



JRC TECHNICAL REPORTS

Quality Checks for Meteorological Data (QUACKME)

Architectural & Technical Guide

JRC-Directorate Sustainable Resources

D5 – Food Security Unit, Ispra, Italy

This publication is an [updated version of the](#) Technical Report by the Joint Research Centre (JRC), the European Commission's science and knowledge service [of the same title published in 2021](#). 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, [2023](#)

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: [Elena Tarnavsky](#) (*D5 – Food Security Unit, Ispra, ITALY*), Marian Bratu (Fincons Spa), *Quality Checks for Meteorological Data (QUACKME) – Technical Guide*, EUR, [August 2023](#)

All images © European Union [2023](#)

Printed in Italy

Contents

Annotations table	4
Version summary	5
1. Requirements	7
2. Architecture	8
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	25
3.4 Heavy checks	30
3.4.1 Exceptions management.....	32
3.5 Threshold checks	33
3.5.1 Threshold file generator	35
3.5.2 Convex Hull Exceptions management	36
3.6 Quality checks flags management.....	37
3.7 SConverter module	38
3.8 RRR file generator module	41
3.9 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	A ctual M eteorological D Atabase C onstruction
QUACKME	Q UAlity C hec K s for M Eteorological Data
SG	S pecialists G roup
MOS	M odel O utput S tatistics
WMO	W orld M eteorological O rganization
ECMWF	E uropean C entre for M edium-Range W eather F orecasts
ROI	R egion O f I nterest

Version summary

Version	Description
1.2.0	More explanation in the Aggregation.XML configuration List all interpolated properties before aggregation Management of empty KO files inside modules
1.2.1	Added interpolation configuration inside Aggregation.XML Added new aggregated property: Absolute Daily Minimum Temperature 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) 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 Aggregation – Add Expected.Values parameter HeavyChecks – Measuread 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
1.4.8	Aggregation: fix the interval to sum the 1H precipitations
1.4.9	Fix some problems inside WeakCheks module on Precipitation & WindSpeed for MOS checks
1.4.10	Switch off the Sunshine (daily) calculation using hourly radiation if ≥ 120 W/mq.
1.5.0	New WeakCheck version, precipitation cross checks inside WeakCheck. Remove interpolation for 1H precipitation. Add exceptions management for HeavyChecks module (chapter 3.4.1), Convex Hull exceptions management inside ThresholdChecks (chapter 3.5.2), generate quality checks flag in all modules (chapter 3.6).
1.5.1	WeakChecks: Improve checks for consecutive hourly values (TT, TD), raise error on both parameters (TD, TT) on TD is higher than TT, change the threshold values for hourly precipitation. HeavyChecks: raise the error for both parameters in correspondence of the cross-checks N/RRR, TN/TX, change the thresholds per precipitation daily value.
1.5.2	HeavyChecks – change daily precipitation checks for suspicious value from 200 to 100 mm, New checks for TMIN > TMAX and visibility validity interval, WeakChecks – added new checks for precipitation in correspondence of TR = 1, 3, 6, 12 hours, Aggregation

	– recalculate TMIN and TMAX using 24h specific interval when TMIN > TMAX after the standard formulas, ThrehsoldChecks – use 2° C tolerance for temperature checks.
--	---

1. Requirements

The main purpose of JRC is the development and maintenance of a new Quality Checks for Meteorological Data Software (QUACKME) to process the meteorological observations from ground weather stations, remove the erroneous values, identify suspicious values, and fill the gaps whenever possible. QuackMe addresses the issues and the technical/scientific gaps of the AMDAC software (previously in place at JRC), providing at the same time a flexible and easy to configure tool.

QUACKME satisfies some key essential criteria:

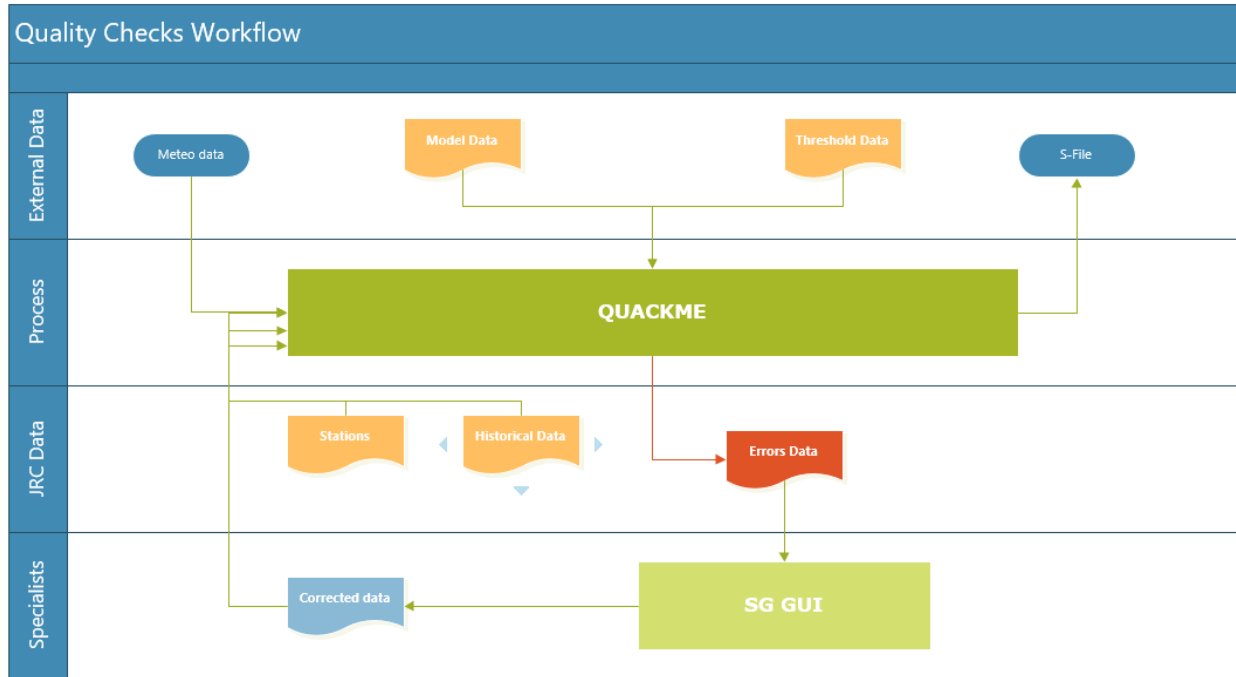
- open source;
- flexible to modify and easy to configure;
- clear and easy to modify code;
- able to use external data whenever necessary or required;
- easily adaptable to different format of input data;
- able to trace errors for fast and easy problem solving;
- adaptable to different platform;
- integrated with a user-friendly interface;
- able to report wrong or suspicious observations in a clear way.

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

Property description	Note	Unit of Measure
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		KJ/m ²
Sunshine duration		Minutes
Precipitation		mm
Relative humidity	Percentage between 0 and 100	
Atmospheric pressure		hPa
Cloud cover		Octas
Visibility		km
Snow depth		cm

2. Architecture

The architectural design of QuackMe.



Architecture details

The quality checks software part is implemented as a workflow where each step produces a part of the input for the next step.

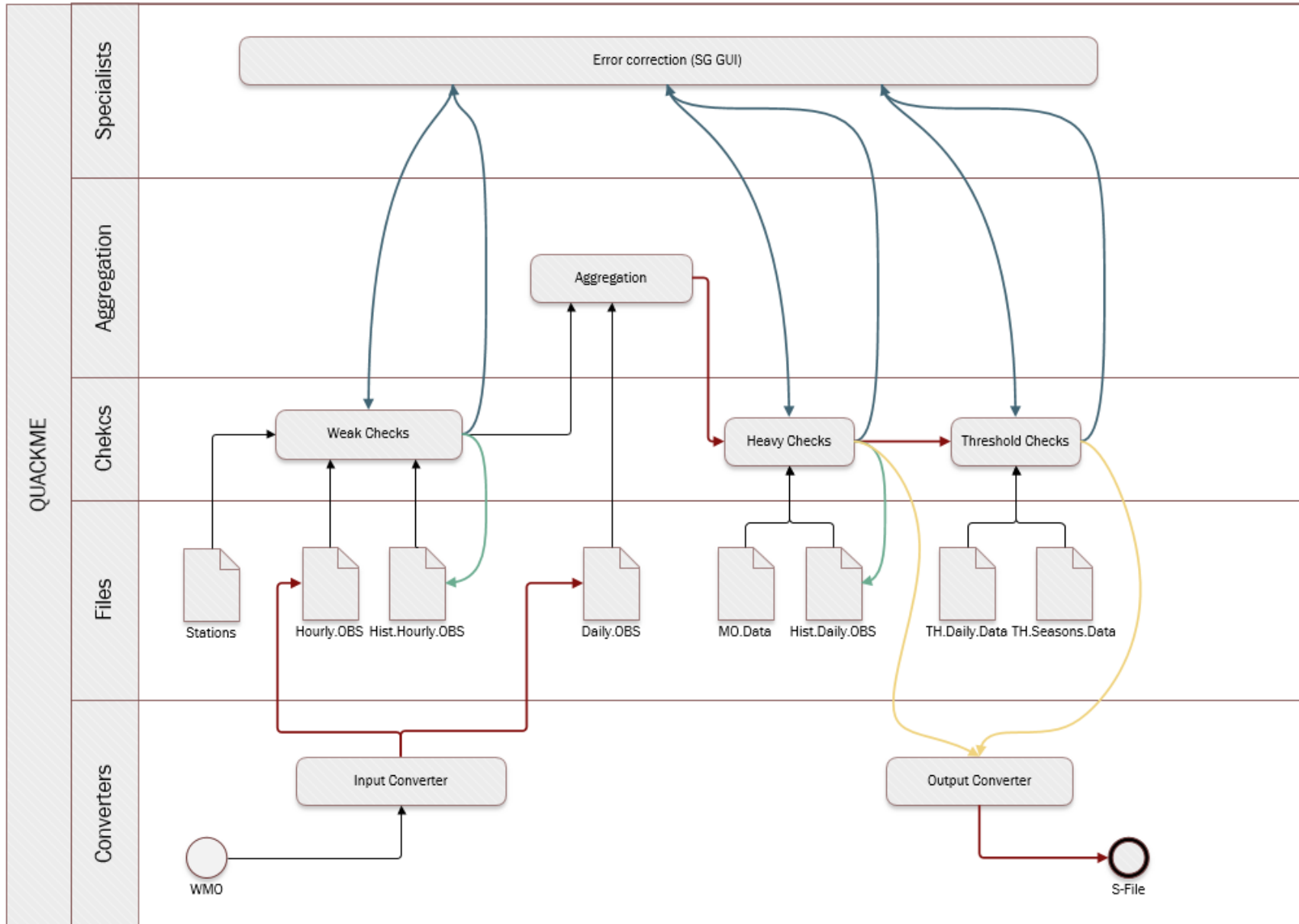
In general, for a given step there are:

- 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 quality checks flag
 - file with errors/problems (xml format)
 - historical file, if necessary
 - data for suppliers (S-File)

In details, the workflow contains the following steps:

- **Input Converter** - converts data from the meteorological data provider format to an internal data format (see section 3.2).
- **Weak Checks** –executes the weak checks
- Input:
 - Input data from meteorological data providers (converted into the internal data format) and, eventually, corrected data (from the SG)
 - It can use both historical hourly observations and the NRT station data

- Output:
 - File with correct hourly observation data (internal data format)
 - New historical hourly observations relative to current date
 - File with errors (xml format)
- **Aggregation** –transforms hourly observations into daily aggregated data and derives some specific agro-climate indicators:
- Input:
 - Correct hourly observation data
 - Daily observational data, if given
- Output:
 - Daily aggregated data
- **Heavy Checks** – executes 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** – executes complex checks:
- Input:
 - Result of heavy checks data
 - Daily historical data
 - Daily and Seasonal Threshold-data
- Output:
 - File with correct data (internal format)
 - File with errors (xml format)
- **SConverter** – converts the output of Quackme to the S-File format :
- Input:
 - Correct data as produced by Heavy Checks or Model Checks
- Output:
 - S-File



3. QUACKME Steps

3.1 Input Converter

To have more flexibility on managing different formats of input data an input converter was developed. For each type of input format, it creates an interface (to be called from the main workflow) to convert the input data into the internal data format.

This module gives the possibility to use more suppliers of meteorological data without changing the workflow and the checks.

The converter also checks for duplicated lines in correspondence of <Station, Day&Time> and raises an error message when one is found.

Please note that it also replaces values equal to 9 in the N and L columns with 8.

3.1.1 Input Data Format

The input data format depends on the data provider.

3.1.2 Internal Data Format

The internal data format is a text file organized by columns. Its structure contains:

- First line dedicated to the header, with the name of the columns
- Subsequent lines with values (associated to each available observation) separated by a single space character
- The file cannot contain empty lines
- Two formats are used: one for hourly observations and one for daily observations

Conversion table between the CSV columns and the DAT file:

Property description	DAT columns	Note	CSV Columns	Unit of Measure
Station number	Station	Numerical value	INDEX	
Day Time	DayTime	Numerical value (YYYYMMDDHH24)	YYYY, MM, DD, HHmm	
Air Temperature	TT	Numerical value	TL	Celsius Degree
Dew point temperature	TD	Numerical value	TD	Celsius Degree
Maximum temperature 1 Hour	TX1	Numerical value	TX1	Celsius Degree
Minimum temperature 1 Hour	TN1	Numerical value	TN1	Celsius Degree
Maximum temperature 6 Hour	TX6	Numerical value	TX6	Celsius Degree
Minimum temperature 6 Hour	TN6	Numerical value	TN6	Celsius Degree
Maximum temperature 12 Hour	TX12	Numerical value	TX12	Celsius Degree

Property description	DAT columns	Note	CSV Columns	Unit of Measure
Minimum temperature 12 Hour	TN12	Numerical value	TN12	Celsius Degree
Hourly Precipitation	PREC	Numerical value	RR1h	mm
6 Hour Precipitation	PR06	Numerical value	RR06	mm
24 Hour Precipitation	PR24	Numerical value	RR24	mm
Precipitation	RR	Numerical value	RR	mm
Precipitation interval	TR	Numerical value (Hour number)	TR	
Snow depth	Snow	Numerical value	SNO	cm
Wind direction	DIR	Numerical value	DIR	Degree
Wind speed at 10 m	FF	Numerical value	FF	m/s
Cloud cover	N	Numerical value	N	Octas
Low cloud cover	L	Numerical value	L	Octas
Radiation 1 Hour	RD	Numerical value	GL1h	J/cm2
24 Hour Radiation	RD24	Numerical value	GL24	J/cm2
Atmospheric pressure at station level	AP	Numerical value	QFE	
Atmospheric pressure at station level	QFF	Numerical value	QFF	
Sunshine 1 Hour	SH	Numerical value	Sh	minutes
Sunshine 24 Hour	SH24	Numerical value	SS24	hours
Visibility	VIS	Numerical value	VIS	km
Relative Humidity	RH	Numerical value	(added by code)	%
Soil State	SOIL	Numerical value	GS	

3.1.3 MOS file Format

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

Conversion table between the CSV columns and the DAT file:

Property description	DAT columns	Note	CSV Columns	Unit of Measure
Station number	Station	Numerical value	INDEX	

Day Time	DayTime	Numerical value (YYYYMMDDHH24)	YYYY, MM, DD, HHmm	
Air Temperature	TT	Numerical value	TL	Celsius Degree
Dew point temperature	TD	Numerical value	TD	Celsius Degree
Hourly Precipitation	PREC	Numerical value	RR1h	mm
6 Hour Precipitation	PR06	Numerical value	RR06	mm
24 Hour Precipitation	PR24	Numerical value	RR24	mm
Wind speed at 10 m	FF	Numerical value	FF	m/s
1H Maximum precipitation	RRRX1	Numerical value	LAYER1	mm
6H Maximum precipitation	RRRX6	Numerical value	LAYER2	mm
12 Maximum precipitation	RRRX12	Numerical value	LAYER3	mm
24 Maximum precipitation	RRRX24	Numerical value	LAYER4	mm

3.2 Weak checks level

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

- 2-m air temperature
- Minimum and maximum 2-m air 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 checks for visibility and state of the soil are not configured.

The following weak checks are implemented:

- 2-m air temperature checks
 - 2-m air temperature values outside -80 / +60 Celsius degrees. Values flagged as WRONG.
 - Difference of 2-m air temperature between two consecutive records (hourly observations) = 0 for stations with integer values or < 0.1 degree for station with not integer values. The alert is raised after the error has been identified at least 8 times for a set of consecutive timestamps. All the values identified with the previous criteria will be flagged as WRONG.

- Difference of 2-m air temperature between two consecutive records (either hourly or 3-hourly) > 15 degree. Values flagged as SUSPICIOUS.
 - If the absolute difference between the observed 2-m air temperature and the correspondent MOS 2-m air temperature is greater than 12 degrees the observed air temperature is flagged as WRONG
 - If the absolute difference between the observed 2-m air temperature and the correspondent MOS air temperature is greater than 8.5 degrees and less than/equal to 12 degrees then the observed air temperature is flagged as SUSPICIOUS
- All 2-m air minimum temperature (hourly, 6 hourly, 12 hourly) checks
 - Minimum temperature value outside -80 / +40 Celsius degrees. Value is flagged as WRONG (for TN1, TN6, Tn12)
 - Difference of minimum temperature between two consecutive records (hourly observations – TN1) < 0.1 degree. Minimum temperature difference between two consecutive records (either hourly or 3-hourly) > 15 degree. Values are flagged SUSPICIOUS.
- All 2-m air maximum temperature (hourly, 6 hourly, 12 hourly) checks
 - Maximum temperature value outside -80 / +60 Celsius degrees. The value is flagged WRONG (for TX1, TX6, TX12).
 - Difference of maximum temperature between two consecutive records (hourly observations – TX1) < 0.1 degree.
 - Maximum temperature difference between two consecutive records (either hourly or 3-hourly) > 15 degree. Values flagged SUSPICIOUS.
- Dew point temperature checks
 - Dew Point temperature outside -80 / 35 Celsius degrees. Values flagged WRONG.
 - Dew point temperature > Air Temperature. If the difference between dew point temperature and air temperature is less than 1 degree then the dew point temperature is replaced with the air temperature, else the value is flagged WRONG for both parameters (TD, TT)
 - Dew point temperature difference between two consecutive records (hourly observations) > 11 degree. Values are flagged SUSPICIOUS.
 - Dew point temperature difference between two consecutive records (hourly observations) = 0 for stations with integer values and < 0.1 degree for stations with not integer values. The alert is raised after the error has been identified at least 8 times for a set of consecutive timestamps. All values reporting error will be flagged as WRONG.
 - 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 is flagged SUSPICIOUS
 - If the absolute difference between the 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 is flagged SUSPICIOUS
- Relative humidity checks
 - Relative humidity outside 0 - 100 %. The value is flagged WRONG.
 - Relative humidity difference between two consecutive records (hourly observations) < 0.5% and both values being less than 95%. The alert is raised after the error has been identified at least 20 times. Values are flagged SUSPICIOUS and the error raised in correspondence of the 21-th observation. This check is not active for the observational networks reporting integer values.
- Atmospheric pressure checks

- Atmospheric Pressure outside 500 – 1100 hPa. Value is flagged WRONG.
 - Atmospheric Pressure difference between two consecutive records (either hourly or 3-hourly) > 15 hPa. Value is flagged SUSPICIOUS.
- MSL atmospheric pressure checks
 - MSL atmospheric pressure outside 950 – 1060 hPa. The value is flagged WRONG
- Wind direction checks
 - Wind direction outside 0-360 degrees. The value is flagged WRONG.
 - Wind direction difference between two consecutive records (hourly observations) < 5 degree conditional on having both wind speed values greater than 4.6 m/s. The alert message is raised after the 20-th identified value and only if the altitude of the station is less than/equal to 1000m. Values are flagged SUSPICIOUS.
 - Wind direction difference between two consecutive records (hourly observations) > 40 degree, conditional on having both wind speed values greater than 4.6 m/s. The alert message is raised after the 5-th identified value and only if the altitude of the station is less than/equal to 1000m. Values are flagged SUSPICIOUS.
- Wind speed checks
 - Wind speed outside 0 – 75 m/sec. The value is flagged WRONG.
 - Wind speed difference between two consecutive records (hourly observations) < 0.5 m/sec, conditional on having both wind speed values greater than 4.6 m/s. The alert message is raised after the 15-th identified message. Values are flagged SUSPICIOUS.
 - Wind speed difference between two consecutive records (hourly observations) > 15 m/sec. The alert message is raised after the 10-th identified value. Values are flagged SUSPICIOUS.
 - Wind speed > 3.5 m/ and Wind direction is NA. In this case, the value will be considered SUSPICIOUS.
 - Absolute difference between the observed wind speed and the correspondent MOS wind speed greater than 40 knots. The value is flagged WRONG
 - Absolute difference between the observed wind speed and the correspondent MOS wind speed greater than 25 knots and less than 40 knots. The value is flagged 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 flagged 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. The value is flagged WRONG.
 - Cloud cover is 0 and precipitation > 0. The value is flagged SUSPICIOUS.
 - Cloud cover 8 and sunshine duration > 0. The value is flagged SUSPICIOUS.
- Low Cloud cover checks
 - Low Cloud Cover outside 0-8. The value is flagged WRONG.
- Solar radiation checks (hourly observation)
 - Solar radiation outside 0 -1600 W/m². The value is flagged WRONG.
 - Sunshine duration > 0 and radiation is 0. The value is flagged SUSPICIOUS.
 - ~~○ Solar radiation 24 hours outside 0 – 38400 W/m². The value is flagged WRONG.~~

- Sunshine checks (hourly observation)
 - Sunshine duration > 60 minutes. The value is flagged WRONG.
 - Sunshine duration > 24 hours. The value is flagged WRONG.
- Precipitation checks (hourly observation)
 - Precipitation outside [0-400] mm. The value is flagged WRONG.
 - Precipitation ≥ 40 mm and < 400 mm. The value is flagged SUSPICIOUS.
 - If the TR = 1 and RR are present and valid the same check applied to hourly observation is applied to RR values
 - If the TR in (3, 6, 12) and RR are valid the following checks are done:
 - If RR < 0 or ≥ 400 mm the value is considered WRONG
 - If RR ≥ 80 mm and < 400 mm the value is considered SUSPICIOUS
 - The same check is applied to the field RR6h (6-hour precipitation)
 - For each precipitation property (1H, 6H, 12H – not available yet, 24H) the following two parameters are derived, if the corresponding MOS data are available:
 - Standard deviation (SD): $\min[(5 * \max[0.2, RRR_{mos}]), \max[1, \text{abs}(RRR_{X1/6/12/24} - RRR_{mos})]]$
 - Scaled difference (F): $(RRR_{mos} - RRR) / RRR_{mos}$

Once these two parameters are calculated the following conditions must be checked:

- When $\text{abs}(RRR_{obs} - RRR_{mos}) > (6 * SD)$ AND $RRR_{obs} > 20$ AND $F > 5$, the observed precipitation property is flagged WRONG
- When $\text{abs}(RRR_{obs} - RRR_{mos}) > (6 * SD)$ AND $RRR_{obs} > RRR_{mos}$ AND
 - ($(RRR_{obs} \leq 10$ AND $RRR_{mos} \geq 0.5$) OR
 - ($RRR_{obs} \leq 15$ AND $RRR_{mos} \geq 2.5$)) is not satisfied, then the observed precipitation property is flagged SUSPICIOUS
- Cross checks precipitation: these checks are carried out on the 24H precipitation:
 - Check if at 06:00 a.m. of the next day the 24H precipitation is valid and in that case, use the interval 06 current day – 06 next day for the precipitation cross checks
 - If at 06:00 a.m. of the next day the 24H precipitation is not available, check if at 00 of the next day the 24H precipitation is valid and in that case, use the interval 00 current day – 00 next day for the precipitation cross checks
 - Assuming that a 24H precipitation value can be derived by using the corresponding time interval the following precipitation cross checks are performed:
 - Check if the PR06 values in the interval (12, 18, 00 of the next day, 06 of the next day or 06, 12, 18, 00 of the next day) are all valid and compare the sum of these PR06 values with PR24. If the difference is greater than 10 mm, a SUSPICIOUS flag is raised for the PR24 value and for all PR06 values
 - If not all the PR06 values are valid, the same check is done by using RR/TR combination. If for specific time the TR = 6 will use the RR value, if the TR = 12 than will try to get the value by difference with the RR value at [current time - 6H] if that observation presents TR = 6, otherwise the value will not be considered. In all four RR values was identified and the difference between their sum and 24H precipitation value is great than 10 mm, a SUSPICIOUS alert will be raise for the PR24 value and for all RR values.

- If all 1H precipitation values are available and valid, if the difference between their sum and the 24H precipitation value is greater than 10 mm a SUSPICIOUS flag is raised for the PR24 value and for all the 24 hourly precipitation values.
- In case the 24H precipitation are not available at 06:00 next day, the application check for a valid value at 00:00 next day. If at this time the 24H precipitation are valid the cross checks are done using the interval 00 current – 00 next day.
- If 24H precipitation value is not available the precipitation cross checks are not performed

The following properties are derived during the WeakChecks step:

- **D_E** = Saturation vapor pressure. To calculate this property we use TT (2-m air temperature) and TD (dew point temperature) properties into the formula (**ESAT(TT, TD)**) :

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

Dove:

- **LVAP** Latent Heat of Vaporization, derived by using the formula :

$$(597 - 0.566 * TT) * 4186$$

- **RV** it's a constant value = **461.51**
- **Celsius2Kelvin** transforms temperature from Celsius to Kelvin degree
- **D_VPD** = Vapor pressure deficit. To calculate this property we use TT and TD properties. The vapor pressure deficit is calculated as the difference between two saturation values:
 - One, obtained by using the saturation vapor pressure with only TT
 - One, obtained by using the saturation vapor pressure with TT and TD
 - The formula : $ESAT(TT, TT) - ESAT(TT, TD)$
- **D_SLOPE**= Slope of saturation vapor pressure. The slope is calculated by using the following formula:

$$238.102 * 17.32491 * (D_E + D_VPD) / (TT + 238.102)**2$$

- **D_RH** = Relative humidity. The relative humidity is calculated by using the following formula:

$$100 * ESAT(TT, TD) / ESAT(TT, TT)$$

The WeakChecks module can use in input the error file too (XML format), after the errors are fixed by a 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 **does not execute the checks for the date when data are missing**.

Another input possibility for the WeakChecks module is to specify the "-m <hour>" option. This option allows 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 contain the hourly observation from <hour> + 1 of the day in elaboration, meanwhile from the hourly history file we read the observation of the day in elaboration with time <= <hour>
- for example a command line of WeakChecks with -d "20191101" -m "06" has the following meaning:
 - o day in elaboration 01.11.2019
 - o from the H.20191101.hist file all the observations with DayTime <= 2019110106 are taken
 - o the CSV must contain observations with DayTime >= 2019110107

An output file of WeakChecks module contains the following properties:

Property description	Acronim	Note	Unit of Measure
Station number	Station	Numerical value	
Day Time	DayTime	Numerical value (YYYYMMDDHH24)	
Air Temperature	TT	Numerical value	Celsius Degree
Dew point temperature	TD	Numerical value	Celsius Degree
Maximum temperature 1 Hour	TX1	Numerical value	Celsius Degree
Minimum temperature 1 Hour	TN1	Numerical value	Celsius Degree
Maximum temperature 6 Hour	TX6	Numerical value	Celsius Degree
Minimum temperature 6 Hour	TN6	Numerical value	Celsius Degree
Maximum temperature 12 Hour	TX12	Numerical value	Celsius Degree
Minimum temperature 12 Hour	TN12	Numerical value	Celsius Degree
Hourly Precipitation	PREC	Numerical value	mm
24 Hour Precipitation	PR24	Numerical value	mm
Precipitation	RR	Numerical value	mm
Precipitation interval	TR	Numerical value (Hour number)	
Snow depth	Snow	Numerical value	cm
Wind direction	DIR	Numerical value	Degree
Wind speed at 10 m	FF	Numerical value	m/s
Cloud cover	N	Numerical value	Octas
Low cloud cover	L	Numerical value	Octas
Radiation 1 Hour	RD	Numerical value	J/m ²
24 Hour Radiation	RD24	Numerical value	J/m ²

Property description	Acronim	Note	Unit of Measure
Atmospheric pressure at station level	AP	Numerical value	
Atmospheric pressure at station level	QFF	Numerical value	
Sunshine 1 Hour	SH	Numerical value	minutes
Sunshine 24 Hour	SH24	Numerical value	minutes
Visibility	VIS	Numerical value	km
Relative Humidity	RH	Numerical value	%
Saturation vapor pressure	D_E	Numerical value, Calculated	hPa
Relative Humidity	D_RH	Numerical value, Calculated	%
Vapor pressure deficit	D_VPD	Numerical value, Calculated	hPa
Slope of saturation vapor pressure	D_SLOPE	Numerical value, Calculated	hPa/Celsius Degree
Soil status	SOIL	Numerical value	

3.3 Aggregation

The main scope of this step is the aggregation of hourly observations to daily scale.

As WeakChecks step, the Aggregation level uses a specific configuration file (**Aggregation.XML**). The file contains the definition of 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 the configuration on the aggregation process from hourly daily values. It is an XML file containing a definition for each property to be processed.

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

In the *PROPERTY* tag it is possible to define the following tags:

- NAME – (mandatory) it defines the name of the property (column) in the output file
- FormulaValue – (mandatory) it defines the formula to be applied. This tag can get the following attributes:
 - Type (mandatory) with two possible values:
 - "Fixed" when the property value is retrieved from a specific property and a specific *datetime*. When the TYPE is Fixed, it is necessary to add the other attributes:
 - "Time" –hour of the observation (e.g., 00, 06 or 23)
 - Day –day of observation. It contains the following possible values:
 - 0 = day in elaboration

- 1 = next day
 - -1 = previous day
- "ReferenceProperty" – it specifies the name of the property to read
- "Formula" when it is necessary to call a method to calculate the property's value
- "RoundedDigits" – not mandatory, specify the number of digits to round the daily value
- "Parameter" - this attribute can be added when Type="Formula" and can get two possible values:
 - "V" – vector
 - "M" – matrix
- In the "FormulaValue" tag, when Type = "Formula" it is necessary to specify the following tags:
 - *Formula* – (mandatory) name of the formula, that must correspond to a method implemented into the "Aggregation.Formulas.R" file
 - Parameter – (optional) allows to specify an acronym that will pass to the method some additional parameters. The following acronyms are defined at the moment:
 - Current.Date – day in elaboration
 - ~~Inter.Data~~ – ~~interpolated data~~
 - Day.Agg – all the properties aggregated for the day in elaboration
 - Expected.Values – number of expected values for the MEAN method
- For each formula, sometimes it is 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 time of hourly observations, "Day" contains the day of the observation (relative value). If values from more days are needed, then 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 the "Start" tag defines the start of the interval and the "End" tag defines the end of the interval. With these tags the application retrieves all the observations between 03 of the day in elaboration until 06 of the next day.
- When the "Parameter" tag is "V" there is a need to specify the "ReferenceProperty" tag with the name of the property to extract and use as an input to the method. When the value is "M" the matrix with all the properties is given to the method
- Independently of the "Parameter" tag (either "V" or "M"), the temporal tags (Specific, Start, End) are applied to extract only the values for the specific times of observation.
- Another tag that can be set within the "Property" tag is "NAFormula" (optional). It defines an alternative formula to be applied if the formula specified by "FormulaValue" returns NA.

Samples:

1. <Property>


```

      <Name>TT_09</Name>
      <FormulaValue Type="Fixed" Time="09" Day="00" ReferenceProperty="TT"
      RoundedDigits="1"/>
      </Property>,
      
```

The "Aggregation" module creates a property with name "TT_09". The value is retrieved from the property TT for the day in elaboration at time 09:00 (from the input file O.WeakChecks.<day>.dat)

2. <Property>

```
<Name>N</Name>
<FormulaValue Type="Formula" Parameter="V" ReferenceProperty="N"
RoundedDigits="2">
    <Specific Time="00, 03, 06, 09, 12, 15, 18, 21" Day="00" />
    <Formula>Aggregation.Mean.80</Formula>
    <Parameter>Expected.Values</Parameter>
</FormulaValue>
</Property>
```

The "Aggregation" module creates a property with name "N". The value is calculated using the formula "Aggregation.Mean.80" that receives in input:

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

At the same time the method receives in input the number of expected values for the mean calculation (in this case Expected.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>
```

the "Aggregation" module creates a property with name "RH06". Its value is retrieved from the property RH at time 06 of the day in elaboration. If the returned value is NA, the application tries to retrieve the value of the property from the property D_RH at time 06 of the day in elaboration (NAFormula).

Within each **Property** tag it is possible to define an 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** represents the name of the property to be interpolated (property available in the WeakCheks files)
- The START tag contains two attributes: Time = the hour, Day = relative value for the day (00 = current day, -01 = previous day, 01 = next day). The tag specifies the start of the interval to be interpolated for the Reference Property
- The END tag contains two attributes: Time = the hour, Day = the relative value for the day (00 = current day, -01 = previous day, 01 = next day). The tag specifies the end of the interval to be interpolated for the Reference Property

- The **AllowNegative** tag allows to specify when the negative interpolated values are accepted. If the tag is missing, its value is equal to FALSE (i.e. the negative interpolated values are not accepted) and the negative values are reset to 0.

The Aggregation.XML changes according to the ROI. At the moment, there are two versions, one for Europe and one for China.

3.3.1.1 Aggregated properties for ROI EUROPE

Properties produced by the Aggregation step for the ROI – Europe are shown in the table below:

Property description	Acronim	Note	Unit of Measure
Station number	Station	Numerical value	
Day Time	DayTime	Numerical value (YYYYMMDD)	
Total cloud cover	N	Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC next day	Octas
Low cloud cover	L	Daytime mean using observation from 06, 09, 12, 15, 18 UTC	Octas
Cloud cover daytime mean	NDT	Daytime mean using observation from 06, 09, 12, 15, 18 UTC	Octas
Measured Sunshine	MSUN	See the formula	Hours
Measured Radiation	MRAD	See the formula	MJ/m ²
Calculated Radiation	CRAD	See the formula	MJ/m ²
Minimum Temperature	TN	See the formula	Celsius Degree
Maximum Temperature	TX	See the formula	Celsius Degree
Daily Mean Vapor pressure	MVP	Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC	hPa
Daily Mean Wind speed	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
Precipitation	RRR	See the formula	mm

Property description	Acronim	Note	Unit of Measure
Air temperature	TT<hour>	Air Temperature at 06, 09, 12, 15 and 18 of current day	Celsius Degree
Relative Humidity	RH<hour>	Relative Humidity at 06, 09, 12, 15 and 18 of current day	%
Daily Mean Vapor Pressure Deficit	VPD	Daily mean using observation from 00, 03, 06, 09, 12, 15, 18, 21 UTC of current day, 00 UTC of next day	hPa
Daily Mean of Slope Saturation	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
ANGOT Radiation	ANGRAD	See the formula	KJ/m ²
Angstrom Prescott Radiation	APRAD	See the formula	KJ/m ²
Supit-Van Kappel Radiation	SVKRAD	See the formula	KJ/m ²
Hargreaves Radiation	HGVRAD	See the formula	KJ/m ²
Penman potential evapotranspiration from a free water surface	E0	See the formula	mm/day
Penman potential evapotranspiration from a moist bare soil surface	ES0	See the formula	mm/day
Penman potential evapotranspiration from a crop canopy	ET0	See the formula	mm/day
Snow depth	SNOW	Instantaneous value at 06 UTC	cm
Visibility	VIS	Formula using observation from 06, 09, 12, 15, 18 UTC	km
Absolut daily minimum temperature	TND	See the formula	Celsius Degree

3.3.1.2 Aggregated properties for ROI CHINA

Properties produced by the Aggregation step for the ROI – CHINA

Property description	Acronim	Note	Unit of Measure
Station number	Station	Numerical value	
Day Time	DayTime	Numerical value (YYYYMMDD)	
Total cloud cover	N	Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day	Octas
Low cloud cover	L	Not available	Octas
Cloud cover daytime mean	NDT	Daytime mean using observation from 00, 03, 06, 09, 12 UTC	Octas
Measured Sunshine	MSUN	Not available	Hours
Measured Radiation	MRAD	Not available	MJ/m ²
Calculated Radiation	CRAD	See the formula	MJ/m ²
Minimum Temperature	TN	See the formula	Celsius Degree
Maximum Temperature	TX	See the formula	Celsius Degree
Daily Mean Vapor pressure	MVP	Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day	hPa
Daily Mean Wind speed	FF	Daily mean using observation from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day	m/s
Precipitation	RRR	See the formula	mm
Air temperature	TT<hour>	Air Temperature at 00, 03, 06, 09, 12 of current day	Celsius Degree
Relative Humidity	RH<hour>	Relative Humidity at 00, 03, 06, 09, 12 of current day	%
Daily Mean Vapor Pressure Deficit	VPD	Daily mean using from 18, 21 UTC of previous day, 00, 03, 06, 09, 12, 15, 18 UTC current day	hPa

Property description	Acronim	Note	Unit of Measure
Daily Mean of Slope Saturation	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
ANGOT Radiation	ANGRAD	See the formula	KJ/m ²
Angstrom Prescott Radiation	APRAD	See the formula	KJ/m ²
Supit-Van Kappel Radiation	SVKRAD	See the formula	KJ/m ²
Hargreaves Radiation	HGVRAD	See the formula	KJ/m ²
Penman potential evapotranspiration from a free water surface	E0	See the formula	mm/day
Penman potential evapotranspiration from a moist bare soil surface	ES0	See the formula	mm/day
Penman potential evapotranspiration from a crop canopy	ET0	See the formula	mm/day
Snow depth	SNOW	Instantaneous value at 00 UTC	cm
Visibility	VIS	Formula using observation from 00, 03, 06, 09, 12	km
Absolut daily minimum temperature	TND	See the formula	Celsius Degree

3.3.2 Aggregation formulas

For each property produced by the Aggregation step, the module is calling a formula.

Before applying those aggregation formulas, some hourly properties undergo 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 in the Aggregation.XML file.

Since version 1.5.0, the interpolation is not anymore available for the 1H precipitation property.

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 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 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 observations (property L): 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC – current day
- **Measured sunshine:**
 - The system searches, first, a 24-hour Sunshine observation between observation of 06 of the next day and sunset time of the current day.
 - If the previous step does not return a valid value, the system tries to derive the measured sunshine as follows: it retrieves the sunrise and the sunset time for the current day. If not available, it considers the interval 06 – 21 of the current day.
 - If all the hourly observations in the interval determined at the previous point are available, the measured sunshine are calculated as the sum of hourly sunshine observations.
- **Measured radiation:**
 - The system searches, first, a 24-hour Radiation observation between the 00 of the next day and the sunset time of the current day.
 - If the previous step does not return a valid value, the following algorithm is applied: it retrieves the sunrise and the sunset time of the current day. If not available, it considers the interval 06 – 21 of the current day.
 - If all the hourly observations are available in the interval determined at previous step, the measured radiation is calculated as the sum of hourly radiation observations.
- **Calculated radiation:**
 - If the measured radiation is correctly calculated it is also used to return the calculated radiation
 - Second alternative. If the measured sunshine is valued, then the Angstrom Radiation is calculated and used
 - Third alternative. If cloud cover, maximum and minimum temperatures are valid values then the Supit radiation is derived and used
 - Fourth alternative. If cloud cover is not available but minimum and maximum temperature are, then the Hargreaves radiation is computed and used

- **Minimum temperature:**
 - If the TN12 property is available at 06 of the next day, its value is used
 - If the previous step does not return a valid value, and if the TN6 property is available at 06, 12, 18, 24 of the current day and 06 of the next day, then their minimum value is taken
 - If the previous step does not return a valid value and if the hourly observations are available, get the minimum from these TN1 values in the interval 18 of the previous day and 06 of the current day
 - If previous step does not return a valid value, the application calculates the minimum of all TT values in the interval 18 of the previous day – 06 of the current day
 - If there is an air temperature value in the interval 18 of the previous day – 06 of the current day that is less than the calculated daily minimum temperature and higher than the minimum temperature from the MOS file at 06 UTC, then the daily minimum temperature is replaced by the minimum temperature from the MOS file at 06 UTC
- **Maximum temperature:**
 - Check if the TX12 property is available at 18 of the current day
 - Check if the TX12 property is available at 15 of the current day (if the previous step fails)
 - If the previous step does not return a valid value, check if the TX6 property is available at 12, 18 and 24 of the current day and 06 of the next day and get the maximum value of these four observations
 - If the previous step does not return a valid value and if the hourly observations are available, get the maximum from these TX1 values in the interval 00-23 of the current day
 - If previous step does not return a valid value, the application calculates the maximum between all TT values in the interval 06 – 18 of the current day
 - If there is an air temperature value in the interval 06 of the current day – 18 of the current day that is higher than the calculated daily maximum temperature and lower than the maximum temperature from the MOS file at 18 UTC, then the daily maximum temperature is replaced by the maximum temperature from the MOS file at 18 UTC
- **Daily mean vapor pressure** = mean of hourly valid vapor pressure values (D_E) 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) 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, the system looks for the 24-hour precipitation value (PR24) at 06 of the next day.
 - Second alternative. Sum the 6-hourly precipitation values (PR06) in the interval 06 UTC – 06 UTC of the next day, if all these values are valid
 - Third option. Sum the hourly observation (PREC) if all the 24 observations (in the interval 07 of the current day– 06 of the next day) are valid.
 - Fourth option. Use PR24 at 00 UTC of the next day (if it exists and it is valid)
 - Fifth option. Sum all the 6-hourly observations (PR06) in the interval 00 UTC – 00 UTC of the next day (if all values are valid)
 - Sixth option. Sum all the hourly observations (PREC) in the interval 01 UTC – 00 UTC of the next day (if all 24 observations are valid).
 - If none of the previous step returns a valid value, the algorithm looks for the first hourly observation having both TR & RR (in the interval 06 of the next

day and 06 of the current day). According to the time of the identified observation it adapts the length of the interval as follows:

- If TR & RR is found at either 06 or 03 of the next day, the interval time is 24 hours
- If TR & RR is found at either 24:00 or 21:00 of the current day the interval time is 18 hours
- If TR & RR is found at either 18 or 12 of the current day the interval time is 12 hours
- If all the previous steps fail, the algorithm cannot be applied
- 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 is 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:** values at the following time of observation: 06, 09, 12, 15, 18 of the current day
- **Relative Humidity:** values at the following time of observation: 06, 09, 12, 15, 18 of the current day. In case values are either missing or not valid the system provides an estimated relative humidity (D_RH)
- **Daily mean vapor pressure deficit** = mean of hourly valid vapor pressure deficit values (D_VPD) at 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) at 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 of the current day
- **Visibility** = daytime visibility. Hourly observations in the interval 06 – 18 of the current day.
- **Absolute Daily Minimum Temperature:**
 - First option. Get the minimum of TN12 at 12 UTC of the current day and 00 UTC of the next day
 - Second option. Get the minimum of TN6 at 06, 12, 18 of the current day and 00 UTC of the next day
 - Third option. Get the minimum value of TN1 between 01 of the current day and 00 of the next day
 - Last option. Get the minimum value of TT between 01 of the current day and 00 of the next day

Concerning the mean values, they are calculated only if at least 80% of valid values (in relation to the number of expected values) are available.

After all formulas are applied and all parameters are calculated, the **Aggregation** module checks if for any stations the minimum temperature is greater than maximum temperature. For these cases, the module will try to recalculate the minimum and maximum temperature as follows:

- A. **Minimum temperature:**

- for stations reporting hourly minimum air temperature is calculated using the interval 19 UTC previous day – 18 UTC current day, if at least 80% of the 24 values are valid

- for stations reporting every 3 hours, minimum air temperature is calculated from the observations at previous day (21 UTC, 00 UTC) and current day (03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC) if at least 80% of these 8 values are valid

- once the new value is calculated it becomes the new minimum temperature, if it is lower than the value calculated by the formula

- B. Maximum temperature:

- for stations reporting hourly maximum air temperature is calculated using the interval 07 UTC current day – 06 UTC next day, if at least 80% of the 24 values are valid

- for stations reporting every 3 hours maximum air temperature is calculated from the observations at current day (09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC) and next day (00 UTC, 03 UTC, 06 UTC) if at least 80% of these 8 values are valid

- once the new value is calculated it becomes the new maximum temperature, if it is greater than the value calculated by the formula

3.3.2.2 Aggregation rules for ROI – CHINA

- **Total cloud cover** = mean of hourly cloud cover 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 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 it is also used for calculated radiation
 - Second alternative. If the measured sunshine is valued, the Angstrom Radiation is calculated
 - Third alternative. If cloud cover, maximum and minimum temperatures are valid then the Supit radiation is computed.
 - Fourth alternative. If cloud cover is not available but minimum and maximum temperature are available, then the Hargreaves radiation is computed
- **Minimum temperature:**
 - Check if the TN12 property is available at 06 of the current day
 - If the previous step does not return a valid value, check if the TN6 property is available at 06, 12, 18 of the previous day and at 06, 12 of the current day, and get the minimum value of all these observations
 - If the previous step does not return a valid value, in case hourly observations are available, get the minimum from all the TN1 values in the interval 06 of the previous day and 06 of the current day
 - If the previous step does not return a valid value, the application calculates the minimum between all TT values in the interval 06 previous day – 06 current day
- **Maximum temperature:**
 - Check if the TX12 property is available at 18 of the current day

- If the previous step does not return a valid value, check if the TX6 property is available at 18 of the previous day and at 00,06,12, 18 of the current day and get the maximum value between the four observations
- If the previous step does not return a valid value, in case hourly observations are available, get the maximum from all the TX1 values in the interval 18 of the previous day and 18 of the current day
- If previous step does not return a valid value, the application calculates the maximum between all the TT values in the interval 18 previous day - 18 current day
- **Daily mean vapor pressure** = mean of hourly valid vapor pressure values (D_E) 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) from 18 UTC, 21 UTC – previous day, 00,03, 06, 09, 12, 15, 18 UTC – next day
- **Precipitation:**
 - If available, use the 24-hour precipitation (PR24) at 00:00 of the next day. If not, use the PR24 at 21 of the current day
- **Air Temperature:** at the following time of observation: 00, 03, 06, 09, 12 of the current day
- **Relative Humidity:** at the following time of observation: 00, 03, 06, 09, 12 of the current day. In case the reported values of relative humidity are missing or not valid (RH), the system retrieves the calculated relative humidity (D_RH)
- **Daily mean vapor pressure deficit** = mean of hourly valid vapor pressure deficit values (D_VPD) at 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) at 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 of the current day
- **Visibility** = daytime mean of visibility hourly observations using the interval 00 – 12 of current date.
- **Absolute Daily Minimum Temperature:** TBD

Mean values are only calculated if at least 80% of valid values (w.r.t. the number of expected values) are available.

3.4 Heavy checks

More complex checks are executed after the aggregation step.

All the errors to be managed are configured in the Messages.HeavyChecks.xml. Each message has a warning level (flag) that determines how the path of the property value. Two possible flags can be configured in the XML file:

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

The “HeavyChecks” step produces three files:

- XML file: list of stations/observations for which at least an error message exists (either W or S)
- DAT file: it contains all stations/observations, transformed as follows:
- If a property has at least one wrong flag, then the property has the value NA and status = "W" (Wrong)
- If a property has only suspicious flags, then the property maintains the value and get the status = "S" (Suspicious)
- If a property has NA and does not have any alerts, then a value from the model file can be used (if exists). Then, in the XML file this property gets the status = "R" (replaced).
- The DAT file is then copied to the history path for future checks (name of the file changed to D.<YYYYMMDD>.hist)

The following heavy checks are currently available:

- **Cloud check**
 - Cloud value outside the interval [0:8] Octas. The value is considered WRONG.
 - Cloud cover is 0 and precipitation > 0. The value is considered SUSPICIOUS and the error reported on both parameters
 - Cloud cover = 8 and sunshine duration > 0.5 * day sunlight duration. The value is considered SUSPICIOUS.
- **Minimum temperature check**
 - Minimum temperature outside the interval [-35:35] Celsius Degrees. The value is considered WRONG.
- **Maximum temperature checks**
 - Maximum temperature outside the interval [-20:50] Celsius Degrees. The value is considered WRONG.
 - Difference between maximum and minimum temperatures outside the interval [0:30] Celsius Degrees. Both min and max temperatures will be considered WRONG.
 - If the maximum temperature is lower than minimum temperature, both values will be considered SUSPICIOUS.
- **Snow check**
 - If the mean temperature is greater than 5 Celsius Degree, the SNOW value is considered WRONG.
- **Measured Sunshine check**
 - If the Measured Sunshine is outside the interval [0:24] hours, its value is considered WRONG.
 - Measured Sunshine > 1 and Measured Radiation is 0. The value is considered SUSPICIOUS.
- **Measured Radiation**
 - If the Measured Radiation is outside the interval (0:36] MJ/m² its value is considered WRONG.
- **Daily Mean Vapor Pressure**
 - If Daily Mean Vapor Pressure is outside the interval [0:38] hPa its value is considered WRONG.
- **Wind speed check**
 - If Wind speed is outside the interval [0:25] m/s its value is considered WRONG
- **Precipitation check**
 - If Precipitation is lower than 0 or greater equal than 500 mm its value is considered WRONG

- If Precipitation is greater equal than 100 but less equal than 500 mm the value will be considered SUSPICIOUS
- **Air Temperature Checks**
 - All Air Temperature values need to be in the interval [-50:50] Celsius Degrees, otherwise are considered WRONG
- **Relative Humidity Checks**
 - All Relative Humidity values need to be in the interval [0:100]%, otherwise are considered WRONG
- **Vapor pressure deficit Check**
 - If Vapor pressure deficit is outside the interval [0:60] hPa then its value is considered WRONG
- **SLOPE check**
 - If SLOPE is outside the interval [0:30] hPa/Celsius Degree, it is considered WRONG
- **Penman Evaporation check**
 - If Penman Evaporation is outside the interval [0:25] mm/day, then its value is considered WRONG
- Visibility check

If the visibility value is lower than 0 or greater than 80 km, then its value is flagged WRONG When the HeavyChecks module generates a KO file it is possible to fix the reported errors and run again the procedure by feeding the HeavyChecks module with the new KO file instead of the output from the Aggregation module. If the KO file is empty (i.e., contains only the tag "<Observations />") the module ends the run without executing any checks.

3.4.1 Exceptions management

The HeavyChecks module can manage exceptions for all its checks. All these exceptions are configured in the file HeavyChecks.Exceptions.xml, stored in the configuration path (specified into the command line of the module through the "-c" option).

The structure of the XML exceptions is the following one:

```
<Exceptions>
  <Exception>
    <Description>Mountain Station - Maximum Temperature</Description>
    <Error Property="TX" Code="001" />
    <Property Name="TX">
      <Condition Type="Include" Target="Value" MinEqual="-40"
MaxEqual="50" />
    </Property>
    <Condition Type="Include" Target="Altitude" MinEqual="1000" />
  </Exception>
  <Exception>
    <Description>Rusia & Kazakhstan - Maximum
Temperature</Description>
    <Error Property="TX" Code="001" />
    <Property Name="TX">
```



```

        <Condition Type="Include" Target="Value" MinEqual="-40"
MaxEqual="50" />
    </Property>
    <Condition Type="Include" Target="Latitude" MinEqual="40" />
    <Condition Type="Include" Target="Longitude" MinEqual="42" />
</Exception>
</Exceptions>

```

Tags explanation:

- Error => it configures the error for which the exception is valid. Two attributes (mandatory) can be used for this tag:
 - Property: name of the property for which the error is raised
 - Code: error code
- Property => it defines the property for which the exception is configured. More than one Property tag can be configured. In case the exception contains more tags Property their evaluation is performed by using an AND logical relation. The Property tag must contain the attribute Name with the name of the property.
- Condition => it defines a condition to check, related to a property or to a specific target:
 - Target: it defines the attributes related to the property or to the station to be verified:
 - Value = value of the property specified in the Property@Name
 - Latitude = latitude of the station
 - Longitude = longitude of the station
 - Altitude = altitude of the station
 - Type: it defines how to interpret the condition: "Include" or "Exclude". The "Exclude" type represents the opposite of "Include".
 - Min: it defines a minimum value to compare with (the value to check must be > Condition@Min)
 - MinEqual: minimum or equal comparison (the value to check must be >= Condition@MinEqual)
 - Max: it defines a maximum value to compare with (the value to check must be < Condition@Max)
 - MaxEqual: maximum or equal comparison (the value to check must be <= @ConditionMax)
 - Equal: fixed value
 - In case both Min and Max (or MinEqual and MaxEqual) are present the control checks that the value lies in the interval (Min, Max) (or [MinEqual, MaxEqual])

All Condition tags in the exceptions are managed by using a logical AND. When at least one exception is verified, the error is not raised and the value is consider correct.

3.5 Threshold checks

On the output of the "HeavyChecks" step, more complex checks are executed using the "threshold" files. Those files are created once *per* year considering historical data from past years and the module TGenerator.

The errors to manage are configured in two files:

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

- Messages.Seasons.ThresholdChecks.xml(file with seasonal threshold values)

All files have the same structure w.r.t. the messages. Each message has a warning level that determines how the property value is transferred to the output. Two possible flags can be configured in the XML file:

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

In input, the "ThresholdChecks" module uses two data files that contain daily and seasonal values:

- Threshold.Daily.dat: daily threshold data for each station
- Threshold.Seasons.dat: seasonal threshold data for each station

Another input of this module consists of history files, produced by the "HeavyChecks" or Aggregation modules (name D.<YYYYMMDD>.hist). Only the file relative to the previous day of elaboration is used.

The "ThresholdChecks" step will produce two files:

- One (XML format) file with the station observations for which exists at least one error message (with flag either W or S)
- One (DAT format) file with all stations' observations, transformed as follows:
 - If one property has at least one wrong alert than the property gets value NA
 - If the property has only suspicious alerts than the property keeps its value

Currently, the following checks are implemented within the "Threshold Checks" step:

- **Checks on daily thresholds**

- [Daily mean temperature greater than its 99th percentile associated with the day under analysis + 2°C](#) (MT99 column from the daily threshold file).
- [Daily maximum temperature greater than its 99th percentile associated with the day under analysis + 2°C](#) (MTX99 column from the daily threshold file)
- [Daily minimum temperature greater than its 99th percentile associated with the day under analysis + 2°C](#) (MTN99 column from the daily threshold file).
- [Daily mean temperature lower than its 1st percentile associated with the day under analysis - 2°C](#) (MT1 column from the daily threshold file)
- [Daily maximum temperature lower than its 1st percentile associated with the day under analysis - 2°C](#) (MTX1 column from the daily threshold file).
- [Daily minimum temperature lower than its 1st percentile associated with the day under analysis - 2°C](#) (MTN1 column from the daily threshold file).
- Daily precipitation greater than its 10-year return level of the season under analysis (RRR5Y column from the season threshold file).
- Daily wind intensity greater than its 10-year return level of the season 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 seasonal thresholds**

- [Difference in the mean temperature of two consecutive days greater than the 99th percentile value associated with the season under analysis + 2°C](#) (MT95 column from season threshold file).

- [Difference in the maximum temperature of two consecutive days greater than the 99th percentile value associated with the season under analysis + 2°C](#) (MTX95 column from season threshold file).
- [Difference in the minimum temperature of two consecutive days greater than the 99th percentile value associated with the season under analysis + 2°C](#) (MTN95 column from season threshold file).
- [Difference in the mean temperature of two consecutive days lower than the 1st percentile value associated with the season under analysis - 2°C](#) (MT5 column from season threshold file).
- [Difference in the maximum temperature of two consecutive days lower than the 1st percentile value associated with the season under analysis - 2°C](#) (MTX5 column from season threshold file).
- [Difference in the minimum temperature of two consecutive days lower than the 1st percentile value associated with the season under analysis - 2°C](#) (MTN5 column from season threshold file).
- Daily wind speed value greater than the 10-year return level of the season (RRR5Y column name). In this case a SUSPICIOUS flag is raised
- Precipitation daily value greater than the 10-year return level of the season (FF5Y column name). In this case a SUSPICIOUS flag is raised

As the ThresholdChecks module generates a KO file it is possible to fix the reported errors and run the module again. Then the input is the KO file with fixed errors. If the KO file sent as input is empty, (contains only the tag "<Observations />") the module ends without executing any checks.

3.5.1 Threshold file generator

The generator of the threshold files (daily and seasonal) is a module that does not run in NRT. It runs once per year and generates two files:

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

Property description	Column	Note	Unit of Measure
Station number	Station	Numerical value	
Day	Day	Julien day value for referenced year	
Mean temperature 99 percentile	MT99	Numerical value	Celsius Degree
Mean temperature 1 percentile	MT1	Numerical value	Celsius Degree
Maximum temperature 99 percentile	MTX99	Numerical value	Celsius Degree
Maximum Temperature 1 percentile	MTX1	Numerical value	Celsius Degree
Minimum Temperature 99 percentile	MTN99	Numerical value	Celsius Degree
Minimum Temperature 1 percentile	MTN1	Numerical value	Celsius Degree
Mean temperature prediction indicators using the mean temperature of 5 closest station [NOT AVAILABLE YET]			

Property description	Column	Note	Unit of Measure
1st indicator	MTC5_1	Numerical value	Celsius Degree
2nd indicator	MTC5_2	Numerical value	Celsius Degree
3rd indicator	MTC5_3	Numerical value	Celsius Degree
4th indicator	MTC5_4	Numerical value	Celsius Degree
5th indicator	MTC5_5	Numerical value	Celsius Degree
6th indicator	MTC5_6	Numerical value	Celsius Degree

The values for the temperature columns are calculated by using a complex approach described in the companion paper with data in the interval 1985 – 2018. The daily threshold file contains, for each station, the mean temperatures values for each day of the analyzed year.

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

Description	Column	Note	Unit of Measure
Station number	Station	Numerical value	
Season	Season	Text value with possible values: winter / spring / summer / autumn	
Mean temperature 99 percentile	MT95	Numerical value	Celsius Degree
Mean temperature 1 percentile	MT5	Numerical value	Celsius Degree
Maximum temperature 99 percentile	MTX95	Numerical value	Celsius Degree
Maximum Temperature 1 percentile	MTX5	Numerical value	Celsius Degree
Minimum Temperature 99 percentile	MTN95	Numerical value	Celsius Degree
Minimum Temperature 1 percentile	MTN5	Numerical value	Celsius Degree
Precipitation 10-year return levels of the season	RRR5Y	Numerical value	mm
Wind speed 10-year return levels of the season	FF5Y	Numerical value	m/s

The values of all columns are calculated by using the approach described in the companion paper with data in the interval 1985 –2018 for each station and for each season.

For each ROI (Europe, China) a set of threshold files (daily, seasonal) exists.

3.5.2 Convex Hull Exceptions management

The threshold checks module does not generate warnings when the same warning message is identified at many stations located within a convex hull with certain attributes.

This step requires the presence of the file ConvexHull.Threhsolds.xml into the CONFIG PATH of the module. The file contains a configuration for each error for which the convex hull approach must be applied. Please find bellow a sample of the configuration:

```
<Exception>
  <Area>Daily</Area>
  <Error Property="TX" Code="001" />
  <Stations_Counter>10</Stations_Counter>
  <Hull_Radius>150000;200000</Hull_Radius>
  <Hull_Area>71000</Hull_Area>
</Exception>, where
```

- <Area> Either Daily or Season
- <Error> contains the attributes to identify the error:
 - Property: the name of the property for which the error is raised
 - Code: error code
- <Stations_Counter> the minimum number of stations
- <Hull_Radius> a list of radius (in meters), separated by `;`, to be used for the cluster calculation
- <Hull_Area> the minimum limit of the convex-hull area

The convex-hullapproach is applied at the end of the Threshold Checks module.

Checks are run only if the number of stations having the error is greater/equal the value configured in the tag "Stations_Counter". Once the stations' number is reached the module derives clusters (based on stations' locations) and for each cluster estimates the convex hull. Then, it checks whether the area of the convex hulls is greater than a value configured into the tag "Hull_Area" (measured in Km²).

Once a valid convex hull area is identified, the SUSPICIOUS flag assigned to all stations within the convex hull is removed.

To keep track of those stations, the list of stations for which the flag was removed is saved into the module log file.

3.6 Quality checks flags management

Quality checks flags provide useful additional information. Quackme produces the following quality flags:

- H (human change): the value of a property was modified by a person (e.g., meteorologist).
- W (wrong value)
- S (suspicious value)
- I (interpolated value)
- R (replaced by MOS value). This flag can be assigned only to minimum and maximum temperature daily values when the MOS value replaces the calculated station value (Aggregation module).

Flags are produced first by the WeakChecks module at hourly level. The flags associated with the sub-daily values used in the aggregation determine the flag for the daily value. For example, if a daily value is derived by using nine hourly values having flags, the daily value will contain a combination of all these hourly flags separated by the character `|`.

The HeavyChecks manages the flags, and it is based on the flags generated by the Aggregation module. At this point the flag 'H' (if entered following changes done by a meteorologist) resets all the other flags associated with the value.

In the same way, the ThresholdChecks manages the flag starting with the flags generated by the HeavyChecks module. The flags file produced by the ThresholdChecks module (Flags.ThresholdChecks.<YYYYMMDD>.dat) represents the final set of flags for the elaboration date.

3.7 SConverter module

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

Structure of the S-File

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

Property description	Length	Note	Unit of Measure	Digits
TEMP_09	78:83	Air temperature late morning (09 UTC in Europe)	Celsius Degree	1
HUM_09	84:89	Relative humidity late morning (09 UTC in Europe)	%	0
TEMP_12	90:95	Air temperature midday (12 UTC in Europe)	Celsius Degree	1
HUM_12	96:101	Relative humidity midday (12 UTC in Europe)	%	0
TEMP_15	102:107	Air temperature afternoon (15 UTC in Europe)	Celsius Degree	1
HUM_15	108:113	Relative humidity afternoon (18 UTC in Europe)	%	0
TEMP_18	114:119	Air temperature evening (18 UTC in Europe)	Celsius Degree	1
HUM_18	120:125	Relative humidity evening (18 UTC in Europe)	%	0
STATE_SOIL	126:131	State of soil		1
VAP_PRES_Def	132:137	Daily mean vapour pressure deficit (mean over 24 hours)	hPa	1
SLOPE_VP_VS_T	138:143	daily mean slope saturation vapour pressure vs. temperature curve	hPa/Celsius Degree	3
CLOUD_DAYTIME_TOTAL	144:149	daytime mean of total cloud cover (mean over daytime hours)	Octas	1
CLOUD_DAYTIME_LOW	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
CLOUD_SHADOW	156:161	daytime mean amount of shadow clouds (mean over daytime hours)	Octas	1

Property description	Length	Note	Unit of Measure	Digits
	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
VISIBILITY	180:185	daytime mean visibility (daytime mean)	km	1
SNOW_DEPTH	186:191	Snow depth (instantaneous value 06 UTC in Europe)	cm	1
RAIN_MOS	192:197	indicator for reference period and usage of short range forecast in daily sum of precipitation RAIN		N.A.

The **SConverter** module ensures that the conversion from the JRC format to the S-File format respects the following convention name: *S<YYYYMMDD>.dat*, where YYYYMMDD represents the day of the input file (the day of the observations contained in the input file)

3.8 RRR file generator module

This module produces the 3h/6h precipitation data files. It can run only after the WeakChecks module because it needs to use the hourly history files.

As an input, the module takes

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

where both files are stored in the CONFIG path (specified through parameters into the command line).

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

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

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

Flags 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), precipitation values are derived as follows:

- 3-hourly interval:
 - o If for the specific time (09, 12, 15, 18, 21, 00 next day, 03 next day, 06 next day) all the observation from the previous 3 hours are not available (specific time -2 hours), the date is ignored
 - o If both RR & TR are correctly valued and RR = 3 then its value is used.
 - o If all the 3 observations (starting for the time under elaboration) in the interval [elaboration time – 2 hours: elaboration time] are available with valid values, then report the sum of these values. In case one value was interpolated, associate flag 4 else 3.
 - o If only one of the 3 observations is not available (with the other 2 are with valid values), then try to retrieve the missing value from MOS. If available and valid then report the sum of the two valid observations and the MOS value. In this case the associated flag is 5.
 - o Try to calculate 3H precipitation value if the RR/TR are both valid and TR is equal either to 6 or 12:

- If TR = 6, calculate the ratio of the precipitation using MOS 1H precipitation values as follows:
 - Ratio for the first 3-hour 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 the second 3-hour interval = (sum of 1H precipitation of the second 3 hours in the interval / sum of 1H precipitation of the 6 hours interval) * RR value
 - If TR = 12 apply the same logics of the previous point but the RR Value is now the difference between the current RR (the one with TR = 12) and the RR observation corresponding to 6 hours before the current time (if that one has TR = 6).
 - In both cases the flag is 5
- 6-hour interval:
- If for the specific times (12, 18, 00 next day 06 next day) observations do not exist, the station is ignored and NA is reported
 - For the observation row corresponding to the time in elaboration, check if the PR06 column contains a valid value and use it
 - Second option. If TR = 6, then use the value of RR if valid
 - Third option. Check if all the 6 observations (starting from the time in elaboration) in the interval [elaboration time – 5 hours: elaboration time] (e.g. [19:00 previous day – 00:00 current day]) are available:
 - Verify if TR = 1. If the RR values are all valid, then the SUM(RR) is used.

Else, if the PREC column contains valid values for all the 6 observations, then SUM(PREC) is used. If one of these values is interpolated, then the flag is set to **4**.
- Last step. In case the previous one does not produce a valid value, check the possibility of 12-hour precipitation split. Two cases are possible:
- When TR = 12 at the current elaboration time, we consider the MOS ratio of precipitation using MOS values (RR) at the current elaboration time and at the current elaboration time - 6 hours. The reported value is calculated as follows:
 - The MOS ratio = RR MOS value at the same time of the observed data in elaboration / (sum of RR MOS values found at the current elaboration time and current elaboration time – 6 hours).
 - The reported value = RR of observed data * MOS ratio
 - If TR = NA at the current elaboration time, we use the MOS ratio of precipitation using MOS values (RR) at the current elaboration time and at the current elaboration time + 6 hours, only if the observation corresponding to +6 hours has TR = 12 and a valid RR. The reported value is calculated as follows:
 - The MOS ratio = the RR MOS value at the same time of the observed data in elaboration / (sum of RR MOS values at the r current elaboration time and at current elaboration time + 6 hours).
 - The reported value = RR of observed data * MOS ratio
 - When the MOS data are used, the reported flag is 5.

The names of the output files adhere 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.

3.9 SG GUI

The GUI has been designed to enable meteorologists to correct the identified errors events, validate the flags and in case remove values.

Both input and output use the XML format.

The XML file has 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 XML tags are:

- TAG is an XML tag with the name of the property (for example: TT – air temperature, Snow – snow depth)
- Status can get the following values:
 - C – correct
 - W – wrong
 - S – suspicious
 - R – replaced
- Inside the ALERT, the LEVEL attribute provides the flag of the alert: W – wrong value, S – suspicious value, R – replaced (this value is used only for the HeavyChecks output)
- Property – contains the name of property (must exist 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 alert levels:

- S (suspicious)
- W (wrong)

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


- **M** (modified) the operator has inserted a new value different from the reported one
- **F** (forced) the operator has confirmed the reported value is correct
- **W** (wrong) the operator leaves unaltered the flag (W). In this case, during the subsequent run, the value is replaced with NA
- **S** (suspicious) the operator leaves unaltered the original status (S)

Show entries Search:

File name	Checks level	Reference date	Actions
O.KO.WeakChecks.20171102.xml	WeakChecks	20171102	Edit
O.KO.WeakChecks.20171102Backup.xml	WeakChecks	20171102Backup	Edit

Showing 1 to 2 of 2 entries Previous Next

Europe meteo data checks



Kalmar Flygplats (Sweden) [Edit](#)

Show entries Search:

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 Previous 2 3 Next

Showing 1 to 2 of 2 entries

Show10▼entries

Search:

File name	Checks level	Reference date	Actions
O.KO.WeakChecks.20171102.xml	WeakChecks	20171102	✎
O.KO.WeakChecks.20171102Backup.xml	WeakChecks	20171102Backup	✎

Previous1Next



Alajar (Spain)

creato file: ./Lock/_8380_201803090811.lock
[Save](#) [Cancel](#)

Show10▼entries

Search:

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

Showing 1 to 10 of 21 entries

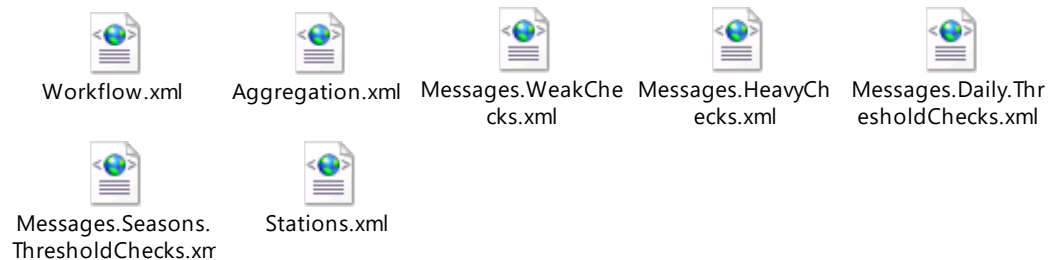
Previous123Next

4. QUACKME Configurations

The QUACKME software contains configuration files associated to each level of execution:

- A. Weak Checks
 - One file to configure messages in the current step (Messages.WeakChecks.xml)
 - One file to configure the routines to execute at each step (Workflow.xml)
 - One file that contains stations' data (Stations.xml)
- B. Aggregation
 - One file to configure the properties and associated formulas to be generated aggregated properties (Aggregation.xml)
- C. Heavy Checks
 - One file to configure the messages in the current step (Messages.HeavyChecks.xml)
 - Use of Aggregation configuration file (Aggregation.xml)
- D. Threshold Checks
 - Two files to configure the messages at each step:
 - 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 for every day of the year
 - ii. Threshold.Seasons.YYYY.dat, which contains the seasonal threshold data relative to the past year.
 - Use of Aggregation or HeavyChecks output: file of the day under analysis.
 - Use of history file (produced at the Aggregation step)

All configuration files have XML format, except for the threshold file having text format.



4.1 Messages structure

The "Messages.xml" file contains a list of MESSAGE tags and it has 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 (a sequential number related to each property)
- "Text" contains the full description of the error
- "Level" contains the warning level: A – alert, S – suspicious, W – wrong value.
- "Values" (only for Messages.HeavyChecks.xml, Messages.Daily.ThresholdChecks.xml, Messages.Seasons.ThresholdChecks.xml) contains a list of property to write (like <PropertyName>=<PropertyValue>) into the output KO XML file.

The text message can contain dynamic values such as #0#, #1# ... #n# that must be transferred inside the source code. The #0# will be replaced by the first value received, the #1# by 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 merges the original input data and the events corrected by the operator through the QUACKME GUI. The following rules apply:

- If a property is reported and corrected by the operator with status **F** then it is considered correct
- If a property is reported and corrected by the operator with status **M** then the new value (assigned by the operator) is moved to the input data and considered correct
- If a property is reported with the status **W** (wrong) and left unchanged by the operator, then its value is discarded and replaced by **NA** (NOT DEFINED)
- If a property is reported with the status **S** (suspicious) and left unchanged by the operator, then its value remains the same for the next run

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

- If the property is corrected by the operator with status **F** (forced) then the message is ignored
- If the property is corrected by the operator with status **M** (modified) the message is then reported again

4.3 Workflow configuration structure

The "Workflow.xml" is organized by levels: Weak Checks, Heavy Checks and so on. All levels 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 needs to be considered, OFF – the check module is 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