

# EARLINET Data Quality Check

## Action report

---

1. Introduction.....	3
2. QC procedure v3.1.....	5
3. QC procedure v3.0.....	6
4. QC procedure v2.0.....	7
5. QC procedure v1.1.....	8
6. QC procedure v1.0.....	9
7. QC extensive procedure v1.0 .....	11
8. QC extensive procedures v0.0 .....	13
9. QC procedure v0.0.....	13

---

*V3.0*

*25<sup>th</sup> March 2024*

---

## 1. Introduction

Quality Check procedure on datafile is a useful and needed tool which serves two main groups of users: Data Originator and External users.

Data Originators take advantage of QC procedures because this avoids the presence of non-useful and in some cases error-contaminated data in the dataset they are responsible for. In the era of Horizon 2020, during which Open Data is the keyword, QC procedures are a first tool for Data Originator for limiting and assuring an established quality of the data, a tool to be accomplished by quality assurance procedures at instrumental and retrieval level in the general framework of complete traceability.

On the other hand, data are one of the most important outputs of a Research Infrastructure and they are potentially used by experts but also by administrative bodies and finally common citizens. The QC procedures could drastically reduce the time needed for screening the data and will assure the quality of data to the final users.

As concerns EARLINET, the network started its activities in 2000 based on innovative lidar instruments for aerosol detection over Europe which agreed on basic format of the data and the database was set up in a collaborative way. Since then, the network has enlarged its activities, detected parameters and stations. The responsibility of the quality of the data always relied on PIs, but the situation changed dramatically with the increase of the EARLINET participants. Going now in the Research Infrastructure path, the setup of automatic QC procedure is a stringent request to be meet. The needs of the network were different in different periods, and based on that some QC procedures were set up.

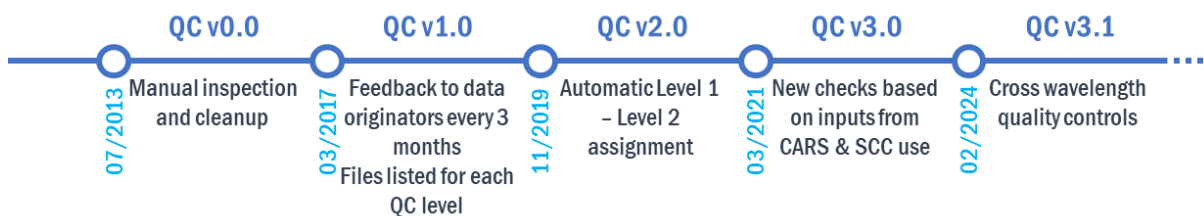


Figure 1. On-fly QC temporal evolution

The whole EARLINET operation period can be divided into two main periods: from EARLINET beginning to 23 February 2017 and after 23 February 2017. On 23 February 2017, a new more comprehensive set of on-fly quality checks have been implemented. Before, an extensive quality check was performed only once and through manual inspection of the data in July 2013. On 23 February 2017, all the data were the object of an extensive off-line automatic QC procedure. Since March 2017, an off-line quality check procedure has been systematically performed every 3 months. The QC procedures running since March 2017 provides direct feedback to Data originators and information to the users about the level of QC compliance for each file.

In November 2019, a new release of the QC accomplished with the need of automatic and fast assignment of Level 2 label to the data.

A new release of the QC occurred on 25 March 2021. This date opens a new chapter into the QC procedures using for the first time ever the expertise of the ACTRIS Center for Aerosol Remote Sensing (CARS) for approving the configuration used for the retrieval of the optical properties.

This document shortly describes the evolution of these procedures.

## 2. QC procedure v3.1

On February 2024, the multiproduct quality control has been implemented for the data that is automatically uploaded to the EARLINET database after it is processed by the SCC.

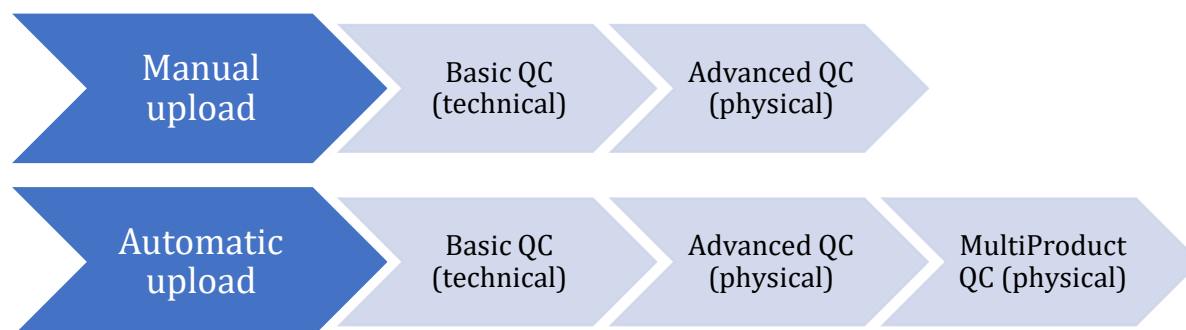
This procedure aims to improve the quality of the optical products uploaded automatically into the EARLINET database, and it is applied to all the optical products corresponding a single measurement in a combined way, to select the products with more information and to check the consistency between products at different wavelengths. The MPQC (MultiProduct Quality Control) does the following:

- a) If the dataset contains multiple files at the same wavelength, it selects the product that contains the highest amount of information.
- b) It performs cross-wavelength measurement QC checks on backscatter, lidar ratio, and Ångström exponent.

The automatically uploaded products can only be labeled as level 2 if they pass these new quality controls, as well as the already existing basic and advanced quality controls from v3.0.

For the optical products uploaded through the [data.earlinet.org](https://data.earlinet.org) interface, the quality control procedures applied are the same than in the version 3.0.

Therefore, these are the current quality control procedures applied to the optical products since February 1<sup>st</sup>, 2024, depending on how they are uploaded into the EARLINET database:



*Figure 2. Quality Controls applied to optical products depending on their upload procedure into the EARLINET database.*

More detailed quality control information can be found [in the new QC document \(replace with link\)](#).

### 3. QC procedure v3.0

On March 2021, the following new QC procedures have been implemented:

- a) all data submitted to the database after 25 March 2021 obtained not using the Single Calculus Chain with configuration approved by CARS will be labeled as Level 1 (AQC10).
- b) all data submitted to the database after 25 March 2021 will be labeled as Level 1 if the atmospheric density is evaluated using standard atmosphere model (AQC08).
- c) old data labeled as cirrus and not reporting the cloud mask as vertical information are labeled as Level 1 data because special attention is needed when a user would like to use it for aerosol investigation (AQC09).
- d) a technical control about the altitude range is added limiting the validity range from 0m asl, up to 50 km. Data not compliant with such technical control will not be available in the release of the updated version data (BQC12). Data submitted to the EARLINET database after 25 March 2021 not compliant with this technical control will be rejected.
- e) technical control will not allow the presence of data with Stop Time equal to Start Time. Data not compliant with such technical control will not be available in the release of the updated version data (BQC 09). Data submitted to the EARLINET database after 25 March 2021 not compliant with this technical control will be rejected.
- f) a bug was fixed about the quality control procedure that with the previous version wrongly required as mandatory the presence of backscatter profile inside e-products (AQC04)

For the sake of completeness, the detailed description of complete quality check procedures is described in *EARLINET\_QC\_v3\_0\_20210325.pdf*.

## 4. QC procedure v2.0

On November 2019, the QC procedure has been reviewed to better accomplish the EARLINET/ACTRIS database reshaping. The Quality Control procedures are conducted exclusively on-fly during the uploading process.

There are two types of Quality Control procedures:

- a) Basic quality control (BQC): technical quality control on the submitted product.
- b) Advanced quality control (AQC): series of physical checks applied to the input product.

If one or more basic quality control fails, the submitted product is rejected and detailed feedback is shown to the data originator.

If the product overcomes the basic quality controls, it is sent in batch to the advanced quality controls. If the input product fails at least one advanced quality control the product is labeled as Level 1 otherwise Level 2.

## 5. QC procedure v1.1

On June 2018, the QC procedure has been reviewed to better accomplish the EARLINET/ACTRIS needs in terms of acceptance for the value of the station altitude.

The criterium n. 5 of the On Fly Quality Control Procedure (TQC\_05) has been changed to accept a difference in altitude respect to the official declared station altitude up to 60m. The altitude criterium needs to be more flexible with respect to what was decided at the beginning, to allow moving the system, for example from a lab on the ground floor to the roof. A maximum difference of 60 m is now allowed since 60m is typically the vertical resolution of the lidar data optical products.

For the sake of completeness, the detailed description of complete on-fly and off-line quality check procedures are described in *EARLINETQCcon\_fly\_v1\_1.pdf* and *EARLINETQCoff\_line\_v1\_1.pdf*.

Log files on the EARLINET web page reporting data:

- Passing on-fly + off-line QC procedures – QC 2.1 data
- Passing only on-fly QC procedures – QC 1.1 data
- Not compliant to on-fly QC procedures – QC 0.1 data



## 6. QC procedure v1.0

Since 23 February 2017, an on-fly QC procedure works directly in the file submission phase. The detailed description of on-fly quality check procedures is described in *EARLINETQCcon\_fly\_v1\_0.pdf*. In general, the on-fly QC procedure checks the conformity of the file content respect to the EARLINET file structure (<https://www.earlinet.org/index.php?id=125>) and is embedded in the developed technical quality check procedures.

Data which is not compliant to the EARLINET standards is rejected already in the submission phase. Automatic feedback is provided to the Data Originator reporting all the problems incurred for each rejected file, fostering the prompt resubmission of the data.

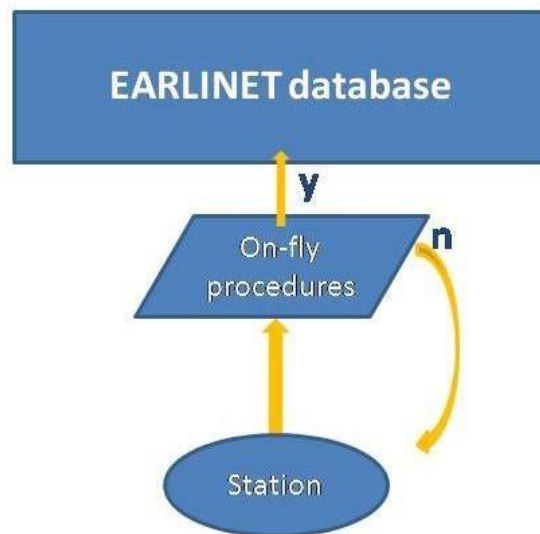


Figure 3. On-fly QC workflow

Because of the on-fly QC, since 23 February 2017, all the data reported at the EARLINET/ACTRIS database are automatically compliant to these standards.

Off-line quality checks (*EARLINETQCoff\_line\_v1\_0.pdf*) are run systematically on the data (every 3 months) and include scientific content QCs related to the content of the EARLINET files in terms of validity of the EARLINET measured parameters. Feedback is provided to the Data Originators. Information about the files compliant to the physical quality checks will be reported in log files on the EARLINET web page. All the data which are compliant to both all technical and physical QCs are labeled as QC2.0 data, while data passing only the on-fly QCs are labeled as QC1.0 data (*EARLINETQCcon\_fly\_v1\_0.pdf*).

Old data still stored on the database but not compliant to the on-fly quality checks (*EARLINETQCextended\_v1\_0.pdf*) are labeled as QC0.0 level data.

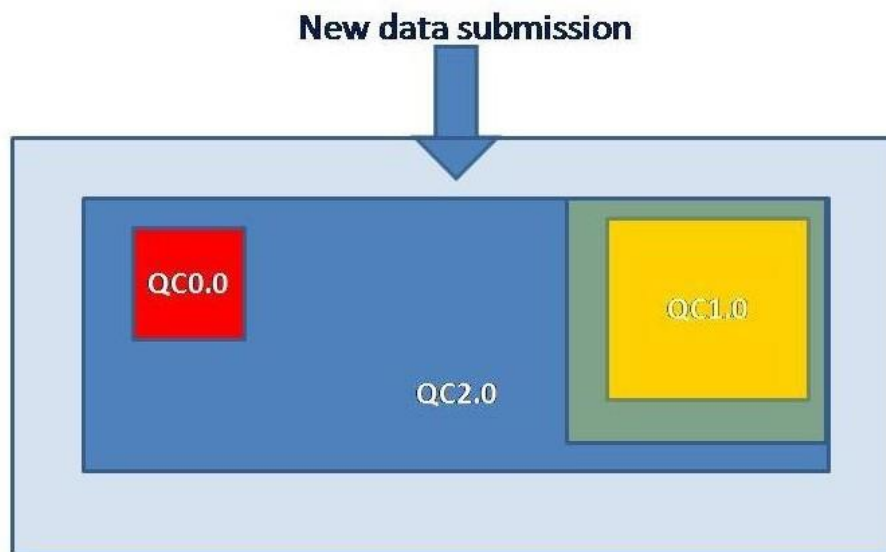


Figure 4. Schematic view of the EARLINET database. The three data groups related to different compliance criteria are reported in distinct colors. Shaded areas indicate the potential growth in the number of corresponding data groups.

Since 23 February 2017, therefore the data group QC0.0 will not be fed anymore thanks to the on-fly QCs, while both the QC1.0 and QC2.0 can increase in number. A reduction of additional QC1.0 is expected thanks to the feedback provided to the data originators.

### *Future developments*

When data versioning and new level products are implemented, an updated version of the EARLINET database interface will be released and will make data explorable through QC results. This could be implemented in diverse ways at the moment under investigation (through flag reported into the file or info into the relational database overlaying the EARLINET database) but however with information available to external users.

Additional checks more related to the physical aspects and to comparison vs climatological info and /or other ACTRIS information could be implemented. Further QCs could be implemented making use of the output of Lidar Calibration center and of the instrumental quality check procedures.

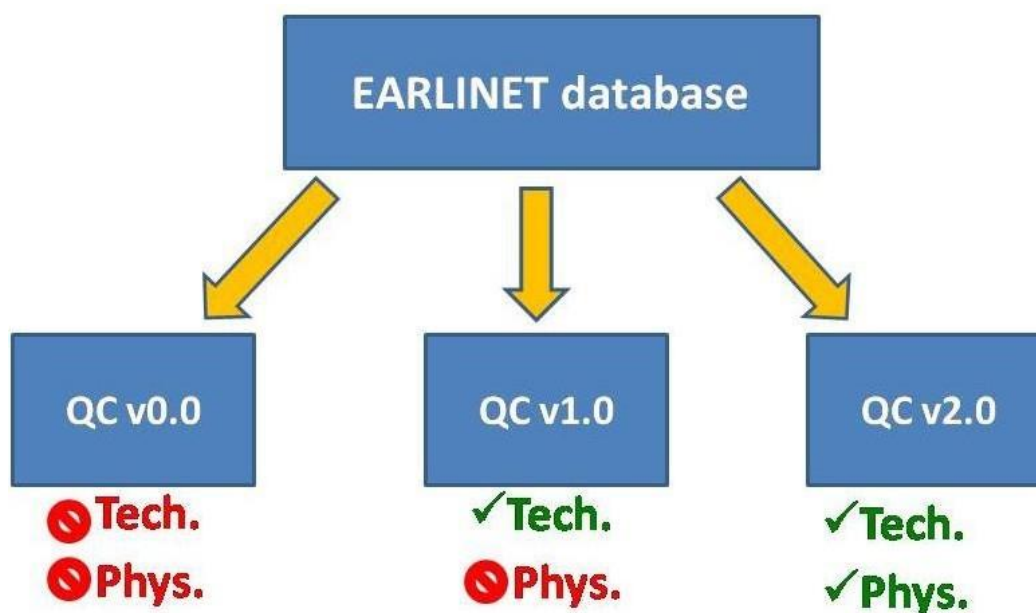
## 7. QC extensive procedure v1.0

Based on previous experiences on the screening of EARLINET data done file by file by eye, it is 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.

The details of all specific checks are reported in the devoted documentation (*EARLINETQCextended\_v1\_0.pdf*).

Data included in the EARLINET/ACTRIS database until 21 February 2017 were the object of an offline extended QC procedures (both technical and physical). The whole database (both internal and publicly available datasets) can be grouped into three subsets on the base of the QC procedure results.



*Figure 5. schematic view of the EARLINET database and how it consists of 3 data groups related to the compliance to technical and physical quality checks.*

Each file can be flagged as QC 2.0 (highest level of QC), as v1.0 and v0.0. If all the QC procedures are successfully passed, data are labeled as QC v2.0. If data passes technical QC but fails the physical ones it is flagged as QC1.0 and finally data failing the essential QC are kept on the database as QC v0.0.

All PIs, responsible for data not compliant with these procedures, are informed about the identified problems. This gives the possibility to inspect their own data (which mostly are affected by small technical

problems) and to be ready for a new data version submission as soon as the versioning of the data is implemented in the EARLINET/ACTRIS database.

In the meantime, the information about the QC run on the database is provided to the users through QClists available on the EARLINET webpage.

## **8. QC extensive procedures v0.0**

On July 2013, an extensive quality check was performed on the pre-existing EARLINET database(*EARLINETQC\_extended\_v0\_0.pdf*).

Some basic automatic checks were applied to the data in the uploading phase. Additionally, a file-by-file manual quality check of the climatological / CALIPSO /dust files has been conducted.

The Principal Investigators (PIs) were informed about this, and some data were corrected based on the manual inspection. All data published on the CERA database are the results of these procedures.

Data publicly available on the EARLINET database since July 2013 are persistently available at the EARLINET/ACTRIS database and through the ACTRIS Data Portal.

## **9. QC procedure v0.0**

Some basic automatic checks were applied to the data in the uploading phase since the beginning of EARLINET. The list of these checks is reported in the on-fly QCs section of the *EARLINETQC\_extended\_v0\_0.pdf*.