley, G. Rappa, H. Hock, A. Pamment, M. Juckes, M. Raspaud: NetCDF Climate and Forecast (CF) Metadata Conventions – Version 1.7 - http://cfconventions.org/Data/cf-conventions/cf-conventions-1.7/cf-conventions.html
2. V. Matthias, et. al. (2004): The vertical distribution over Europe: Statistical analysis of Raman lidar data from 10 EARLINET stations, J. Geophys. Res., 109, D18201, doi:10.1029/2004JD004638
3. L. Mona, A. Amodeo, M. Pandolfi, and G. Pappalardo (2006), Saharan dust intrusions in the Mediterranean area: Three years of Raman lidar measurements, J. Geophys. Res., 111, D16203, doi:10.1029/2005JD006569
4. D. M. Winker, J. L. Tackett, B. J. Getzewich, Z. Liu, M. A. Vaughan, and R. R. Rogers: The global 3-D distribution of tropospheric aerosols as characterized by CALIOP, Atmos. Chem. Phys., 13, 3345-3361, https://doi.org/10.5194/acp-13-3345-2013, 2013.