



## JRC TECHNICAL REPORTS

# Quality Checks Meteorological data (QUACKME)

*Architectural & Technical Guide*

Bratu, M.

This publication is a Technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

JRC Science Hub

<https://ec.europa.eu/jrc>

© European Union, 2018

The reuse of the document is authorised, provided the source is acknowledged and the original meaning or message of the texts are not distorted. The European Commission shall not be held liable for any consequences stemming from the reuse.

How to cite this report: Bratu M., *Quality Checks Meteorological Data (QUACKME) – Technical Guide*, EUR, November 2018

All images © European Union 2018

*Printed in Italy*

# Contents

|   |    |
|---|----|
| Annotations table.....                    | 4  |
| Version summary.....                      | 5  |
| 1. Requirements.....                      | 6  |
| 2. Architecture.....                      | 7  |
| 3. QUACKME Steps.....                     | 11 |
| 3.1 Input Converter.....                  | 11 |
| 3.1.1 Input Data Format.....              | 11 |
| 3.1.2 Internal Data Format.....           | 11 |
| 3.1.3 MOS file Format.....                | 12 |
| 3.2 Weak checks level.....                | 13 |
| 3.3 Aggregation.....                      | 19 |
| 3.3.1 Aggregation.XML structure.....      | 19 |
| 3.3.2 Aggregation formulas.....           | 26 |
| 3.4 Heavy checks.....                     | 31 |
| 3.5 Threshold checks.....                 | 33 |
| 3.5.1 Threshold file generator.....       | 35 |
| 3.6 SConverter module.....                | 37 |
| 3.7 RRR file generator module.....        | 40 |
| 3.8 SG GUI.....                           | 43 |
| 4. QUACKME Configurations.....            | 47 |
| 4.1 Messages structure.....               | 47 |
| 4.2 Messages behavior.....                | 48 |
| 4.3 Workflow configuration structure..... | 48 |

## Annotations table

| Annotation | Description  |
|------------|--|
| AMDAC      | Actual Meteorological DAtabase Construction        |
| QUACKME    | QUAlity ChecKs MEteorological Data                 |
| SG         | Specialists Group                                  |
| MOS        | Model Output Statistics                            |
| WMO        | World Meteorological Organization                  |
| ECMWF      | European Centre for Medium-Range Weather Forecasts |
| ROI        | Region Of Interest                                 |

## Version summary

| Version | Description   |
|---------|---|
| 1.2.0   | More explanation about Aggregation.XML configuration<br>List all interpolated properties before aggregation<br>Management of empty KO files inside modules  |
| 1.2.1   | Added interpolation configuration inside Aggregation.XML<br>Added new aggregated property: Absolute Daily Minimum Temperature<br>Change on aggregation of precipitation logics, modify periods for daily mean values (Vapour pressure, Total Cloud, Vapour Pressure Deficit, Wind Speed, Slope) |
| 1.2.2   | Change logics for –m option of WeakChecks module  |
| 1.3.0   | Added ROI – CHINA (3.3.1.2, 3.3.2.2, 3.5)<br>Added the RRR Generator (3.7)  |
| 1.3.1   | New weak check - automatic correction of TD after comparison with TT, RH difference of consecutive observations changed (3.2)   |
| 1.3.2   | Aggregation formulas – maximum temperature, check the TX12 at 15 UTC  |
| 1.4.0   | MOS management  |
| 1.4.1   | Using MOS file inside RRRGenerator  |
| 1.4.2   | Fixing some ISSUE on various modules (3.2, 3.3.1, 3.4)  |
| 1.4.3   | Manage check of consecutive hourly values for integer values of TT, TD, D_RH  |
| 1.4.4   | Change the interval for minimum temperature checks (WeakChecks)   |
| 1.4.5   | WeakChecks – use altitude for DIR consecutive check, check of hourly Sunshine > 0 & Hourly Radiation = 0 assigned to Radiation<br>Aggregation – Add Expected.Values parameter<br>HeavyChecks – Measured Sunshine > 1 and Measured Radiation = 0   |
| 1.4.6   | WeakChecks – Replace total cloud with low cloud when total cloud is missing and low cloud is valid  |
| 1.4.7   | 3H precipitation generation, Unit of measure for Visibility parameter   |

# 1. Requirements

The main purpose of JRC is to realize a new Quality Checks Meteorological Data Software (QUACKME) to manage the meteorological observations from ground stations, in the way to remove the gap and problems of the actual AMDAC package and, at the same moment, to give more flexibility to configure and manage the QUACKME behavior.

QUACKME must satisfy some criteria:

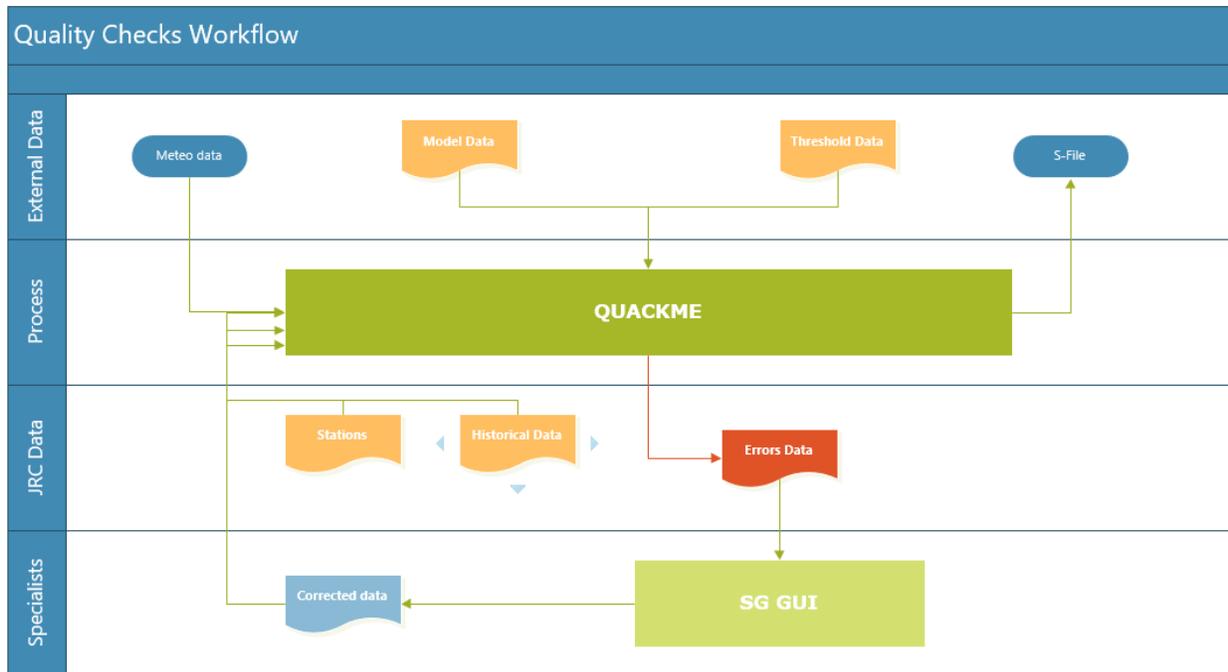
- need to be realize using open source programming language (like R, Python or similar)
- need to be flexible to modify and configure checks
- the sources must be clear and easy to modify
- can allow to use external data for checks where is necessary
- can easily be adapted to different format of input data
- must be able to trace errors for quickly and easy problem solving
- can run on different platform without many changes on source codes
- friendly user interface to allow the specialist to correct reported events
- report wrong or suspicious observations in the clear way for the specialists

Input data need to include hourly observation for the following properties:

| <b>Property description</b> | <b>Note</b>                    | <b>Unit of Measure</b>         |
|-----------------------------|--------------------------------|--------------------------------|
| Station number              | Numerical value                |                                |
| Day Time                    | Numerical value (YYYYMMDDHH24) | Including hour with 24H format |
| Air temperature             |                                | Celsius degrees                |
| Dew point temperature       |                                | Celsius degrees                |
| Wind direction              |                                | Degrees                        |
| Wind speed                  |                                | Knots                          |
| Solar radiation             |                                | J/m <sup>2</sup>               |
| Sunshine duration           |                                | Minutes                        |
| Precipitation               |                                |                                |
| Relative humidity           | Percentage between 0 and 100   |                                |
| Atmospheric pressure        |                                | hPa                            |
| Cloud cover                 |                                | Octas                          |
| Visibility                  |                                | KiloMeters                     |
| Snow depth                  |                                | Centimeters                    |

## 2. Architecture

Below it is possible to view an architectural design for the Quality Checks Software area.



### Architecture detail

The quality checks software part will be implemented like a workflow where each step will produce a part of the input of the next step.

In general, for a step will have:

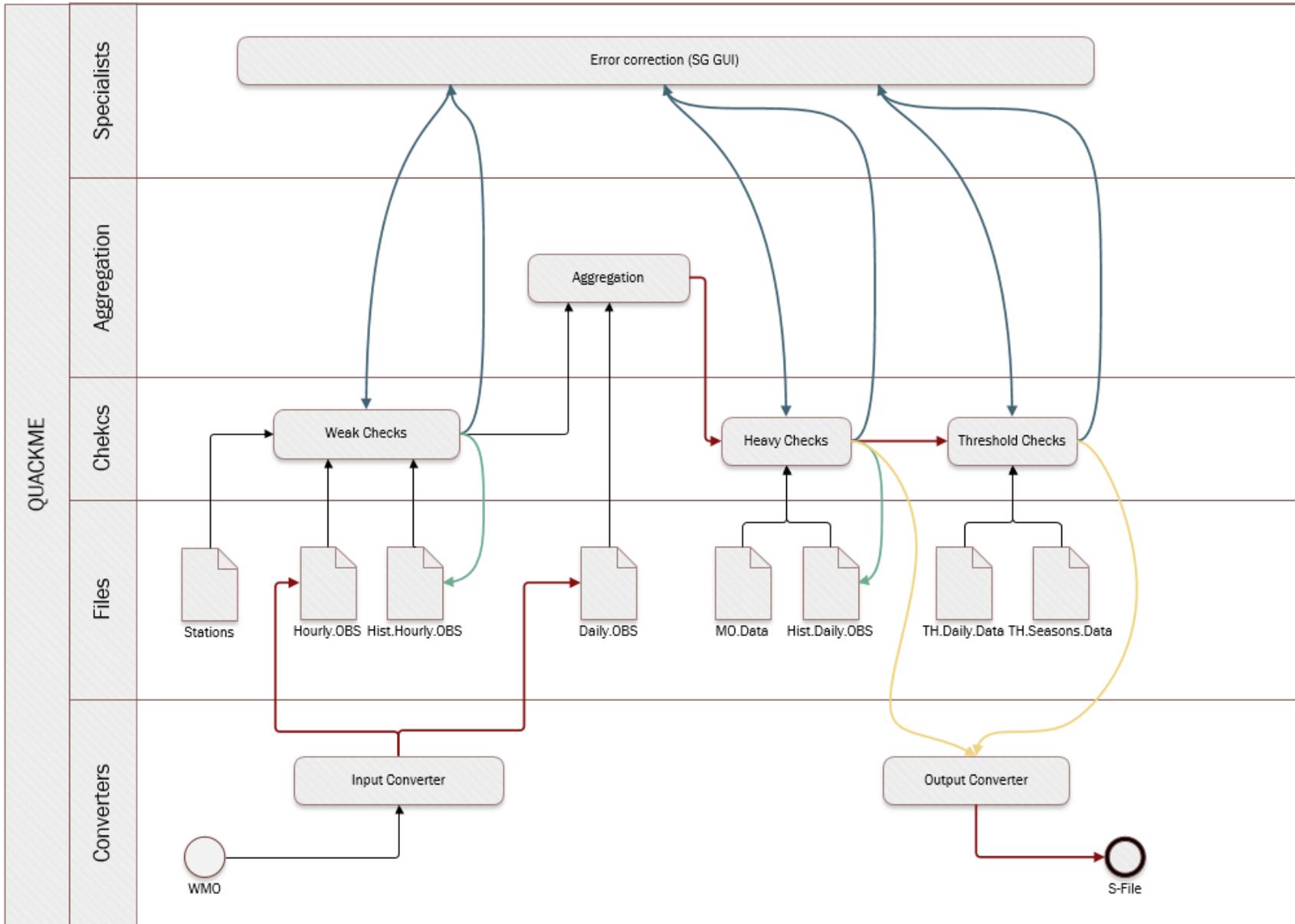
- Input:
  - the output with correct data of the previous step if exists
  - historical data, if necessary
  - external data from suppliers, if necessary
  - threshold data
- Output
  - file with correct data (using an internal data format)
  - file with errors/problems (xml format)
  - historical file, where necessary
  - data for suppliers (S-File)

In detail, the workflow will contains the following step:

- **Input Converter** - with the role to convert data from meteorological data provider (hourly observation) to an internal data format (see the chapter 3.2).
  - **Weak Checks** – with the role to execute the weak checks
  - Input:

- Input data from meteorological data provider (converted into an internal data format) and, eventually, corrected data (from the SG)
  - It can use the historical hourly observation data and the stations data
- Output:
  - File with correct hourly observation data (internal format)
  - New historical hourly observation data relative to current date
  - File with errors (xml format)
- **Aggregation** – with the role to transform hourly observation into daily aggregated data:
- Input:
  - The correct hourly observation data
  - Daily observation data, if presents
- Output:
  - Daily aggregated data
- **Heavy Checks** – with the role to executed complex checks:
- Input:
  - Daily aggregated data
  - Daily historical data
  - Model Output Statistics data
- Output:
  - File with correct data (internal format)
  - File with errors (xml format)
- **Threshold Checks** – with the role to executed complex checks:
- Input:
  - Result of heavy checks data
  - Daily historical data
  - Daily and Season Threshold data
- Output:
  - File with correct data (internal format)
  - File with errors (xml format)
- **SConverter** – with the role to convert the output of Quackme chain to the S-File format (ALTERRA):
- Input:
  - Correct data produce by Heavy Checks step or Model Checks step
- Output:
  - S-File





### 3. QUACKME Steps

#### 3.1 Input Converter

To have more flexibility on managing different format of input data was decide to create an input converter. In fact, for each type of input format, will create an interface (to be called from the main workflow) to convert the input data into internal format data.

This operation will give the possibility to JRC to use more suppliers for meteorological data without change the workflow and checks.

The converter are checking duplicate lines in correspondence of <Station, Day&Time> and will raise an error message if will find at least one duplicate line.

At the same moment, during the conversion step the value = 9 for N and L column will be replaced with value 8.

##### 3.1.1 Input Data Format

The input data format will depend by the provider choose by JRC.

##### 3.1.2 Internal Data Format

The internal data format is a text file organized by columns. His structure will include:

- First line is dedicated to the header, with the name of the columns
- Any other line will contain values for each observation separated by a single space character
- The file cannot contains empty line
- Two formats will be used: one for hourly observation and one for daily observation

Bellow you can the find the conversion table between the CSV columns and DAT file:

| Property description              | DAT columns | Note                           | CSV Columns        | Unit of Measure |
|-----------------------------------|-------------|--------------------------------|--------------------|-----------------|
| <b>Station number</b>             | Station     | Numerical value                | INDEX              |                 |
| <b>Day Time</b>                   | DayTime     | Numerical value (YYYYMMDDHH24) | YYYY, MM, DD, HHmm |                 |
| <b>Air Temperature</b>            | TT          | Numerical value                | TL                 | Celsius Degree  |
| <b>Dew point temperature</b>      | TD          | Numerical value                | TD                 | Celsius Degree  |
| <b>Maximum temperature 1 Hour</b> | TX1         | Numerical value                | TX1                | Celsius Degree  |
| <b>Minimum temperature 1 Hour</b> | TN1         | Numerical value                | TN1                | Celsius Degree  |
| <b>Maximum temperature 6 Hour</b> | TX6         | Numerical value                | TX6                | Celsius Degree  |

|  |      |                               |                 |                |
|--|------|-------------------------------|-----------------|----------------|
| <b>Minimum temperature 6 Hour</b>            | TN6  | Numerical value               | TN6             | Celsius Degree |
| <b>Maximum temperature 12 Hour</b>           | TX12 | Numerical value               | TX12            | Celsius Degree |
| <b>Minimum temperature 12 Hour</b>           | TN12 | Numerical value               | TN12            | Celsius Degree |
| <b>Hourly Precipitation</b>                  | PREC | Numerical value               | RR1h            | mm             |
| <b>6 Hour Precipitation</b>                  | PR06 | Numerical value               | RR06            | mm             |
| <b>24 Hour Precipitation</b>                 | PR24 | Numerical value               | RR24            | mm             |
| <b>Precipitation</b>                         | RR   | Numerical value               | RR              | mm             |
| <b>Precipitation interval</b>                | TR   | Numerical value (Hour number) | TR              |                |
| <b>Snow depth</b>                            | Snow | Numerical value               | SNO             | cm             |
| <b>Wind direction</b>                        | DIR  | Numerical value               | DIR             | Degree         |
| <b>Wind speed at 10 m</b>                    | FF   | Numerical value               | FF              | m/s            |
| <b>Cloud cover</b>                           | N    | Numerical value               | N               | Octas          |
| <b>Low cloud cover</b>                       | L    | Numerical value               | L               | Octas          |
| <b>Radiation 1 Hour</b>                      | RD   | Numerical value               | GL1h            | J/m2           |
| <b>24 Hour Radiation</b>                     | RD24 | Numerical value               | GL24            | J/m2           |
| <b>Atmospheric pressure at station level</b> | AP   | Numerical value               | QFE             |                |
| <b>Atmospheric pressure at station level</b> | QFF  | Numerical value               | QFF             |                |
| <b>Sunshine 1 Hour</b>                       | SH   | Numerical value               | Sh              | minutes        |
| <b>Sunshine 24 Hour</b>                      | SH24 | Numerical value               | SS24            | minutes        |
| <b>Visibility</b>                            | VIS  | Numerical value               | VIS             | km             |
| <b>Relative Humidity</b>                     | RH   | Numerical value               | (added by code) | %              |
| <b>Soil State</b>                            | SOIL | Numerical value               | GS              |                |

### 3.1.3 MOS file Format

The MOS file (daily file) contains the hourly MOS values for the interval: 00 of day in elaboration – 06 next day.

Bellow you can the find the conversion table between the CSV columns and DAT file:

| Property description            | DAT columns | Note                           | CSV Columns        | Unit of Measure |
|---------------------------------|-------------|--------------------------------|--------------------|-----------------|
| <b>Station number</b>           | Station     | Numerical value                | INDEX              |                 |
| <b>Day Time</b>                 | DayTime     | Numerical value (YYYYMMDDHH24) | YYYY, MM, DD, HHmm |                 |
| <b>Air Temperature</b>          | TT          | Numerical value                | TL                 | Celsius Degree  |
| <b>Dew point temperature</b>    | TD          | Numerical value                | TD                 | Celsius Degree  |
| <b>Hourly Precipitation</b>     | PREC        | Numerical value                | RR1h               | mm              |
| <b>6 Hour Precipitation</b>     | PR06        | Numerical value                | RR06               | mm              |
| <b>24 Hour Precipitation</b>    | PR24        | Numerical value                | RR24               | mm              |
| <b>Wind speed at 10 m</b>       | FF          | Numerical value                | FF                 | m/s             |
| <b>1H Maximum precipitation</b> | RRRX1       | Numerical value                | LAYER1             | mm              |
| <b>6H Maximum precipitation</b> | RRRX6       | Numerical value                | LAYER2             | mm              |
| <b>12 Maximum precipitation</b> | RRRX12      | Numerical value                | LAYER3             | mm              |
| <b>24 Maximum precipitation</b> | RRRX24      | Numerical value                | LAYER4             | mm              |

### 3.2 Weak checks level

The weak checks level will execute checks (based only on hourly observations data) for the following parameters:

- Air temperature
- Minimum and maximum temperature
- Dew point temperature
- Cloud cover
- Amount of precipitation
- Pressure at sea level
- Wind speed & direction
- Duration of sunshine
- Relative humidity
- Atmospheric pressure
- Solar radiation

Currently are not configured checks for visibility and state of the soil.

In detail, the following weak checks need to be implement:

- Air temperature checks

- Air Temperature values outside -80 / +60 Celsius degrees. In this case, the value will be consider WRONG.
- Air temperature between two consecutive records (hourly observations) < 0.1 degree. The alert will be raise after the error has identifier at least 15 times. In this case, the value will be consider SUSPICIOUS and the error raise in correspondence of the 16-th observations, only if the temperature values are not integer values
- Air temperature between two consecutive records (either hourly or 3-hourly) > 15 degree. In this case, the value will be consider SUSPICIOUS.
- If the absolute difference between the observed air temperature and the correspondent MOS air temperature is greater than 12 degrees the observed air temperature will be considered wrong
- If the absolute difference between observed air temperature and the correspondent MOS air temperature is greater than 8.5 degress and less or equal to 12 degress then the observed air temperature will be considered suspicious
- All Minimum temperature (one hour, 6 hours, 12 hours) checks
  - Minimum temperature value outside -80 / +40 Celsius degrees. In this case, the value will be consider WRONG.
  - Minimum temperature between two consecutive records (hourly observations) < 0.1 degree. ~~For TN6 and TN12 the alert will be raise after the error has identifier at least 15 times.~~ In this case, the value will be consider SUSPICIOUS and the error raise in correspondence of the 16-th observations.
  - Minimum temperature between two consecutive records (either hourly or 3-hourly) > 15 degree. In this case, the value will be consider SUSPICIOUS.
- All Maximum temperature (one hour, 6 hours, 12 hours) checks
  - Maximum temperature value outside -80 / +60 Celsius degrees. In this case, the value will be consider WRONG.
  - Maximum temperature between two consecutive records (hourly observations) < 0.1 degree. ~~For TX6 and TX12 the alert will be raise after the error has identifier at least 15 times.~~ In this case, the value will be consider SUSPICIOUS and the error raise in correspondence of the 16-th observations.
  - Maximum temperature between two consecutive records (either hourly or 3-hourly) > 15 degree. In this case, the value will be consider SUSPICIOUS.
- Dew point temperature checks
  - Dew Point Temperature outside -80 / 35 Celsius degrees. In this case, the value will be consider WRONG.
  - Dew point temperature > Air Temperature. If the difference between dew point temperature and air temperature is less tan 1 Celsius degree than the dew point temperature are replaced with the air temperature, else the value will be consider WRONG.
  - Dew point temperature between two consecutive records (hourly observations) > 11 degree. In this case, the value will be consider SUSPICIOUS.
  - Dew point temperature between two consecutive records (hourly observations) < 0.05 degree. The alert will be raise after the error has identifier at least 20 times. In this case, the value will be consider SUSPICIOUS and the error raise in correspondence of the 16-th observations only if the dew point temperatures are not integer values

- If the absolute difference between the observed dew point temperature and the correspondent MOS dew point temperature is greater than 15 degrees the observed dew point temperature will be considered wrong
- If the absolute difference between observed dew point temperature and the correspondent MOS dew point temperature is greater than 11 degrees and less or equal to 15 degrees then the observed dew point temperature will be considered suspicious
- Relative humidity checks
  - Relative Humidity outside 0 - 100 %. In this case, the value will be consider WRONG.
  - Relative humidity between two consecutive records (hourly observations) < 0.5% and the both values are less than 95%. The alert will be raise after the error has identifier at least 20 times. In this case, the value will be consider SUSPICIOUS and the error raise in correspondence of the 16-th observations, only if the air temperature and dew point temperatures are not integer values.
- Atmospheric pressure checks
  - Atmospheric Pressure outside 500 – 1100 hPa. In this case, the value will be consider WRONG.
  - Atmospheric Pressure between two consecutive records (either hourly or 3-hourly) > 15 hPa. In this case, the value will be consider SUSPICIOUS.
- Atmospheric pressure reduce to MSL checks
  - Atmospheric Pressure reduce to MSL outside 950 – 1060 hPa. In this case the values will be consider WRONG
- Wind direction checks
  - Wind direction outside 0-360 degrees. In this case, the value will be consider WRONG.
  - Wind direction between two consecutive records (hourly observations) < 5 degree. This control also check if both wind speed are great than 4.6 m/s. The alert message will be raise after the 20-th error and only if the altitude of the station is less or equal than 1000m. In this case, the value will be consider SUSPICIOUS.
  - Wind direction between two consecutive records (hourly observations) > 40 degree. This control also check if both wind speed are great than 4.6 m/s. The alert message will be raise after the 5-th error and only if the altitude of the station is less or equal than 1000m. In this case, the value will be consider SUSPICIOUS.
- Wind speed checks
  - Wind speed outside 0 – 75 m/sec. In this case, the value will be consider WRONG.
  - Wind speed between two consecutive records (hourly observations) < 0.5 m/sec. This control also check if both wind speed are great than 4.6 m/s. The alert message will be raise after the 15-th error. In this case, the value will be consider SUSPICIOUS.
  - Wind speed between two consecutive records (hourly observations) > 15 m/sec. The alert message will be raise after the 10-th error. In this case, the value will be consider SUSPICIOUS.
  - ~~○ Wind speed > 3.5 m/ and Wind direction is NA. In this case, the value will be consider SUSPICIOUS.~~

- If the absolute difference between the observed wind speed and the correspondent MOS wind speed is greater than 40 knots then the observed wind speed is considered wrong
  - If the absolute difference between the observed wind speed and the correspondent MOS wind speed is greater than 25 knots and less than 40 knots then the observed wind speed is considered suspicious
  - If the observed wind speed is 0 and the correspondent MOS wind speed is greater than 19 18 knots then the observed wind speed is considered suspicious
  - If the ratio between observed wind speed and correspondent MOS wind speed lies between 1.75 and 2.25 then the observed wind speed is considered suspicious
  - Total Cloud cover checks
    - Total cloud cover is replaced with low cloud cover value when total cloud is missing and low cloud value is valid
    - Cloud Cover outside 0-8. In this case, the value will be consider WRONG.
    - Cloud cover is 0 and precipitation > 0. In this case, the value will be consider SUSPICIOUS.
    - Cloud cover 8 and sunshine duration > 0. In this case, the value will be consider SUSPICIOUS.
  - Low Cloud cover checks
    - Low Cloud Cover outside 0-8. In this case, the value will be consider WRONG.
  - Solar radiation checks (hourly observation)
    - Solar radiation outside 0 -1600 W/m<sup>2</sup>. In this case, the value will be consider WRONG.
    - Sunshine duration > 0 and radiation is 0. In this case, the value will be consider SUSPICIOUS.
    - Solar radiation checks (24 hours observation)
    - Solar radiation outside 0 -38400 W/m<sup>2</sup>. In this case, the value will be consider WRONG.
  - Sunshine checks (hourly observation)
    - Sunshine duration > 60 minutes. In this case, the value will be consider WRONG.
    - Sunshine checks (24 hours observation)
    - Sunshine duration > 24 hours. In this case, the value will be consider WRONG.
  - Precipitation checks (hourly observation 6 hours observation, 24 hours observation)
    - Precipitation outside 0-400. In this case, the value will be consider WRONG.
    - Precipitation > 200. In this case, the value will be consider SUSPICIOUS.
    - For each precipitation properties (1H, 6H, 12H – not available yet, 24 H) need to calculate the following parameters, if the corresponding MOS data are available:
      - Standard deviation (SD):  $\min[(5 * \max[0.2, RRR_{mos}]), \max[1, \text{abs}(RRR_{1/6/12/24} - RRR_{mos})]]$
      - Scaled difference (F):  $(RRR_{mos} - RRR) / RRR_{mos}$
- Once these two parameters are calculate the following conditions must be

check:

- When  $\text{abs}(\text{RRRobs} - \text{RRRmos}) > (6 * \text{SD})$  AND  $\text{RRRobs} > 20$  AND  $F > 5$ , the observed precipitation property is considered wrong
- When  $\text{abs}(\text{RRRobs} - \text{RRRmos}) > (6 * \text{SD})$  AND  $\text{RRRobs} > \text{RRRmos}$  AND  
 $( (\text{RRRobs} \leq 10 \text{ AND } \text{RRRmos} \geq 0.5) \text{ OR } (\text{RRRobs} \leq 15 \text{ AND } \text{RRRmos} \geq 2.5))$  isn't satisfied , then the observed precipitation property is considered suspicious

A part the properties listed below, during the WeakChecks step are calculate some new properties useful to the Aggregation step:

- o **D\_E** = Saturation vapor pressure. To calculate this property we are using TT (air temperature) and TD (dew point temperature) properties into the formula (**ESAT(TT, TD)**) :  

$$6.1078 * \exp(\text{LVAP}(\text{TT}) / \text{RV} * (1./\text{Celsius2Kelvin}(0.) - 1./\text{Celsius2Kelvin}(\text{TD})),$$

Dove:

- o **LVAP** are calculated the Latent Heat of Vaporization, using the formula :  
 $(597 - 0.566 * \text{TT}) * 4186$
- o **RV** it's a constant value = **461.51**
- o **Celsius2Kelvin** it's a method that transform temperature from Celsius to Kelvin degree
- o **D\_VPD** = Vapor pressure deficit. To calculate this property we are using TT and TD properties. The vapor pressure deficit is calculated like difference between to saturation value:
  - o One, obtained using the saturation vapor pressure with only TT value
  - o Second, obtained using the saturation vapor pressure with TT and TD
  - o The formula :  $\text{ESAT}(\text{TT}, \text{TT}) - \text{ESAT}(\text{TT}, \text{TD})$
- o **D\_SLOPE**= Slope of saturation vapor pressure. The slope are calculated using the following formula:  

$$238.102 * 17.32491 * (\text{D}_E + \text{D\_VPD}) / (\text{TT} + 238.102)**2$$
- o **D\_RH** = Relative humidity. The relative humidity are calculated using the following formula:  

$$100 * \text{ESAT}(\text{TT}, \text{TD}) / \text{ESAT}(\text{TT}, \text{TT})$$

The WeakChecks module can use in input the errors file, too (XML format), after the errors was fixed by meteorologist. The "-r" option must include at least one KO file name and maximum two, separated by ";". When the KO file sent through the option "-r" are empty (contains only the tag "<Observations />") or are missing the module **will not executed the checks for the date for which the data are missing**.

Another input possibility for WeakChecks module is to specify the "-m <hour>" option. This option allow to realize an append operation between the input CSV file and the H.<day in elaboration>.hist file (present into the History Path):

- can be used for both area of interest : Europe & China

- the input CSV file must contains the hourly observation from <hour> + 1 of day in elaboration, meanwhile from the hourly history file will be read the observation of day in elaboration with time <= <hour>
- for example a command line of WeakChecks with -d "20191101" -m "06" has the following significant:
  - o day in elaboration 01.11.2019
  - o from the H.20191101.hist file will be read all observation with DayTime <= 2019110106
  - o the CSV must contains observations with DayTime >= 2019110107

An output file of WeakChecks module will contains the following properties:

| Property description               | Acronim | Note                           | Unit of Measure  |
|------------------------------------|---------|--------------------------------|------------------|
| <b>Station number</b>              | Station | Numerical value                |                  |
| <b>Day Time</b>                    | DayTime | Numerical value (YYYYMMDDHH24) |                  |
| <b>Air Temperature</b>             | TT      | Numerical value                | Celsius Degree   |
| <b>Dew point temperature</b>       | TD      | Numerical value                | Celsius Degree   |
| <b>Maximum temperature 1 Hour</b>  | TX1     | Numerical value                | Celsius Degree   |
| <b>Minimum temperature 1 Hour</b>  | TN1     | Numerical value                | Celsius Degree   |
| <b>Maximum temperature 6 Hour</b>  | TX6     | Numerical value                | Celsius Degree   |
| <b>Minimum temperature 6 Hour</b>  | TN6     | Numerical value                | Celsius Degree   |
| <b>Maximum temperature 12 Hour</b> | TX12    | Numerical value                | Celsius Degree   |
| <b>Minimum temperature 12 Hour</b> | TN12    | Numerical value                | Celsius Degree   |
| <b>Hourly Precipitation</b>        | PREC    | Numerical value                | mm               |
| <b>24 Hour Precipitation</b>       | PR24    | Numerical value                | mm               |
| <b>Precipitation</b>               | RR      | Numerical value                | mm               |
| <b>Precipitation interval</b>      | TR      | Numerical value (Hour number)  |                  |
| <b>Snow depth</b>                  | Snow    | Numerical value                | cm               |
| <b>Wind direction</b>              | DIR     | Numerical value                | Degree           |
| <b>Wind speed at 10 m</b>          | FF      | Numerical value                | m/s              |
| <b>Cloud cover</b>                 | N       | Numerical value                | Octas            |
| <b>Low cloud cover</b>             | L       | Numerical value                | Octas            |
| <b>Radiation 1 Hour</b>            | RD      | Numerical value                | J/m <sup>2</sup> |

|  |         |                                |                    |
|--|---------|--------------------------------|--------------------|
| <b>24 Hour Radiation</b>                     | RD24    | Numerical value                | J/m <sup>2</sup>   |
| <b>Atmospheric pressure at station level</b> | AP      | Numerical value                |                    |
| <b>Atmospheric pressure at station level</b> | QFF     | Numerical value                |                    |
| <b>Sunshine 1 Hour</b>                       | SH      | Numerical value                | minutes            |
| <b>Sunshine 24 Hour</b>                      | SH24    | Numerical value                | minutes            |
| <b>Visibility</b>                            | VIS     | Numerical value                | km                 |
| <b>Relative Humidity</b>                     | RH      | Numerical value                | %                  |
| <b>Saturation vapor pressure</b>             | D_E     | Numerical value,<br>Calculated | hPa                |
| <b>Relative Humidity</b>                     | D_RH    | Numerical value,<br>Calculated | %                  |
| <b>Vapor pressure deficit</b>                | D_VPD   | Numerical value,<br>Calculated | hPa                |
| <b>Slope of saturation vapor pressure</b>    | D_SLOPE | Numerical value,<br>Calculated | hPa/Celsius Degree |
| <b>Soil status</b>                           | SOIL    | Numerical value                |                    |

### 3.3 Aggregation

The main scope of this step is to aggregate hourly observation data to daily observation data for each station and for each observed date.

Like WeakChecks step, the Aggregation level are using specific configuration file (**Aggregation.XML**). The file contains the definitions for the daily properties to generate using the hourly observations of the current day and, eventually, of the previous and next day.

#### 3.3.1 Aggregation.XML structure

The **Aggregation.XML** file contains configurations about how to aggregate hourly values and to produce daily values. Is an XML file that contains a definition (XML structure) for each property to generate.

A property is identified by the tag *PROPERTY*, that is a simple tag, without attributes.

Inside the *PROPERTY* tag it's possible to define the following tags:

- NAME – (mandatory) that will define the name of the property (column) into the output file
- FormulaValue – (mandatory) that will define the formula to be applied to obtain the property's value. This tag can have the following attributes:
- Type (mandatory) that can have two values:

- "Fixed" when the property value will be retrieved from specific property and specific datetime. When the TYPE is Fixed is necessary to add the other attributes:
  - "Time" – that indicate the hour of the observation (like: 00, 06 or 23)
  - Day – that will indicate the day of observation. This attributes can contains the following values :
    - 0 = for day in elaboration
    - 1 = next day
    - -1 = previous day
  - "ReferenceProperty" – will specify the name of the property to read
- "Formula" when will be necessary to call a method to calculate the property's value
- "Parameter" , this attribute can be added when the Type="Formula" and can have two values:
  - "V" – vector
  - "M" – matrix
- Inside the "FormulaValue" tag, when the Type = "Formula" is necessary to specify the following tags :
- *Formula* – (mandatory) name of the formula, that must correspond to a method implemented into "Aggregation.Formulas.R" file
- Parameter – (not mandatory) allow to specify an acronym that will pass to the method some additional parameters. The following acronyms are defined for the moment:
  - Current.Date – day in elaboration
  - Inter.Data – interpolated data
  - Day.Agg – all the properties aggregated to day in elaboration
  - Expected.Values – number of expected values for the MEAN method
- For each formula, sometime it's possible to specify the time intervals. That can be realized in two ways:
  - Using one or more "Specific" tags like: <Specific Time="00, 03, 06, 09, 12, 15, 18, 21" Day="00" />, where "Time" contains the times of hourly observations, "Day" contains the day of the observations (relative value). If you want to get values from more days need to add a "Specific" tag for each day.
  - Using a combination of tags "Start" and "End" like : <Start Time="03" Day="00"/> <End Time="06" Day="01" />, where "Start" tag will define the start of the interval and "End" tag will define the end of the interval. With this type of tags the application will retrieve all the observations between 03:00 of day in elaboration until 06:00 of next day (considering the sample).
- When the "Parameter" tag is "V" need to specify the "ReferenceProperty" tag with the name of property to extract and to pass in input to the method, meanwhile the value is "M" the matrix with all properties are passed to the method
- Independent if the "Parameter" tag is "V" or "M" the temporal tags (Specific, Start, End) will be applied to extract only the values for specific times observations and to pass to the methods only observations for specific times.
- Another tag that can be present inside the "Property" tag is "NAFormula" (is not mandatory). This one has the scope to define an alternative formula in case the formula specified by "FormulaValue" return NA.

## Samples:

1. <Property>

```
<Name>TT_09</Name>  
<FormulaValue Type="Fixed" Time="09" Day="00" ReferenceProperty="TT" />  
</Property>,
```

, has the following explanation: the "Aggregation" module will create a property with name "TT\_09" for which value will be retrieved from the property TT for the day in elaboration at time 09:00 (inside the input file O.WeakChecks.<day>.dat)

2. <Property>

```
<Name>N</Name>  
<FormulaValue Type="Formula" Parameter="V" ReferenceProperty="N">  
  <Specific Time="00, 03, 06, 09, 12, 15, 18, 21" Day="00" />  
  <Formula>Aggregation.Mean.80</Formula>  
  <Parameter>Expected.Values</Parameter>  
</FormulaValue>  
</Property>
```

, has the following explanation: the "Aggregation" module will create a property with name "N" for which value will be calculate using the formula "Aggregation.Mean.80" that will receive in input:

An array with values of the property "N" in correspondence of the hourly observation 00, 03, 06, 09, 12, 15, 18, 21 of day in elaboration

At the same moment the method will receive in input number of expected values for the mean calculation (in that case the Expcted.Values = 9)

3. <Property>

```
<Name>RH06</Name>  
<FormulaValue Type="Fixed" Time="06" Day="0" ReferenceProperty="RH" />  
<NAFormula>  
  <FormulaValue Type="Fixed" Time="06" Day="0" ReferenceProperty="D_RH"  
</NAFormula>  
</Property>
```

, has the following explanation: the "Aggregation" module will create a property with name "RH06" for which value will be retrieve from the property RH at time 06:00 of day in elaboration. If the returned value is NA, the application will try to retrieve the value of the property from the property D\_RH at time 06:00 of day in elaboration (NAFormula).

Inside each **Property** tag it's possible to define interpolation configuration, using the following structure:

```

<Interpolation>
    <ReferenceProperty>N</ReferenceProperty>
    <Start Time="00" Day="00" />
    <End Time="06" Day="01" />
    <AllowNegative>YES</AllowNegative>
</Interpolation>

```

, where:

- The **ReferenceProperty** represent the name of the property to be interpolated (property present into WeakCheks files)
- The START tag contains two attributes: Time = contains the hour, Day = contains relative value for the day (00 = current day, -01 = previous day, 01 = next day). The tag specify the start of the interval that will be interpolate for Reference Property
- The END tag contains two attributes: Time = contains the hour, Day = contains relative value for the day (00 = current day, -01 = previous day, 01 = next day). The tag specify the end of the interval that will be interpolate for Reference Property
- The **AllowNegative** tag allow specifying when the negative interpolated values are accepted or not. If the tag is missing his value is equivalent to FALSE (that means the negative interpolated values will not be accepted – in that case, the negative values will be reset to 0).

The Aggregation.XML is different between various ROI: there are a version for Europe and a different version for CHINA.

### **3.3.1.1 Aggregated properties for ROI EUROPE**

The properties that Aggregation step produce for the ROI – Europe are show in the table below:

| <b>Property description</b>     | <b>Acronim</b> | <b>Note</b>  | <b>Unit of Measure</b> |
|---------------------------------|----------------|--|------------------------|
| <b>Station number</b>           | Station        | Numerical value  |                        |
| <b>Day Time</b>                 | DayTime        | Numerical value (YYYYMMDD)   |                        |
| <b>Total cloud cover</b>        | N              | Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC next day | Octas                  |
| <b>Low cloud cover</b>          | L              | Daytime mean using observation from 06, 09, 12, 15, 18 UTC   | Octas                  |
| <b>Cloud cover daytime mean</b> | NDT            | Daytime mean using observation from 06, 09, 12, 15, 18 UTC   | Octas                  |

|  |          |   |                    |
|--|----------|---|--------------------|
| <b>Measured Sunshine</b>                 | MSUN     | See the formula   | Hours              |
| <b>Measured Radiation</b>                | MRAD     | See the formula   | MJ/m <sup>2</sup>  |
| <b>Calculated Radiation</b>              | CRAD     | See the formula   | MJ/m <sup>2</sup>  |
| <b>Minimum Temperature</b>               | TN       | See the formula   | Celsius Degree     |
| <b>Maximum Temperature</b>               | TX       | See the formula   | Celsius Degree     |
| <b>Daily Mean Vapor pressure</b>         | MVP      | Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC                                    | hPa                |
| <b>Daily Mean Wind speed</b>             | FF       | Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC of next day | m/s                |
| <b>Precipitation</b>                     | RRR      | See the formula   | mm                 |
| <b>Air temperature</b>                   | TT<hour> | Air Temperature at 06, 09, 12, 15 and 18 of current day   | Celsius Degree     |
| <b>Relative Humidity</b>                 | RH<hour> | Relative Humidity at 06, 09, 12, 15 and 18 of current day   | %                  |
| <b>Daily Mean Vapor Pressure Deficit</b> | VPD      | Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC of next day | hPa                |
| <b>Daily Mean of Slope Saturation</b>    | SLOPE    | Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC of next day | hPa/Celsius Degree |
| <b>ANGOT Radiation</b>                   | ANGRAD   | See the formula   | KJ/m <sup>2</sup>  |
| <b>Angstrom Prescott Radiation</b>       | APRAD    | See the formula   | KJ/m <sup>2</sup>  |
| <b>Supit-Van Kappel Radiation</b>        | SVKRAD   | See the formula   | KJ/m <sup>2</sup>  |
| <b>Hargreaves Radiation</b>              | HGVRAD   | See the formula   | KJ/m <sup>2</sup>  |

|   |          |   |                |
|---|----------|---|----------------|
| <b>Penman potential evapotranspiration from a free water surface</b>      | E0       | See the formula                                       | mm/day         |
| <b>Penman potential evapotranspiration from a moist bare soil surface</b> | ES0      | See the formula                                       | mm/day         |
| <b>Penman potential evapotranspiration from a crop canopy</b>             | ET0      | See the formula                                       | mm/day         |
| <b>Snow depth</b>   | SNOW     | Instantaneous value at 06 UTC                         | cm             |
| <b>Visibility</b>   | VIS      | Formula using observation from 06, 09, 12, 15, 18 UTC | km             |
| <b>Absolut daily minimum temperature</b>                                  | Absolute | See the formula                                       | Celsius Degree |

### 3.3.1.2 Aggregated properties for ROI CHINA

The properties that Aggregation step produce for the ROI – CHINA are show in the table below:

| <b>Property description</b>      | <b>Acronim</b> | <b>Note</b>  | <b>Unit of Measure</b> |
|----------------------------------|----------------|--|------------------------|
| <b>Station number</b>            | Station        | Numerical value  |                        |
| <b>Day Time</b>                  | DayTime        | Numerical value (YYYYMMDD)   |                        |
| <b>Total cloud cover</b>         | N              | Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day | Octas                  |
| <b>Low cloud cover</b>           | L              | Not available  | Octas                  |
| <b>Cloud cover daytime mean</b>  | NDT            | Daytime mean using observation from 00, 03, 06, 09, 12 UTC   | Octas                  |
| <b>Measured Sunshine</b>         | MSUN           | Not available  | Hours                  |
| <b>Measured Radiation</b>        | MRAD           | Not available  | MJ/m <sup>2</sup>      |
| <b>Calculated Radiation</b>      | CRAD           | See the formula  | MJ/m <sup>2</sup>      |
| <b>Minimum Temperature</b>       | TN             | See the formula  | Celsius Degree         |
| <b>Maximum Temperature</b>       | TX             | See the formula  | Celsius Degree         |
| <b>Daily Mean Vapor pressure</b> | MVP            | Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day | hPa                    |

|   |          |  |                    |
|---|----------|--|--------------------|
| <b>Daily Mean Wind speed</b>  | FF       | Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day | m/s                |
| <b>Precipitation</b>  | RRR      | See the formula  | mm                 |
| <b>Air temperature</b>  | TT<hour> | Air Temperature at 00, 03, 06, 09, 12 of current day   | Celsius Degree     |
| <b>Relative Humidity</b>  | RH<hour> | Relative Humidity at 00, 03, 06, 09, 12 of current day   | %                  |
| <b>Daily Mean Vapor Pressure Deficit</b>                                  | VPD      | Daily mean using from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day             | hPa                |
| <b>Daily Mean of Slope Saturation</b>                                     | SLOPE    | Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day | hPa/Celsius Degree |
| <b>ANGOT Radiation</b>  | ANGRAD   | See the formula  | KJ/m <sup>2</sup>  |
| <b>Angstrom Prescott Radiation</b>  | APRAD    | See the formula  | KJ/m <sup>2</sup>  |
| <b>Supit-Van Kappel Radiation</b>   | SVKRAD   | See the formula  | KJ/m <sup>2</sup>  |
| <b>Hargreaves Radiation</b>   | HGVRAD   | See the formula  | KJ/m <sup>2</sup>  |
| <b>Penman potential evapotranspiration from a free water surface</b>      | E0       | See the formula  | mm/day             |
| <b>Penman potential evapotranspiration from a moist bare soil surface</b> | ES0      | See the formula  | mm/day             |
| <b>Penman potential evapotranspiration from a crop canopy</b>             | ET0      | See the formula  | mm/day             |
| <b>Snow depth</b>   | SNOW     | Instantaneous value at 00 UTC  | cm                 |
| <b>Visibility</b>   | VIS      | Formula using observation from 00, 03, 06, 09, 12  | km                 |

|  |     |                 |                |
|--|-----|-----------------|----------------|
| <b>Absolut daily minimum temperature</b> | TND | See the formula | Celsius Degree |
|--|-----|-----------------|----------------|

### 3.3.2 Aggregation formulas

For each property produced by Aggregation step the application are calling a formula.

The formulas are apply for each station, but before to apply those aggregation formulas for some hourly properties are applied an Interpolation temporal rule:

- Air temperature (TT)
- Dew point temperature (TD)
- Maximum 1h temperature (TX1)
- Minimum 1h temperature (TN1)
- Relative humidity (RH)
- Calculated relative humidity (D\_RH)
- Wind speed (FF)
- Slope (D\_SLOPE)
- Cloud cover (N)
- Low cloud cover (L)
- Vapor pressure deficit (D\_VPD)
- Vapor pressure (D\_E)
- Visibility (VIS)

To apply the interpolation is necessary to have at least 80% of valid values in the interval configured into the Aggregation.XML file.

#### 3.3.2.1 Aggregation rules for ROI – EUROPE

We would describe below some logics to calculate aggregated values for the ROI - Europe:

- **Total cloud cover** = mean of hourly cloud cover observation where is valued, using the values from the observations (property N) : 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC – current day, 00 UTC – next day
- **Cloud Cover Daytime mean** = mean of cloud cover values for specific times, using the values from the observations (property N) : 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC – current day
- **Low Cloud Cover Daytime mean** = mean of low cloud cover values for specific times, using the values from the observations (property L) : 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC – current day
- **Measured sunshine:**
  - The system will search, first, a Sunshine 24 hour observation between observation of 06:00 of next day and sunset time of current day. If this observation exists, will be use like Measured Sunshine.
  - If previous step does not return a valid value, the system try to calculate the measured sunshine like follow:

- Retrieve the sunrise and sunset time for current day. If not available, consider the interval 06 – 21 for the current day.
- If all the hourly observations related to the interval determined the previous point are presents the measured sunshine are calculate like the sum of hourly sunshine observation for that interval.
- Like last step, if the previous two are not returning a valid value and if all hourly radiation observations for the interval determined in the previous step, are presents, will consider 1 hour of sunshine for each hourly radiation observation is great or equal than 120 W/mq.
- **Measured radiation:**
  - The system will search, first, a Radiation 24 hour observation between the observation of 00:00 of next day and sunset time of current day. If this observation exists, will be use like Measured Radiation.
  - If previous step does not return a valid value the following algorithm will be applied:
  - Retrieve the sunrise and sunset time for current day. If not available, consider the interval 06 – 21 for the current day.
  - If all the hourly observations related to the interval determined the previous point are present the measured radiation are calculate like the sum of hourly radiation observation for that interval.
- **Calculated radiation:**
  - If the measured radiation is correctly calculated his are used to return the calculated radiation
  - Second alternative, if the measured sunshine is valued, the value of Angstrom Radiation is calculate and return
  - Third alternative, if cloud cover, maximum and minimum temperature are valid values then the Supit radiation are compute and return
  - Forth alternative, when clouder cover is not available but minimum and maximum temperature are valid values than the Hargreaves radiation are compute and return
- **Minimum temperature:**
  - Check if the TN12 property is present at 06 next day and that value
  - If previous step does not return a valid value, check if the TN6 property is present at 06, 12, 18 ,24 current day and 06 next day, and get the minimum value between the these observations
  - If previous step does not return a valid value, in case of hourly observations, get the minimum from TN1 values for interval 18 previous day and 06 current day
  - If previous step does not return a valid value the application will calculate the minimum between TT values for interval 18 previous day – 06 current day
  - If exist only one air temperature in the interval 18 previous day – 06 current day that is less than the calculated daily minimum temperature and the same air temperature is higher than minimum temperature from MOS file at 06 UTC then the daily minimum temperature will be replaced by minimum temperature from MOS file at 06 UTC
- **Maximum temperature:**
  - Check if the TX12 property is present at 18 current day
  - Check if the TX12 property is present at 15 current day if the previous step fail

- If previous step does not return a valid value, check if the TX6 property is present at 12, 18 and 24 current day and 06 next day and get the maximum value between the four observations
- If previous step does not return a valid value, in case of hourly observations, get the maximum from TX1 values for interval 00-23 of current day
- If previous step does not return a valid value the application will calculate the maximum between TT values for interval 06 – 18 current day
- If exist only one air temperature in the interval 06 current day – 18 current day that is higher than the calculated daily maximum temperature and the same air temperature is lower than maximum temperature from MOS file at 18 UTC then the daily maximum temperature will be replaced by maximum temperature from MOS file at 18 UTC
- **Daily mean vapor pressure** = mean of hourly valid vapor pressure values (D\_E) using observations from 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC –current day, 00 UTC next day
- **Daily mean wind speed** = mean of hourly valid wind speed values(FF) using observations from 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC – current day,00 UTC next day
- **Precipitation:**
  - First step are searching the 24 hours precipitation (PR24) at 06:00 of next. If exists will use like precipitation value
  - The 2<sup>nd</sup> possibility supposed to sum the 6-hourly precipitation values (PR06) from the interval 06 UTC – 06 UTC next day, only if all values are valid
  - The 3<sup>rd</sup> possibility, is to sum the hourly observation precipitation values (PREC) if all 25 observations (considering the interval 06:00 current day– 06:00 next day) are valid.
  - Another possibility is to search for PR24 value at 00 UTC next day. If exists and is valid that will be use like precipitation value
  - Next step suppose to sum 6-hourly observation precipitation values (PR06) from the interval 00 UTC – 00 UTC next day, only if all values are valid
  - Next step suppose to sum hourly observation precipitation values (PREC) from the interval 00 UTC – 00 UTC next day if all 25 observations are valid.
  - If none of the previous step return a valid value, the algorithm will search the first hourly observation for which the TR & RR data are values, into the interval 06:00 of next day and 06:00 of current day. In base of time of the observation where will find the first valid values for TR & RR will decide the length of interval of observation to use for precipitation value, like follow:
  - If the TR & RR was found for 06:00 or 03:00 of next day the interval time is 24 hours
  - If the TR & RR was found for 24:00 or 21:00 of current day the interval time is 18 hours
  - If the TR & RR was found for 18:00 or 12:00 of current day the interval time is 12 hours
  - Otherwise the algorithm cannot be apply
  - Considering that, we found a valid interval value will start to sum the precipitation values (RR) starting from the observation find for a length = interval time. The algorithm will go back until the first observation time of the interval time (find at the previous step) using TR duration value (in hours)

like step. If an observation are not present or the values (TR, RR) are not valid will reduce the current TR to half, but not less than 3 hours.

- **Air Temperature:** are expose the air temperature for the following time observations: 06:00, 09:00, 12:00, 15:00, 18:00 of current date
- **Relative Humidity:** are expose the relative humidity for the following time observations: 06:00, 09:00, 12:00, 15:00, 18:00 of current date. In case the reported values of relative humidity are missing or is not valid (RH) the system will retrieve the calculated relative humidity (D\_RH) for the same time observation
- **Daily mean vapor pressure deficit** = mean of hourly valid vapor pressure deficit values (D\_VPD) for the interval 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC – current day, 00 UTC – next day
- **Daily mean of slope saturation** = mean of hourly valid slope saturation values (D\_SLOPE) for the interval 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC – current day, 00 UTC – next day
- Potential evapotranspiration:
  - Penman potential evapotranspiration from a free water surface (E0)
  - Penman potential evapotranspiration from a moist bare soil surface (ES0)
  - Penman potential evapotranspiration from a crop canopy (ET0)
- **Snow depth** = Instantaneous value of SNOW at 06 current day
- **Visibility** = daytime of visibility hourly observations using the interval 06:00 – 18:00 of current date.
- **Absolute Daily Minimum Temperature:**
  - First step consider to get the minimum of TN12 between the observation of 12 UTC current day and 00 UTC next day
  - The 2<sup>nd</sup> step, consider to get the minimum of TN6 between the observation of 06, 12, 18 current day and 00 UTC next day
  - The 3<sup>rd</sup> step, consider to get the minimum value of TN1 between 01 current day and 00 next day for hourly observation
  - The 4<sup>th</sup> step, consider to get the minimum value of TT between 01 current day and 01 next day for hourly observation

For the mean values, the mean are calculate only if the application found at least 80% of valid values in relation to the number of expected values.

### **3.3.2.2 Aggregation rules for ROI – CHINA**

We would describe below some logics to calculate aggregated values for the ROI - Europe:

- **Total cloud cover** = mean of hourly cloud cover observation where is valued, using the values from the observations (property N) : 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day
- **Cloud Cover Daytime mean** = mean of cloud cover values for specific times, using the values from the observations (property N) : 00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC – current day
- Low Cloud Cover Daytime mean:
  - Not calculated
  - Measured sunshine:
    - Not calculated

- Measured radiation :
- Not calculated
- Calculated radiation:
- If the measured radiation is correctly calculated his are used to return the calculated radiation
- Second alternative, if the measured sunshine is valued, the value of Angstrom Radiation is calculate and return
- Third alternative, if cloud cover, maximum and minimum temperature are valid values then the Supit radiation are compute and return
- Forth alternative, when Cloud Cover is not available but minimum and maximum temperature are valid values than the Hargreaves radiation are compute and return
- Minimum temperature:
- Check if the TN12 property is present at 06 current day and that value
- If previous step does not return a valid value, check if the TN6 property is present at 06, 12, 18 previous day and 06, 12 current day, and get the minimum value between these observations
- If previous step does not return a valid value, in case of hourly observations, get the minimum from TN1 values for interval 06 previous day and 06 current day
- If previous step does not return a valid value the application will calculate the minimum between TT values for interval 06 previous day – 06 current day
- Maximum temperature:
- Check if the TX12 property is present at 18 current day and get that value
- If previous step does not return a valid value, check if the TX6 property is present at 18 previous day and 00,06,12, 18 current day and get the maximum value between the four observations
- If previous step does not return a valid value, in case of hourly observations, get the maximum from TX1 values for interval 18 of previous day and 18 current day
- If previous step does not return a valid value the application will calculate the maximum between TT values for interval 18 previous day - 18 current day
- **Daily mean vapor pressure** = mean of hourly valid vapor pressure values (D\_E) using observations from 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day
- **Daily mean wind speed** = mean of hourly valid wind speed values(FF) using observations from 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day
- Precipitation:
- Search the 24 hours precipitation (PR24) at 00:00 of next day. If exists will use like precipitation value, otherwise will check for PR24 at 21 current day and get that value if valid
- **Air Temperature:** are expose the air temperature for the following time observations: 00:00, 03:00, 06:00, 09:00, 12:00 of current date
- **Relative Humidity:** are expose the relative humidity for the following time observations: 00:00, 03:00, 06:00, 09:00, 12:00 of current date. In case the reported values of relative humidity are missing or is not valid (RH) the system will retrieve the calculated relative humidity (D\_RH) for the same time observation
- **Daily mean vapor pressure deficit** = mean of hourly valid vapor pressure deficit values (D\_VPD) for the interval 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day

- **Daily mean of slope saturation** = mean of hourly valid slope saturation values (D\_SLOPE) for the interval 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day
- Potential evapotranspiration:
- Penman potential evapotranspiration from a free water surface (E0)
- Penman potential evapotranspiration from a moist bare soil surface (ES0)
- Penman potential evapotranspiration from a crop canopy (ET0)
- **Snow depth** = Instantaneous value of SNOW at 00 current day
- **Visibility** = daytime mean of visibility hourly observations using the interval 00:00 – 12:00 of current date.
- Absolute Daily Minimum Temperature: TBD

For the mean values, the mean are calculate only if the application found at least 80% of valid values in relation to the number of expected values.

### 3.4 Heavy checks

Using aggregation results, more complex checks will be execute inside this step. At the same moment, it is possible to use a MOS file to fill the gap for missing observations.

All the errors to manage are configure into Messages.HeavyChecks.xml. Each message has a gravity level that will determine how the property value will be send to the output. Two possible level error can be configure into the XML file:

- W (wrong value)
- S (suspicious value)

The “HeavyChecks” step will produce three files:

- One (XML file) with the station observation for which exists at least on error message (with W or S gravity level)
- One (DAT file) with all station observation, transformed like follow:
- If one property has at least one wrong alert than the property will have the value NA and status = “W” (Wrong)
- If the property has only suspicious alerts than the property will maintain the value and status = “S” (Suspicious)
- If property has NA and do not have alerts than a possible value are searched inside the model file, if was specified in the command line. The model value are search for station and date. If exists, will replace the current property value. On the XML file this property will have the status = “R” (replaced).
- The same DAT file are copied to the history path to be used for the future checks( name of the file changed to D.<YYYYMMDD>.hist

When the HeavyChecks module are using Model Data and if some property was not valued from the previous step, that value of that property with be replaced with correspondent value from Model Data.

In detail, the following heavy checks need to be implement:

- Cloud check

- Cloud value outside of interval [0:8] Octas. In this case, the value will be consider WRONG.
  - Cloud cover is 0 and precipitation > 0. In this case, the value will be consider SUSPICIOUS.
- Cloud cover = 8 and sunshine duration > 0.5 \* day sunlight duration. In this case, the value will be consider SUSPICIOUS.
  - Minimum temperature check
  - The minimum temperature outside of the interval [-35:35] Celsius Degrees. In this case, the value will be consider WRONG.
- Maximum temperature checks
  - The maximum temperature outside of the interval [-20:50] Celsius Degrees. In this case, the value will be consider WRONG.
  - The difference between maximum and minimum temperature outside of interval [0:30] Celsius Degrees. In this case the maximum temperature will be consider wrong
- Snow check
  - If the mean temperature are great than 5 Celsius Degree the SNOW value will be consider WRONG
- Measured Sunshine check
  - If the Measured Sunshine are outside of interval [0:24] hours his value will be consider WRONG
  - Measured Sunshine > 1 and Measured Radiation is 0. In this case, the value will be consider SUSPICIOUS.
- Measured Radiation
  - If the Measured Radiation are outside of interval (0:36] MJ/m<sup>2</sup> his value will be consider WRONG. Even the 0 like value will be considered wrong
- Daily Mean Vapor Pressure
  - If Daily Mean Vapor Pressure are outside of interval [0:38] hPa his value will be consider WRONG
- Wind speed check
  - If Wind speed are outside of interval [0:25] m/s his value will be consider WRONG
- Precipitation check
  - If Precipitation are outside of interval [0:1400] mm his value will be consider WRONG
- Air Temperature Checks
  - All Air Temperature values need to be in the interval [-50:50] Celsius Degrees, otherwise will be consider WRONG
- Relative Humidity Checks
  - All Relative Humidity values need to be in the interval [0:100]%, otherwise will be consider WRONG
- Vapor pressure deficit Check
  - If Vapor pressure deficit is outside of interval [0:60] hPa than his value will be consider WRONG
- SLOPE check
  - If SLOPE value are outside of the interval [0:30] hPa/Celsius Degree will be consider WRONG
- Penman Evaporation check

- If Penman Evaporation value are outside of interval [0:25] mm/day than his value will be consider WRONG

If the HeavyChecks module generate a KO file it's possible to fix the reported error and run again, sending into the input of HeavyChecks module the KO file with fixed error instead of output from Aggregation module. In the case of the KO file sent into the input is empty (contains only the tag "<Observations />") the module will end the run without execute any checks.

### 3.5 Threshold checks

Using the output of "HeavyChecks" step, more complex checks will be execute inside this step using the "threshold" files. Those files are create once by year considering meteorological data from past years using the module TGenerator.

The errors to manage are configure into two files:

- Messages.Daily.ThresholdChecks.xml (file with daily threshold values).
- Messages.Seasons.ThresholdChecks.xml(file with seasons threshold values)

All files have same structure for the messages. Each message has a gravity level that will determine how the property value will be send to output. Two possible level error can be configure inside the XML file:

- W (wrong value)
- S (suspicious value)

Always in input, the "ThresholdChecks" module are using two data files that contains data for daily and seasons observations:

- One file (Threshold.Daily.dat) with threshold daily data for each station
- The second file (Threshold.Seasons.dat) with seasons data for each station

Another input of the module will consist on history files, produced by the "HeavyChecks" or Aggregation modules (name D.<YYYYMMDD>.hist). Only the file relative to the previous day of elaboration are use now.

The "ThresholdChecks" step will produce two files:

- One (XML format) file with the stations observation for which exists at least one error message (with W or S gravity level)
- One (DAT format) file with all stations observation, transformed like follow:
- If one property has at least one wrong alert than the property will have the value NA
- If the property has only suspicious alerts than the property will maintain the value

Actually, the following checks will be implement inside "Threshold Checks" step:

- Checks on daily thresholds
- Daily mean temperature greater than the 99th percentile value associated with the day under analysis (MT99 column from daily threshold file).
- Daily maximum temperature greater than the 99th percentile value associated with the day under analysis (MTX99 column from daily threshold file)

- Daily minimum temperature greater than the 99th percentile value associated with the day under analysis (MTN99 column from daily threshold file).
- Daily mean temperature lower than the 1st percentile value associated with the day under analysis (MT1 column from daily threshold file)
- Daily maximum temperature lower than the 1st percentile value associated with the day under analysis (MTX1 column from daily threshold file).
- Daily minimum temperature lower than the 1st percentile value associated with the day under analysis (MTN1 column from daily threshold file).
- Daily precipitation greater than the 5-year return levels of the season (winter, spring, summer and autumn) under analysis (RRR5Y column from season threshold file).
- Daily wind intensity greater than the 5-year return levels of the season (winter, spring, summer and autumn) under analysis (FF5Y column from season threshold file).
- Comparison of daily mean temperature predicted by an autoregressive of order 1, AR(1), model having mean daily temperatures from the 5 closest stations as additional covariates. **[NOT AVAILABLE YET]**
- Checks on season thresholds
- Difference in the mean temperature of two consecutive days greater than the 99<sup>th</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MT95 column from season threshold file).
- Difference in the maximum temperature of two consecutive days greater than the 99<sup>th</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MTX95 column from season threshold file).
- Difference in the minimum temperature of two consecutive days greater than the 99<sup>th</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MTN95 column from season threshold file).
- Difference in the mean temperature of two consecutive days lower than the 1<sup>st</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MT5 column from season threshold file).
- Difference in the maximum temperature of two consecutive days lower than the 1<sup>st</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MTX5 column from season threshold file).
- Difference in the minimum temperature of two consecutive days lower than the 1<sup>st</sup> percentile value associated with the season (winter, spring, summer and autumn) under analysis (MTN5 column from season threshold file).
- Daily wind speed value great than the wind speed threshold of the 10-years return levels of the season (RRR5Y column name). In this case a WRONG alert will be generate
- Precipitation daily value great than precipitation threshold of the 10-years return levels the season (FF5Y column name). In this case a WRONG alert will be generate

If the ThresholdChecks module generate a KO file it's possible to fix the reported error and run again, sending into the input of ThresholdChecks module the KO file with fixed error instead of output from HeavyChecks module. In the case of the KO file sent into the input is empty (contains only the tag "<Observations />") the module will end the run without execute any checks.

### 3.5.1 Threshold file generator

The generator of threshold files (daily and seasons) is a module that are not running inside the Quackme chain. It run once by year and will generate two files:

- Daily threshold file: that contains for each station and for each day, the following properties:

| Property description  | Column  | Note                                 | Unit of Measure |
|---|---------|--------------------------------------|-----------------|
| <b>Station number</b>   | Station | Numerical value                      |                 |
| <b>Day</b>  | Day     | Julien day value for referenced year |                 |
| <b>Mean temperature 99 percentile</b>   | MT99    | Numerical value                      | Celsius Degree  |
| <b>Mean temperature 1 percentile</b>  | MT1     | Numerical value                      | Celsius Degree  |
| <b>Maximum temperature 99 percentile</b>  | MTX99   | Numerical value                      | Celsius Degree  |
| <b>Maximum Temperature 1 percentile</b>   | MTX1    | Numerical value                      | Celsius Degree  |
| <b>Minimum Temperature 99 percentile</b>  | MTN99   | Numerical value                      | Celsius Degree  |
| <b>Minimum Temperature 1 percentile</b>   | MTN1    | Numerical value                      | Celsius Degree  |
| <b>Mean temperature prediction indicators using the mean temperature of 5 closest station [NOT AVAILABLE YET]</b> |         |                                      |                 |
| <b>1<sup>st</sup> indicator</b>   | MTC5_1  | Numerical value                      | Celsius Degree  |
| <b>2<sup>nd</sup> indicator</b>   | MTC5_2  | Numerical value                      | Celsius Degree  |
| <b>3<sup>rd</sup> indicator</b>   | MTC5_3  | Numerical value                      | Celsius Degree  |
| <b>4<sup>th</sup> indicator</b>   | MTC5_4  | Numerical value                      | Celsius Degree  |
| <b>5<sup>th</sup> indicator</b>   | MTC5_5  | Numerical value                      | Celsius Degree  |
| <b>6<sup>th</sup> indicator</b>   | MTC5_6  | Numerical value                      | Celsius Degree  |

The values for the temperature columns are calculate using the R formula **loess** with the data interval: 01.01.1985 – 31.12.2018. The daily threshold file will contains, for each station, the mean temperatures values for each day of the analyzed year.

On the other side, the mean temperature prediction indicators are calculate using both "**quantile**" and "**arima**" functions of R, for the same interval.

- Seasons threshold file: that contains for each station and for each season (winter, spring, summer, autumn), the following properties:

| Description           | Column  | Note   | Unit of Measure |
|-----------------------|---------|--|-----------------|
| <b>Station number</b> | Station | Numerical value  |                 |
| <b>Season</b>         | Season  | Text value with possible values: winter / spring / summer / autumn |                 |

|  |       |                 |                |
|--|-------|-----------------|----------------|
| <b>Mean temperature 99 percentile</b>                    | MT95  | Numerical value | Celsius Degree |
| <b>Mean temperature 1 percentile</b>                     | MT5   | Numerical value | Celsius Degree |
| <b>Maximum temperature 99 percentile</b>                 | MTX95 | Numerical value | Celsius Degree |
| <b>Maximum Temperature 1 percentile</b>                  | MTX5  | Numerical value | Celsius Degree |
| <b>Minimum Temperature 99 percentile</b>                 | MTN95 | Numerical value | Celsius Degree |
| <b>Minimum Temperature 1 percentile</b>                  | MTN5  | Numerical value | Celsius Degree |
| <b>Precipitation 10-year return levels of the season</b> | RRR5Y | Numerical value | mm             |
| <b>Wind speed 10-year return levels of the season</b>    | FF5Y  | Numerical value | m/s            |

The values for the temperature columns are calculate using the R formula "**quantile**" with the data interval: 01.01.1985 – 31.12.2018 for each station and for each season, while the precipitation and wind intensity are calculate using the R formula "**fpot**" with the same interval for each station and each season.

The seasons are covering the following intervals:

- Winter : 01.12 -> 28/29.02
- Spring : 01.03 -> 31.05
- Summer : 01.06 -> 31.08
- Autumn : 01.09 -> 30.11

For each ROI (Europe, China) will exists a set of threshold files (daily, seasons) to be use on the specific workflow chain.

For the daily threshold file, the following logics are used to compute the 99 percentile and 1<sup>st</sup> percentile values:

- Estimate the daily 99th percentile with a 15-day moving window approach centered on the day of interest. Then, we derive the minimum of the set of functions  $\geq$  the one describing the 99th daily percentiles, and finally we apply a smoothing spline with natural basis. Boundaries are extended by adding the first/last 15 days of the year to avoid spurious effects in the smoothing step.
- Estimate the daily 1<sup>st</sup> percentile with a 15-day moving window approach centered on the day of interest. Then, we derive the maximum of the set of functions  $\geq$  the one describing the 1<sup>st</sup> daily percentiles, and finally we apply a smoothing spline with natural basis. Boundaries are extended by adding the first/last 15 days of the year to avoid spurious effects in the smoothing step.

### 3.6 SConverter module

The main scope of this module is to convert the aggregation results (after heavy checks and eventually threshold checks) into the S-File format, to be send to ALTERRA.

On the next table are describe the structure of the S-File:

| Property description  | Length | Note   | Unit of Measure | Digits |
|-----------------------|--------|--|-----------------|--------|
| <b>Station_Number</b> | 01:07  | Station number                                       |                 | N.A.   |
| <b>Day</b>            | 08:17  | Day number of the year (Julien format)               |                 | N.A.   |
| <b>CLOUD_24_TOTAL</b> | 18:23  | Daily mean of total cloud cover (mean over 24 hours) | Octas           | 1      |
| <b>SUNSHINE</b>       | 24:29  | Sunshine duration (sum of 24 hours)                  | Hours           | 1      |
| <b>RAD_MEA</b>        | 30:35  | Measure global radiation (sum of 24 hours)           | MJ/m2           | 1      |
| <b>TEMP_MIN</b>       | 36:41  | Minimum temperature (last 12 hours in Europe)        | Celsius Degree  | 1      |
| <b>TEMP_MAX</b>       | 42:47  | Maximum temperature (last 12 hours in Europe)        | Celsius Degree  | 1      |
| <b>VAP_PRES</b>       | 48:53  | Daily mean vapour pressure (mean over 24 hours)      | hPa             | 1      |
| <b>WIND_10</b>        | 54:59  | Daily mean speed at 10 metters (mean over 24 hours)  | m/s             | 1      |
| <b>RAIN</b>           | 60:65  | Amount of precipitation (sum over 24 hours)          | Mm/d            | 1      |
| <b>TEMP_06</b>        | 66:71  | Air temperature morning (06 UTC in Europe)           | Celsius Degree  | 1      |
| <b>HUM_06</b>         | 72:77  | Relative humidity morning (06 UTC in Europe)         | %               | 0      |
| <b>TEMP_09</b>        | 78:83  | Air temperature late morning (09 UTC in Europe)      | Celsius Degree  | 1      |
| <b>HUM_09</b>         | 84:89  | Relative humidity late morning (09 UTC in Europe)    | %               | 0      |
| <b>TEMP_12</b>        | 90:95  | Air temperature midday (12 UTC in Europe)            | Celsius Degree  | 1      |

|                            |         |  |                    |   |
|----------------------------|---------|--|--------------------|---|
| <b>HUM_12</b>              | 96:101  | Relative humidity midday (12 UTC in Europe)  | %                  | 0 |
| <b>TEMP_15</b>             | 102:107 | Air temperature afternoon (15 UTC in Europe)   | Celsius Degree     | 1 |
| <b>HUM_15</b>              | 108:113 | Relative humidity afternoon (18 UTC in Europe)   | %                  | 0 |
| <b>TEMP_18</b>             | 114:119 | Air temperature evening (18 UTC in Europe)   | Celsius Degree     | 1 |
| <b>HUM_18</b>              | 120:125 | Relative humidity evening (18 UTC in Europe)   | %                  | 0 |
| <b>STATE_SOIL</b>          | 126:131 | State of soil  |                    | 1 |
| <b>VAP_PRES_Def</b>        | 132:137 | Daily mean vapour pressure deficit (mean over 24 hours)  | hPa                | 1 |
| <b>SLOPE_VP_VS_T</b>       | 138:143 | daily mean slope saturation vapour pressure vs. temperature curve  | hPa/Celsius Degree | 3 |
| <b>CLOUD_DAYTIME_TOTAL</b> | 144:149 | daytime mean of total cloud cover (mean over daytime hours)  | Octas              | 1 |
| <b>CLOUD_DAYTIME_LOW</b>   | 150:155 | daytime mean amount of CL clouds or, if no CL clouds are present, the daytime mean amount of CM clouds (mean over daytime hours) | Octas              | 1 |
| <b>CLOUD_SHADOW</b>        | 156:161 | daytime mean amount of shadow clouds (mean over daytime hours)   | Octas              | 1 |
|                            | 162:167 | Calculated sunshine duration (NOT SAVED INTO DB)   | %                  | 1 |
|                            | 168:173 | Highest possible global radiation at clear sky Crad (NOT SAVED INTO DB)  | MJ/m2              | 1 |
|                            | 174:179 | Potential evapotranspiration ETP (NOT SAVED INTO DB)   | mm/day             | 1 |

|                   |         |   |    |      |
|-------------------|---------|---|----|------|
| <b>VISIBILITY</b> | 180:185 | daytime mean visibility (daytime mean)  | km | 1    |
| <b>SNOW_DEPTH</b> | 186:191 | Snow depth (instantaneous value 06 UTC in Europe)   | cm | 1    |
| <b>RAIN_MOS</b>   | 192:197 | indicator for reference period and usage of short range forecast in daily sum of precipitation RAIN |    | N.A. |

The **SConverter** module will ensure the conversion from JRC format to S-File format, for which the file name will respect the following convention name: *S<YYYYMMDD>.dat*, where YYYYMMDD represent the day of input file (the day of observations contained in the input file)

### 3.7 RRR file generator module

This module has the scope to generate the precipitation 3h/6h data files. Can run only after the WeakChecks module, because need to use the hourly history files.

Like input, the module are using:

- A. two hourly history files:
  - o H.<reference date>.hist
  - o H.<reference date - 1 day>.hist

, where both files are presents into CONFIG path (specified through parameter into the command line).

- B. MOS files:
  - a. M.MG.<reference date>.csv
  - b. M.MG.<reference date - 1 day>.csv

Like output, the module can produce 4 files for 6h mode or 8 files for 3h mode, starting with time 00 of the reference date (day for which module is running) for every 3/6 hours. A file of 3/6 hour precipitation has the following structure:

| Description                | Column  | Note  | Unit of Measure |
|----------------------------|---------|---|-----------------|
| <b>Station</b>             | Station | Numerical value   |                 |
| <b>Day time</b>            | DayTime | Day time of observation in the format YYYY-MM-HH H24:mm |                 |
| <b>Precipitation value</b> | RRR     | Numerical value   | mm              |
| <b>Flag</b>                | Flag    | Numerical value   |                 |

The flag can have the following values:

- 1=Changed by meteorologist (not applicable)
- 2=Automatically corrected (not applicable)
- 3=Observation
- 4=Linear interpolation from observations
- 5=Interpolated via MOS from observations
- 6=MOS analyses (not managed yet)

For each time interval (3 or 6 hours), for which need to generate the files, are applied a specific logic to compute the precipitation values:

- 3 hours interval:
  - o If for the specific time (09, 12, 15, 18, 21, 00 next day, 03 next day, 06 next day) are missing all the observation from the previous 3 hours (specific time - 2 hours) the date are ignored
  - o If the RR & TR are correctly valued and RR = 3 then use the value of RR.
  - o Next try, check if all the 3 observations (starting for the time elaboration) in the interval [elaboration time - 2 hours : elaboration time] are presents and if the hourly precipitation values are all 3 valid will report the sum of these

- values. In case at least of this values was interpolated, the flag reported will be 4, else 3.
- If from the 3 observations are missing only one and the two present observations contains valid values for 1H precipitation parameter will try to retrieve the missing value from the MOS. If present and valid than will report the sum of two valid observation and one MOS value. In that case the flag will be 5.
  - Try to calculate a 3H precipitation value id the RR/TR are valid and TR is 6 or 12:
    - In case of TR = 6 calculate the ratio of the precipitation using MOS 1H precipitation values like follow :
      - Ratio for first 3 hours interval = (sum of 1H precipitation of the first 3 hours in the interval / sum of 1H precipitation of the 6 hours interval) \* RR value
      - Ratio for second 3 hours interval = (sum of 1H precipitation of the second 3 hours in the interval / sum of 1H precipitation of the 6 hours interval) \* RR value
    - In case of TR = 12 will apply the same logics of the previous point but the RR Value will be the difference between current RR (where TR = 12) and the RR observation the correspond to 6 hours before current time if that one has TR = 6.
    - On both situation the flag will be 5
- 6 hours interval:
- If for the specific time (12, 18, 00 next daym 06 next day) doesn't exists the row of observation for a station, that station is ignored for that time and NA value will be reported
  - For the observation row corresponding to time in elaboration, check if the PR06 column contains a valid value and use it if valid
  - Next try, are related to the columns RR & TR. If the TR = 6, then use the value of RR if valid
  - Next try, check if all the 6 observations (starting for the time elaboration) in the interval [elaboration time – 5 hours : elaboration time] (for example [19:00 previous day – 00:00 current day] are presents:
    - Verify if the RR = 1 for all of them and if the TR values are valid for all, then the SUM(TR) for all these rows will be used like precipitation value

Else, if the PREC columns contains valid values for all 6 observations time, then the SUM(PREC) will be used like precipitation value. If one of this value was interpolated, then the flag of the reported value will be set to **4**.
  - The last step, in the case the previous step did not produce a valid value, suppose to check for 12-hour precipitation split. We can found two situation:
  - The TR = 12 for the current elaboration time, case when we will consider the MOS ratio of precipitation using MOS values (RR) for current elaboration time and current elaboration time - 6 hour. The reported value will be calculated like follow:

- The MOS ratio = the RR MOS value for the same time of observed data in elaboration / (sum of RR MOS values found for current elaboration time and current elaboration time - 6 hour).
- The reported value = RR of observed data \* MOS ratio
- The TR = NA for the current elaboration time. In that case we can use the logics of the MOS ratio of precipitation using MOS values (RR) for current elaboration time and current elaboration time + 6 hour, only if the observation corresponding to +6 hour relative to current time, present TR = 12 and a valid RR valid. The reported value will be calculated like follow:
- The MOS ratio = the RR MOS value for the same time of observed data in elaboration / (sum of RR MOS values found for current elaboration time and current elaboration time + 6 hour).
- The reported value = RR of observed data \* MOS ratio
- When the MOS data are involved, the reported flag is 5.

The name of the output files are respecting the following naming convention:

rrr\_<yyyyMMddHH>.txt – for 6H precipitation,

rrr3h\_<yyyyMMddHH>.txt – for £H precipitation,

, where:

- yyyy = the year of reference date (4 digits)
- MM = the month of reference date (2 digits)
- dd = the day of reference date (2 digits)
- HH = the time

These files are created on the OUTPUT path specified in the command line.

### 3.8 SG GUI

This step, which can be involved from the many points of the principal workflow; has the role to allow the specialists to correct the errors events intercepted during the data elaboration.

Both input and output are using an XML format.

The XML file will have the following structure:

```
<Observations>
  <Observation>
    <Station> station code </Station>
    <DayTime> day time with the format YYYYMMDDHH </DayTime>
    <TAG Status="status of the value for current property"> value </TAG>
    ....
    <TAG Status="status of the value for current property"> value </TAG>
    <ALERT Message="alert message"
      Type="type of the alert"
      Property="property name to which the alert refers"
      Status="proposed status for the property" />
    ....
    <ALERT Message="alert message"
      Type="type of the alert"
      Property="property name to which the alert refers"
      Status="proposed status for the property" />
  </Observation>
</Observations>
```

The significant of the XML tag are:

- TAG is an XML tag with the name of the property (for example: TT – air temperature, Snow – show depth)
- Status can have the following values :
- C – correct
- W – wrong
- S – suspicious
- R - replaced

Inside the ALERT , the LEVEL attribute provide the gravity of the alert: W – wrong value, S – suspicious value, R – replaced (this value are used only for the HeavyChecks output)

- Property – contains the name of property to which the alert message refers (must exists a tag with same name in the same Observation parent node)
- 

Sample of XML with errors:

```
<Observations>
  <Observation>
    <Station>15490</Station>
```

```

<DayTime>2017110206</DayTime>
<TTT Status="C">-4.46</TTT>
<TD Status="C">-3.46</TD>
<Prec Status="C">0</Prec>
<Snow Status="W">5</Snow>
<FF Status="C">1.7</FF>
<ALERT Level="W" Property="Snow" Message="a decrease of snow depth of more than
50cm is very unlikely under any circumstance"/>
</Observation>
<Observation>
  <Station>15491</Station>
  <DayTime>2017110206</DayTime>
  <TTT Status="C">-4.46</TTT>
  <TD Status="C">-3.46</TD>
  <Prec Status="C">0</Prec>
  <Snow Status="C">50</Snow>
  <FF Status="S">1</FF>
  <ALERT Level="S" Property="FF" Message="Suspicious value when difference relative to
previous observation are great than 25 knots"/>
</Observation>
<Observation>
  <Station>15491</Station>
  <DayTime>2017110212</DayTime>
  <TTT Status="C">0.45</TTT>
  <TD Status="C">0.42</TD>
  <Prec Status="C">0</Prec>
  <Snow Status="C">70</Snow>
  <FF Status="S">0</FF>
  <ALERT Level="S" Property="FF" Message="Suspicious value in correspondence with
previous and after observations"/>
  <ALERT Status="A" Property="Snow" Type="W" Message="Fresh snow must originate
from precipitation. Normally, 1mm precipitation results in 1cm snow. It is not likely that this
ratio is highly exceeded."/>
</Observation>
<Observation>
  <Station>15491</Station>
  <DayTime>2017110221</DayTime>
  <TTT Status="C">2.02</TTT>
  <TD Status="C">1.82</TD>
  <PREC Status="C">101</Prec>
  <SNOW Status="S">500</Snow>
  <FF Status="C">2</FF>
  <ALERT Level="S" Property="Snow" Message="the values are unusually high but yet
possible"/>
</Observation>
</Observations>

```

QUACKME engine can generate errors with two gravity levels:

- S (suspicious)
- W (wrong)

From the QUACKME GUI the operator can change the value of the reported property or can decide that the value reported like "suspicious" or "wrong" it is correct. After the operator correction, the status of the property can change in:

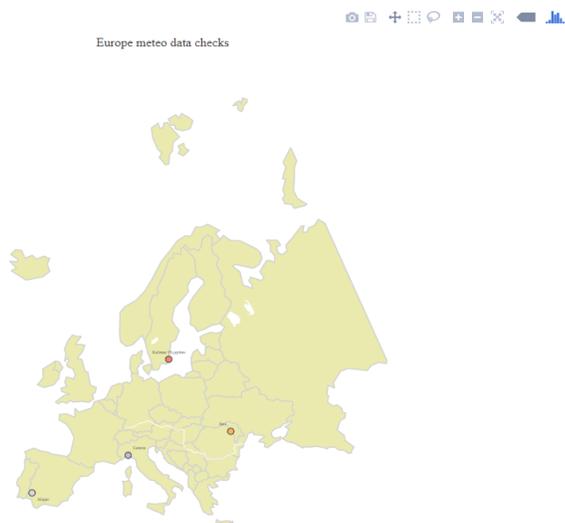
- **M** (modified) if the operator insert a new value different from the reported one
- **F** (forced) if the operator consider the reported value to be right
- **W** (wrong) the operator leaves the original status (W). In that case, on the next run, the property's value will be change to NA (NOT DEFINED)
- **S** (suspicious) the operator leaves the original status (S)

A possible design for the GUI can be view in the imagines bellow:

Showing 1 to 2 of 2 entries

| File name                          | Checks level | Reference date | Actions           |
|------------------------------------|--------------|----------------|-------------------|
| O.KO.WeakChecks.20171102.xml       | WeakChecks   | 20171102       | <a href="#">✎</a> |
| O.KO.WeakChecks.20171102Backup.xml | WeakChecks   | 20171102Backup | <a href="#">✎</a> |

Showing 1 to 2 of 2 entries



Kalmar Flygplats ( Sweden )

[✎ Edit](#)

Showing 1 to 10 of 22 entries

| Time  | Property | Status   | Value |
|-------|----------|----------|-------|
| 03:00 | TD       | Modified | 19    |
| 03:00 | DIR      | Modified | 238   |
| 03:00 | N        | Modified | 15    |
| 03:00 | RH       | Modified | 1     |
| 06:00 | FF       | Forced   | 1     |
| 06:00 | RH       | Forced   | 0     |
| 09:00 | TD       | Modified | 12    |
| 09:00 | FF       | Forced   | 20    |
| 09:00 | N        | Modified | 8     |
| 12:00 | TT       | Modified | 22    |

Showing 1 to 10 of 22 entries

Show 10 entries

Search:

| File name                          | Checks level | Reference date | Actions           |
|------------------------------------|--------------|----------------|-------------------|
| O.KO.WeakChecks.20171102.xml       | WeakChecks   | 20171102       | <a href="#">✎</a> |
| O.KO.WeakChecks.20171102Backup.xml | WeakChecks   | 20171102Backup | <a href="#">✎</a> |

Showing 1 to 2 of 2 entries

Previous **1** Next

Europe meteo data checks



Alajar ( Spain )

creato file: ./Lock/\_8380\_201803090811.Lock  
[Save](#) [Cancel](#)

Show 10 entries

Search:

| Time  | Property | Status     | Value                            | Forced                              |
|-------|----------|------------|----------------------------------|-------------------------------------|
| 03:00 | TD       | Modified   | <input type="text" value="11"/>  | <input type="checkbox"/>            |
| 03:00 | DIR      | Forced     | <input type="text" value="240"/> | <input checked="" type="checkbox"/> |
| 03:00 | RH       | Forced     | <input type="text" value="0"/>   | <input checked="" type="checkbox"/> |
| 03:00 | AP       | Suspicious | <input type="text" value="850"/> | <input type="checkbox"/>            |
| 06:00 | FF       | Suspicious | <input type="text" value="1"/>   | <input type="checkbox"/>            |
| 06:00 | RH       | Suspicious | <input type="text" value="0"/>   | <input type="checkbox"/>            |
| 06:00 | AP       | Suspicious | <input type="text" value="900"/> | <input type="checkbox"/>            |
| 09:00 | FF       | Suspicious | <input type="text" value="20"/>  | <input type="checkbox"/>            |
| 12:00 | TT       | Suspicious | <input type="text" value="37"/>  | <input type="checkbox"/>            |
| 12:00 | FF       | Suspicious | <input type="text" value="0"/>   | <input type="checkbox"/>            |

Showing 1 to 10 of 21 entries

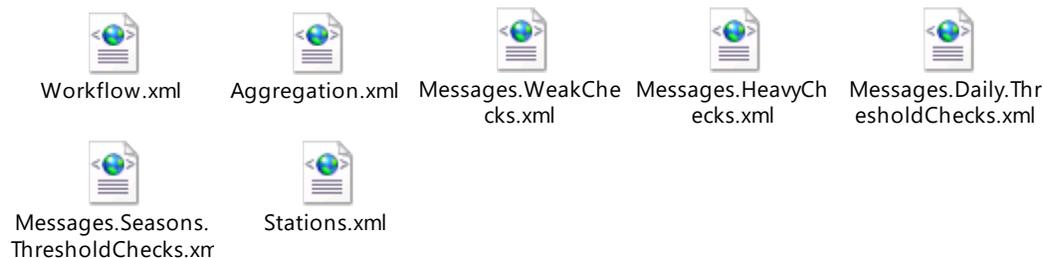
Previous **1** 2 3 Next

## 4. QUACKME Configurations

The QUACKME software will contains more configuration files, for each level of execution:

- A. Weak Checks
  - One file to configure the messages for current step (Messages.WeakChecks.xml)
  - One file to configure the routines to execute for each step of QUACKME (Workflow.xml)
  - One file that contains the stations data (Stations.xml)
- B. Aggregation
  - One file to configure the properties and associated formulas to generate aggregated properties (Aggregation.xml)
- C. Heavy Checks
  - One file to configure the messages for current step (Messages.HeavyChecks.xml)
  - Use of Aggregation configuration file (Aggregation.xml)
- D. Threshold Checks
  - Two files to configure the messages for each step of QUACKME:
    - i. Messages.Daily.ThresholdChecks.xml
    - ii. Messages.Seasons.ThresholdChecks.xml
  - Two files with threshold data
    - i. Threshold.Daily.YYYY.dat, which contains the daily threshold data relative to the past year , for every day
    - ii. Threshold.Seasons.YYYY.dat, which contains the season's threshold data relative to the past year.
  - Use of Aggregation or HeavyChecks output: file of day in exam.
  - Use of history file (produced on Aggregation step)

All configuration files will have XML format, less than threshold file that will have text format.



### 4.1 Messages structure

The "Messages.xml" file contains a list of MESSAGE tag that have the following structure:

```
<Message>  
  <Property>TD</Property>  
  <Code>006</Code>
```

<**Text**>Dew-point temperature exceeds air temperature by 1C or less: dew-point temperature automatically changed to air temperature</**Text**>

<**Level**>A</**Level**>

</**Message**>, where:

- "Property" tag contains the name of the property to which refers
- "Code" represents the code of the message (in fact a sequential number related to each property)
- "Text" contains the full description of the error
- "Level" contains the gravity level: A – alert, S – suspicious, W – wrong value. The scale of priority for gravity level are W, S, A.
- "Values" (only for Messages.HeavyChecks.xml, Messages.Daily.ThresholdChecks.xml, Messages.Seasons.ThresholdChecks.xml) will contains a list of property to expose (like <PropertyName>=<PropertyValue>) into the output KO XML file.

The text message can contains dynamic values like #0#, #1# ... #n# that must be transfer inside the source code. The #0# will be replace with the first value received, the #1# with the second one, and so on.

The KO XML file, if empty, contains only the tag "<Observations />".

## 4.2 Messages behavior

At the start of each check step the engine are making a merge between the original input data and events corrected from QUACKME GUI by the operator. The merge operation are using the following rules:

- If a property was reported and corrected by the operator with status **F** will be consider correct
- If a property was reported and corrected by the operator with status **M** the new values assigned by the operator will be move to the input data and consider correct
- If a property was reported with the status **W** (wrong) and left unchanged by the operator his value will be discarded and change to **NA** (NOT DEFINED)
- If a property was reported with the status **S** (suspicious) and left unchanged by the operator his value will remain the same for next run

For all runs after the first, exists the possibility to generate the same message for a specific property of an observation. In this situation, the application will apply two rules:

- If the property was corrected by the operator with status **F** (forced) the message will be ignore
- If the property was corrected by the operator with status **M** (modified) the message will be report again

## 4.3 Workflow configuration structure

The "Workflow.xml" is organize by levels: Weak Checks, Heavy Checks and so on. All level have the same structure:

<**Check**>

```
<Name>Dew Point</Name>
<Module>DewPoint_WeakChecks.R</Module>
<Method>WeakChecks.CheckDewPoint</Method>
<Name>ON/OFF</Name>
</Check>, where:
```

- **Name:** check description
- **Module:** name of the module that contains the checks
- **Method:** name of the method to call for the current check
- **Status:** ON – the check module need to be considered, OFF – the check module will be ignored

○

A parent tag specific for each level includes all the checks relative to that level:

- **WeakChecks** for the weak checks level
- **HeavyChecks** for the heavy checks level
- **ThresholdChecks** for the checks using threshold data