Quality Control

Data collection and management face the inherent complexity of multiple-sourced errors that, if undetected, can be ingested in the operational version/s of a database. To avoid this situation, and depending on the data source, there are different quality control (QC) layers that can be applied before incorporating a new dataset to any existent operational database (Zahumenský, 2004).

This document presents a preliminar proposal of checkings to be applied along the structure and formatting examination of the downloaded and post-processed files in the framework of the Earth Sciences department. This QC layer is based on the documents from the European Climate Assessment and Dataset (ECA&D, 2013) and the recomendations made by the WMO in the technical report of Zahumenský (2004).

Actually, our aim is to focus on rules that can be applied to any time-step, to ease the construction of general algorithms and the possibility to expand the protocol to new variables and/or other time-frequencies. The addition of spatial consistency and homogeneity tests requires further discussion for they are heavy time consuming and might be difficult to translate to general purpose algorithms. In this first version we deal with the following variables,

1) Precipitation

2) Surface air pressure

3) Geopotential Height

4) Maximum temperature

5) Minimum temperature

6) Temperature

7) Snow depth

8) Cloud cover

9) Humidity

10) Sunshine duration

11) Solar radiation fluxes

12) Mean wind speed

13) Maximum wind gust

Precipitation

- Must be equal or exceed 0 mm

$$\geqslant 0\,\mathrm{mm}$$

- **3-hourly**: must be less than 600 mm

$$\leqslant 600 \, \text{mm}$$

- **6-hourly**: must be less than 850 mm

$$\leqslant 850 \, \text{mm}$$

- **Daily**: must be less than 1800 mm

$$\leqslant 1800 \, \text{mm}$$

- **Monthly**: must be less than 9500 mm

$$\leqslant 9500 \, \text{mm}$$

- Must not be repetitive (i.e. exactly the same amount) for 10 timesteps in a row if amount is larger than 1.0 mm

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row if amount larger than 5.0 mm

Surface air pressure

- Must exceed 850.0 hPa

- Must be less than 1080.0 hPa

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row

Geopotential height

- Must be equal or exceed 0 m

$$\geqslant 0 \, \text{m}$$

- Must be less than 50000 m

$$\leqslant 50000 \, \text{m}$$

- Must not be repetitive (i.e. exactly the same amount) for 10 timesteps in a row

Maximum temperature (atmosphere and ocean)

- Must exceed -90°C

$$\geqslant -90°\text{C}$$

- Must be less than 60°C

$$\leqslant 60°\text{C}$$

- Must exceed or equal the corresponding timestep minimum temperature (if exists)

$$\geqslant T_{mn}$$

- Must exceed or equal the corresponding timestep mean temperature (if exists)

$$\geqslant T_g$$

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row

- Must be less than the long term average maximum temperature for that timestep + 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\leqslant [\bar{T}_{mx} + (5 \times \sigma_{T_{mx}})]_i$$

- Must be greater than the long term average maximum temperature for that timestep − 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\geqslant [\bar{T}_{mx} - (5 \times \sigma_{T_{mx}})]_i$$

Minimum temperature (atmosphere and ocean)

- Must exceed -90°C

$$\geqslant -90°\text{C}$$

- Must be less than 60°C

$$\leqslant 60°\text{C}$$

- Must be less or equal to the corresponding timestep maximum temperature (if exists)

$$\leqslant T_{mx}$$

- Must be less or equal to the corresponding timestep mean temperature (if exists)

$$\leqslant T_g$$

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row

- Must be less than the long term average minimum temperature for that timestep + 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\leqslant [\bar{T}_{mn} + (5 \times \sigma_{T_{mn}})]_i$$

- Must be greater than the long term average minimum temperature for that timestep − 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\geqslant [\bar{T}_{mn} - (5 \times \sigma_{T_{mn}})]_i$$

Temperature (atmosphere and ocean)

- Must exceed -90°C

$$\geqslant -90°\text{C}$$

- Must be less than 60°C

$$\leqslant 60°\text{C}$$

- Must exceed or equal the corresponding timestep minimum temperature (if exists)

$$\geqslant T_{mn}$$

- Must be less or equal to the corresponding timestep maximum temperature (if exists)

$$\leqslant T_{mx}$$

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row

- Must be less than the long term average mean temperature for that timestep + 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\leqslant [\bar{T}_g + (5 \times \sigma_{T_g})]_i$$

- Must be greater than the long term average mean temperature for that timestep − 5 times the standard deviation (computed for a 5 day window centred on that timestep over the whole period)

$$\geqslant [\bar{T}_g - (5 \times \sigma_{T_g})]_i$$

Snow depth

- Must be equal or exceed 0 cm

$$\geqslant 0 \, \text{cm}$$

- Must be less than 1100 cm

$$\leqslant 1100 \, \text{cm}$$

- Must not be repetitive (i.e. exactly the same amount) for 10 timesteps in a row if amount is larger than 1.0 cm

- Must not be repetitive (i.e. exactly the same amount) for 5 timesteps in a row if amount larger than 5.0 cm

Cloud cover

- Must exceed or equal 0 oktas

$$\geqslant 0 \, \text{oktas}$$

- Must be less or equal than 8 oktas

$$\leqslant 8 \, \text{oktas}$$

Humidity

- Must exceed or equal 0.0%

$$\geqslant 0\%$$

- Must be less or equal than 100.0%

$$\leqslant 100.0\%$$

Sunshine duration

- Must exceed or equal 0.0 h

$$\geqslant 0.0 \, \text{h}$$

- Must be less or equal than 24.0 h

$$\leqslant 24.0 \, \text{h}$$

Solar radiation fluxes

- Must exceed or equal 0.0 W/m^2

$$\geqslant 0.0 \, \text{W}/m^2$$

- Must be less or equal than 1362.0 W/m^2

$$\leqslant 1362.0 \, \text{W}/m^2$$

Wind speed (u, v and module)

- Must exceed or equal 0.0 m/s

$$\geqslant 0.0 \, \text{m/s}$$

- Must be less or equal than 140.0 m/s

$$\leqslant 140.0 \, \text{m/s}$$

- Must not be repetitive (i.e. exactly the same value) for 5 timesteps in a row if value is larger than 2.0 m/s

Maximum wind gust (u, v and module)

- Must exceed or equal 0.0 m/s

$$\geqslant 0.0 \, \text{m/s}$$

- Must be less or equal than 140.0 m/s

$$\leqslant 140.0 \, \text{m/s}$$

- Must not be repetitive (i.e. exactly the same value) for 5 timesteps in a row if value is larger than 4.0 m/s

Bibliography

ECA&D (2013). Algorithm theoretical basis document. Tech. rep., Royal Netherlands Meteorological Institute (KNMI).

Zahumenský I. (2004). Guidelines on quality control procedures for data from automatic weather stations. Tech. rep., World Meteorological Organization.