

EARLINET Data Quality Check Extensive Procedure – v1.0

Summary

- 1- Introduction..... 2
- 2- Technical Quality Check v1.0 3
- 3- Physical Quality Check v1.0 5

1- Introduction

Data quality check procedures for EARLINET/ACTRIS database improved over the time. This document reports the description of the extensive QC procedure applied on 21 February 2017.

On the base of the previous experiences on the screening of EARLINET data done file by file by eye, it was established to set up an automatic procedure for avoiding what were recognized as the most common problems in the EARLINET database.

Two main categories of checks are identified: technical checks and scientific content checks. **Technical checks** are procedures for the control of the file content respect to the file structure as defined in the EARLINET database (<https://www.earlinet.org/index.php?id=125>). The **scientific content** QCs are instead related to the content of the EARLINET files in terms of validity of the EARLINET measured parameters.

Details of the specific QC are reported in the following sections.

2- Technical Quality Check v1.0

TQC_00

This procedure checks that each file contains the mandatory product:

QC_00a	applies on b files	==>BackscatterIS NOT NaN or NULL (has at least 1 valid value)
QC_00b	applies on e files	==>ExtinctionIS NOT NaN or NULL (has at least 1 valid value)

TQC_01

This procedure checks that the name of the file is correct respect to expected one (correct ID, Datetime and wavelength).

QC_01a check on [STATIONID [2 chars]]

QC_01b check on [DATETIME [yy][MM][DD][HH][mm]]

QC_01c check if Filename [DATETIME] is consistent with the measurement [starttime] in the database
In particular this checks if the StartTime and the name of the file are compliant.StartTime is considered compliant if it is equal to the time reported in the Filename within 2 minutes.

TQC_02

This procedure performs the following checks :

- 1 - check if file length is 17 or 18 characters
- 2 - check the 1-st and 2-nd file characters are [char]
- 3 - check the 13-th file characters is [.]
- 4 - check the 14-th file characters is [e/b]
- 5 - check the chars from 2 to 11 are decimals [datetime]
- 6 - check the chars from 14 to 16/17 are decimals [wavelength]

TQC_03

Here the procedure checks if file extension is consistent (within 1nm) with the Emission Wavelength in the internal data table of the database, needed for searching procedure on the EARLINET database interface.

TQC_04

This QC procedure checks for Undefined Variables: it is expected that defined variables should contain at least one valid value. Further checks on MixingLayerHeight and DustLayerHeight are performed respect to their definition reported in the EARLINET file content description (<https://www.earlinet.org/index.php?id=125>)

QC_04a All array variables cannot be ALL undefined or negatives

QC_04b If it is a b-file must contain the variables Backscatter, ErrorBackscatter

QC_04c If it is an e-file must contain the variables Extinction, ErrorExtinction

QC_04d If the MixingLayerHeight is present must also be present the DustLayerHeight

QC_04e $\text{MixingLayerHeight} \leq \text{DustLayerHeight}$

QC_04f $\text{Altitude of the station} < \text{MixingLayerHeight}, \text{DustLayerHeight}$

This QC procedure applies also to the time series.

TQC_05

This procedure checks that the coordinates reported in the file are compliant with EARLINET station coordinates.

It checks if (latitude, longitude, altitude) of location are correct within the errors:

- (latitude, longitude) ± 0.05 degrees \Rightarrow 5 km
- altitude ± 20 m

TQC_06

The EARLINET climatological schedule establishes as routine measurements the ones performed on Monday around noon and around sunset and on Thursday around sunset. The schedule was established for providing a climatological dataset with a fixed time during the week and simultaneous measurements around the network. Only measurements performed with this schedule can be considered as climatologic measurements.

A check with calendar is implemented: it checks if Climatology Measurements had been made on Monday and Thursday

TQC_07

A check on cirrus presence is done. It checks that files associated to *cirrus category* reports some comments about the cirrus/cloud.

- Check the presence of the keywords : cirrus, cloud in the cirrus category files.

TQC_08

A check on cirrus presence is done. If cirrus/clouds keywords are reported in the comments, files should be in cirrus category.

TQC_09

If a file is classified as cirrus, all the simultaneously co-located files must be classified as cirrus. A procedure for this check has been set up.

TQC_10

The CALIPSO EARLINET correlative measurements are considered overpasses if the relative distance is maximum 100 km (what within the network was called Case 1 or Case A). All the other CALIPSO related measurements are performed for variability studies and could accomplish the CALIPSO related investigation, but cannot be considered as “collocated” measurements. The CALIPSO category contains the measurements related to the centralized alerted CALIPSO overpasses (Case1-2-3 or Case A-B-C). For the alerted stations, the measurement is considered compliant with the distributed overpass table if the temporal distance between the overpass and the analysed data is less than 12 hours (maximum temporal distance considered for variability study in the ESA-CALIPSO project and Pappalardo et al., JGR 2010).

A check with distributed overpass tables is implemented for this.

3- Physical Quality Check v1.0

PQC-00

This procedure scans each profile and check that the error on the optical properties are positive for all defined value of the corresponding optical property.

A file does not pass the PQC00

* if at least for 1 error point is negative, zero or not defined, for defined optical property value

PQC_01

This procedure scans each profile and check that the aerosol optical properties are positive within 3sigma, that there are not negative peaks and that very extreme values are present only in cirrus cloud cases.

Check on backscatter :

* backscatter plus a Threshold value ≥ 0

* OR $\text{abs}(\text{backscatter}) < 3 * \text{errBackscatter}$

* backscatter < Back_peak(Wavelength) if it does not belong to cirrus category

Checks on extinction :

* extinction plus a Threshold value ≥ 0

* OR abs(extinction) < 3*errExtinction

* extinction < Ext_peak(Wavelength) if it does not belong to cirrus category.

For a first screening of unrealistic negative peak value, a value representative of an aerosol layer is used. In particular $b_{th} = 5 \cdot 10^{-7} \text{ m}^{-1} \text{ sr}^{-1}$ and $a_{th} = 2.5 \cdot 10^{-5} \text{ m}^{-1}$. These threshold values are set as 50 times the calibration value for the 1064nm.Wavelength and IB are scaled in the most conservative way, and threshold are keep constant over the different wavelengths.

Peak values had been set up studying the pdf of optical properties values for cirrus cases: even for cirrus cases the probability of having $ext > 0.005 \text{ m}^{-1}$ is less than 5‰. Correspondingly a $b_{peak} = 1.7 \cdot 10^{-4} \text{ m}^{-1} \text{ sr}^{-1}$ had set up assuming a lidar ratio of 30sr (most conservative value) has been assumed.

PQC_02

Integrated quantities of the aerosol properties should be positive and not exceed very extreme values not realistic for aerosol layer cases. The aerosol optical depth AOD (without any assumption in the lowest troposphere, so evaluated only in the portion of atmosphere covered by the provided profile) should remain under an established threshold for data not belonging to the cirrus category.

AOD > 0

AOD < AOD_{th} if it does not belong to cirrus category

The threshold values AOD_{th} = 1.5 had been set up studying the pdf of AOD for cirrus cases available on the EARLINET database until July 2016: even for cirrus cases the probability of having AOD > 1.5 is less than 5‰. The scaling with the wavelength had been set up in the most conservative way: considering that on average over Europe AERONET report 1.1-1.5 as typical Angstrom exponent, the threshold value was scaled with wavelength considering an Angstrom value of 0, so no wavelength dependence at all.

PQC_03

As for AOD, integrated quantities of the aerosol backscatter should be positive and not exceed very extreme values not realistic for aerosol layer cases. The aerosol integrated backscatter IB (without any assumption in the lowest troposphere, so evaluated only in the portion of atmosphere covered by the provided profile) should remain under an established threshold for data not belonging to the cirrus category.

IB > 0

IB < IB_{th} if it does not belong to cirrus category

Starting from the AOD_{th} discussed above, the $IB_{th}=0.05 \text{ sr}^{-1}$ is defined assuming a low (30 sr) lidar ratio value (most conservative choice) for converting extinction into backscatter and an Angstrom of 0 for scaling with the wavelength.

PQC_04

This procedure checks the Lidar Ratio (S) values when Extinction and Backscatter are provided in the same file (i.e. are provided at the same vertical resolution). Lidar ratio is defined as positive value and values are typically between 10 and 120 sr. A wider window is defined for excluding not realistic values without constraining the pdf of the S values.

$S \in [0; 200] \text{ sr}$ within $3\sigma S$

However lidar ratio is an intensive property so it is defined only where aerosols are significantly present. So the limit on S should be valid only where an aerosol layer is present, so where extinction and/or backscatter exceed a certain value that can be treat as a minimum aerosol layer detection limit and with a limited statistical uncertainty. In particular the following check is performed:

If extinction > $Ext_{dect}(Wavelength)$ AND $(errExtinction/extinction) < 50\%$

AND If backscatter > $Bck_{dect}(Wavelength)$ AND $(errBackscatter/backscatter) < 50\%$

* LidarRatio must be between [0 - 200] sr in $3 * ErrorLidarRatio$

$Bck_{dect} = 5 \cdot 10^{-7} \text{ m}^{-1} \text{ sr}^{-1}$ and $Ext_{dect} = 2.5 \cdot 10^{-5} \text{ m}^{-1}$ and where the uncertainty on extinction and backscatter is lower than 50%.