Automatic quality control of temperature and relative humidity observations from a high-resolution network composed of institutional and amateur weather stations

Francesco Uboldi (1), Andrea Zonato(1), Elena Oberto(1), Martina Lagasio(1), Alessandro Ceppi (2,3), Guido Cioni (2), Gianandrea Peressutti (2), Massimo Milelli (1)

(1) CIMA Foundation, Savona, Italy
(2) Associazione Meteonetwork ODV, Milano, Italy
(3) Department of Civil and Environmental Engineering (D.I.C.A.), Politecnico di Milano, Italy

Eumetnet Data Management Workshop, Oslo, 2025-11-05

www.cimafoundation.org

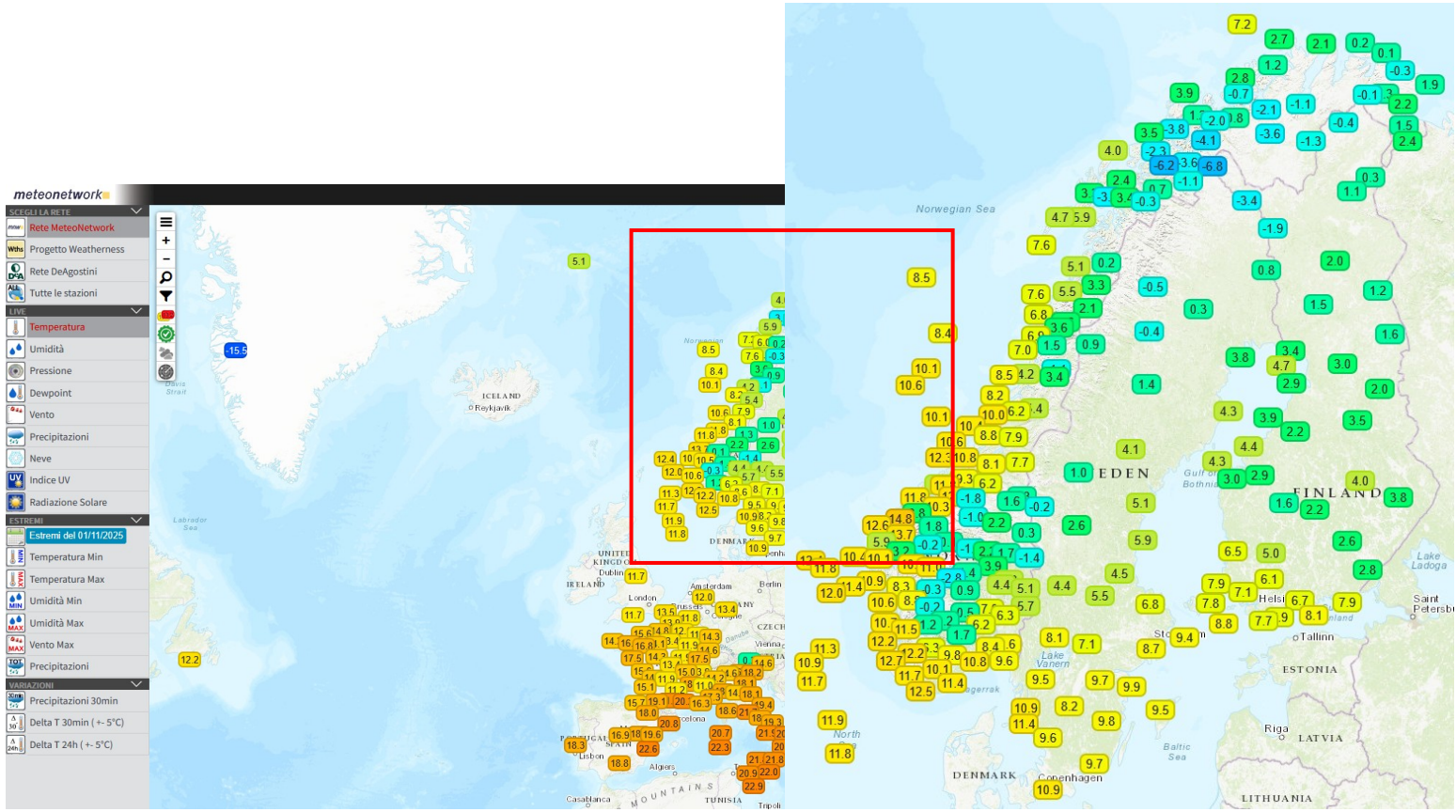
Observational network 1: **Civil Protection Department (DPC)**

- The observational network is a a nation-wide network, resulting from a **collection of several local networks** managed by local (regional) weather services
- It is composed of surface stations, providing hourly observations in near real time of **2-m temperature**, 2-m relative humidiy, 10-m wind, 1-hour accumulated **precipitation** and other variables
- Data are available through the MyDewetra portal: ***https://www.mydewetra.org/*** realised by CIMA for the Italian national Civil Protection Department
- Such network provides **detailed and valuable information** on meteorological fields near the surface
- For model verification, analysis of past events, climatological characterization of recent years up to present time and data assimilation
- There are errors in metadata: frequently on **elevation**, sometimes on **coordinates**.
- Observations are affected by **gross errors** and invevitably (but depending on their intended use) **large representativeness errors**

PROTEZIONE CIVILE

Observational network 2: **Meteonetwork (MNW)** *https://www.meteonetwork.it/*

- The Meteonetwork (MNW) association collects and makes available observations from stations owned (+4000), operated and maintained by private citizens, covering a wide territory with a high spatial density and high redundancy of measures (Giazzi et al., 2022).
- MNW stations are included in the MISTRAL portal (+2000): *https://www.mistralportal.it/*
- *and they are real-time availabe at:*
- *https://www.meteonetwork.it/rete/livemap/*
- MNW data are subjected to some automatic validation and quality control procedures before being made available.
- Supervision is carried out by scientists from its technical staff



Giazzi, M., Peressutti, G., Cerri, L., Fumi, M., Riva, I. F., Chini, A., Ferrari, G., Cioni, G., Franch, G., Tartari, G., Galbiati, F., Condemi, V., & Ceppi, A. (2022). Meteonetwork: An Open Crowdsourced Weather Data System. *Atmosphere*, *13*(6), 928. https://doi.org/10.3390/atmos13060928

Observational network: **Merge DPC and MNW**

DPC observations are professionally taken by local civil protection offices and operationally collected together by CIMA in real time.

MNW observations are taken by private amateurs and collected by Meteonetwork association; they undergo preliminary QC checks under the scientific supervision of internal data analysts.

Neither DPC nor MNW stations are WMO-compliant. Nevertheless they provide useful meteorological information at a high spatial resolution (1 to 10 km).

DPC and MNW are merged together in a unique dataset

DPC and MNW observations undergo the same QC checks

Stations are located in areas with various orographic and topographic characters (urban, rural, mountains, valleys, coastlines). For any intended use, **representativeness errors** are present.

Metadata

An important part of the work has been devoted to ensure quality of basic station metadata, such as geographical coordinates and orographic elevation above mean sea level.

Geographical coordinates:

- Sometimes coordinates are **missing**: these stations (their observations) are **discarded**
- Sometimes coordinates are simply **wrong**. For example, there are duplicates, such as:

  Stations with **same lon, lat** with **significantly different obs series**.

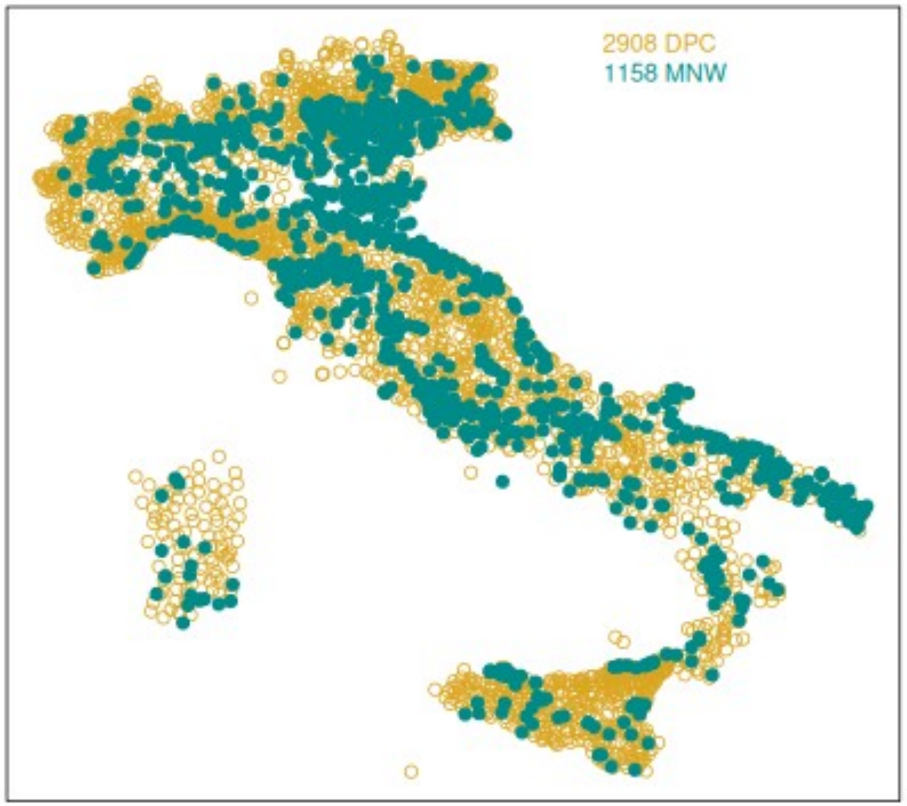
  Stations with **same obs series**, with **locations far apart**.

  Comparison with model forecasts and variability analysis: model forecasts have their errors, but variability is generally acceptable. The variability of differences between observations and model forecasts (estimated at station location) is effective in many cases to solve ambiguous cases.

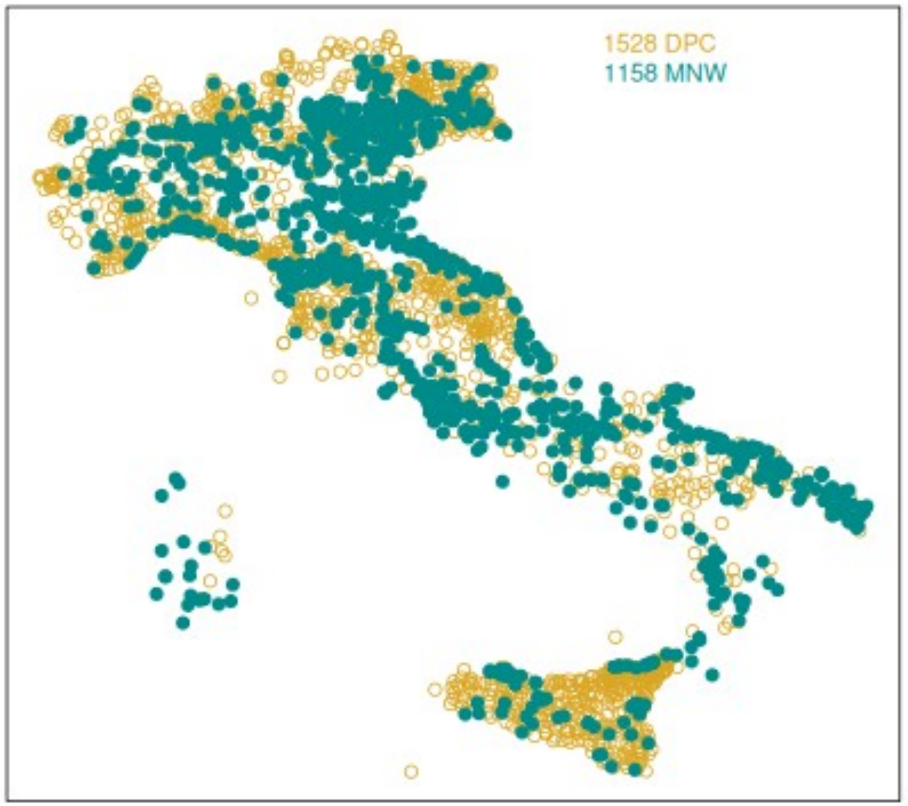
- **Elevation** is often missing or wrong: comparison with a high-resolution Digital Elevation Model and a Sea Mask

Stations active in August 2024

T2M stations August 2024 DPC + MNW

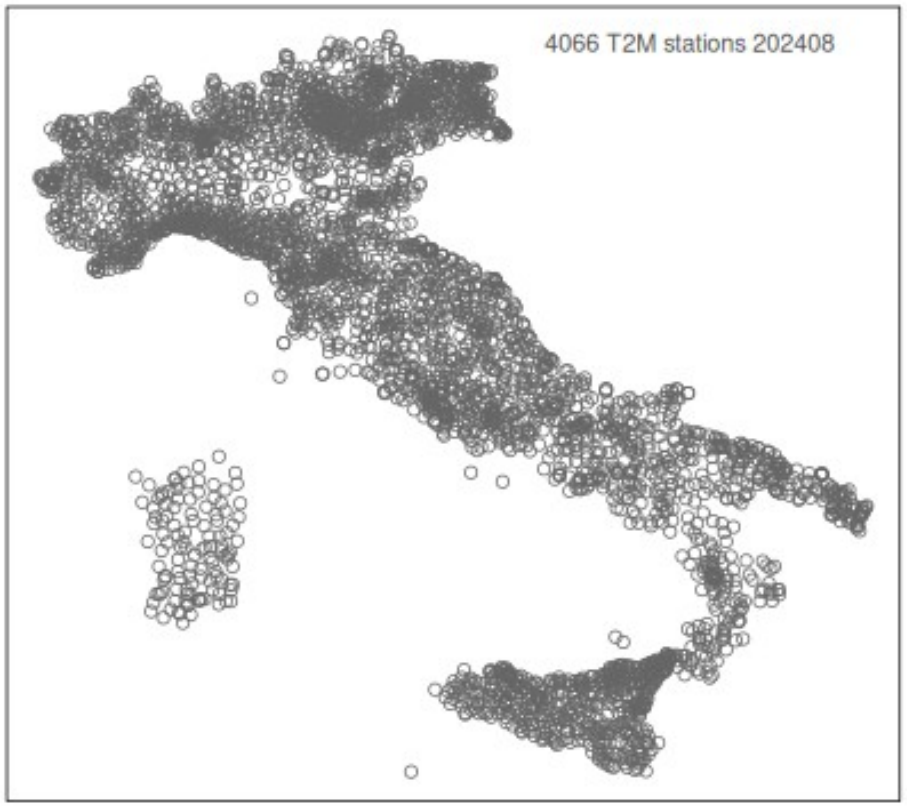


RH2M stations August 2024 DPC + MNW

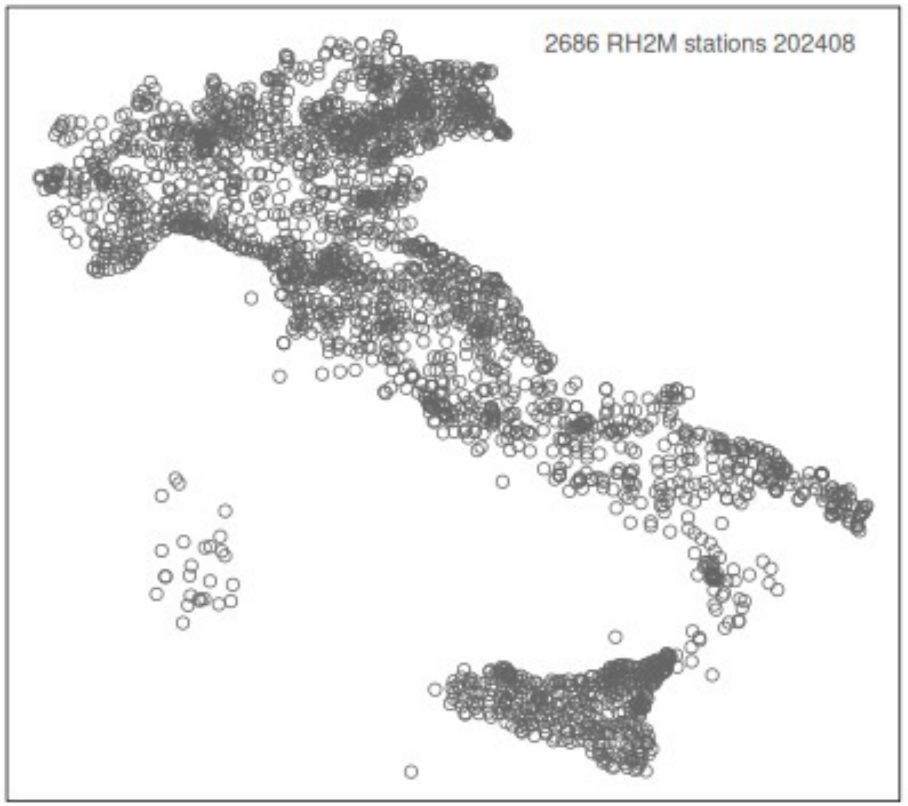


Stations active in August 2024

Nominally available T2M stations 2024-08-23 10:00

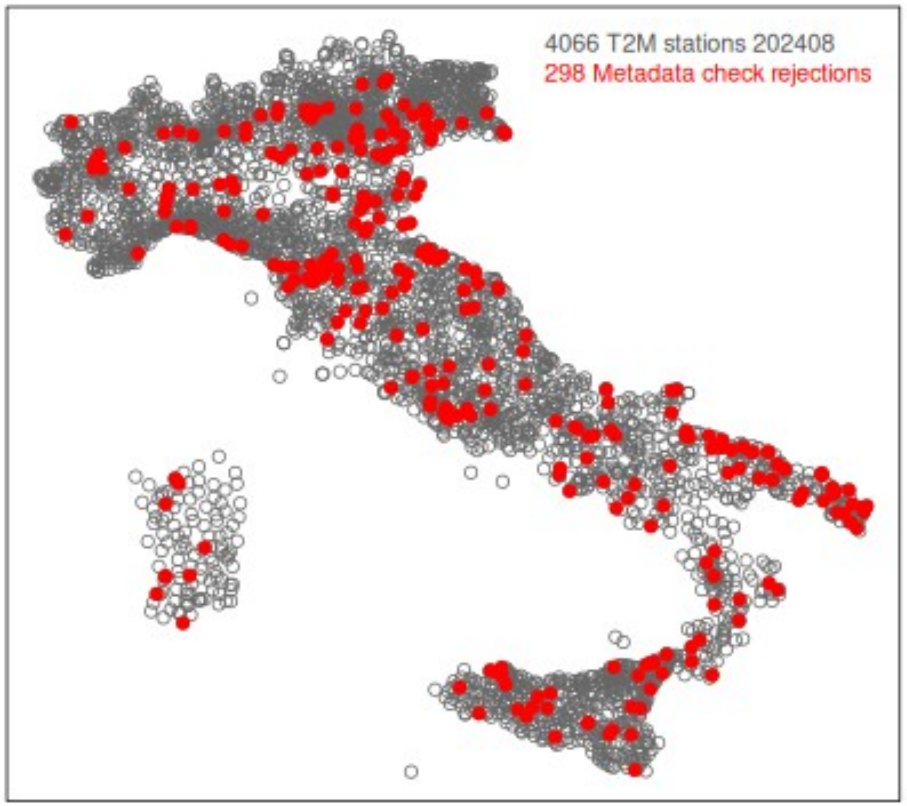


Nominally available RH2M stations 2024-08-23 10:00

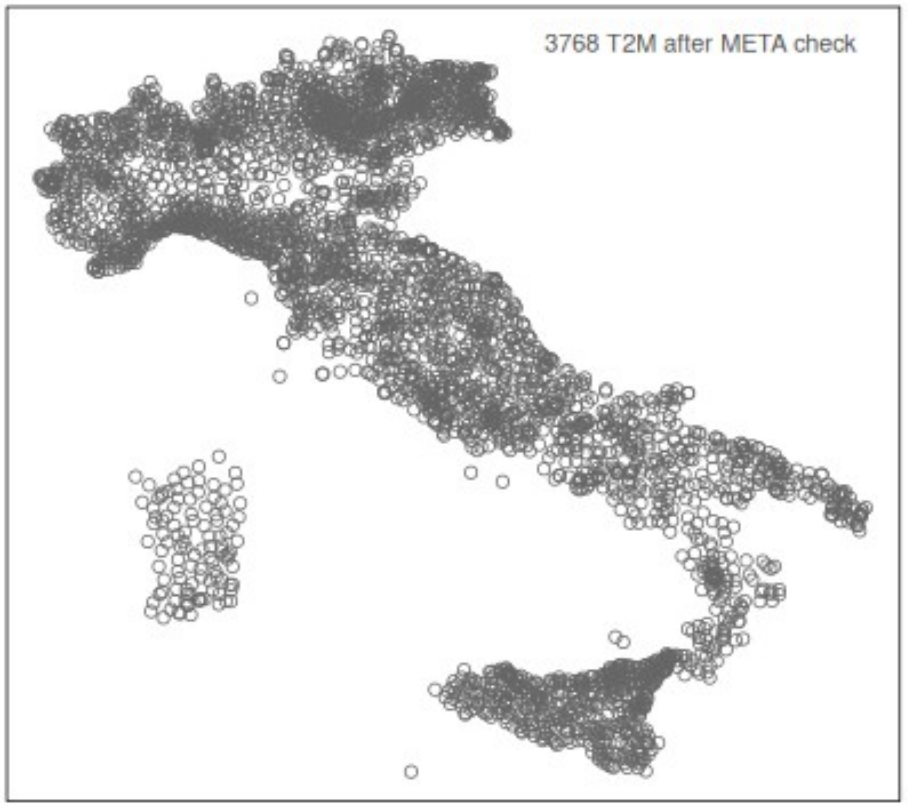


T2M stations: Metadata check rejections on 2024-08-23 at 10:00

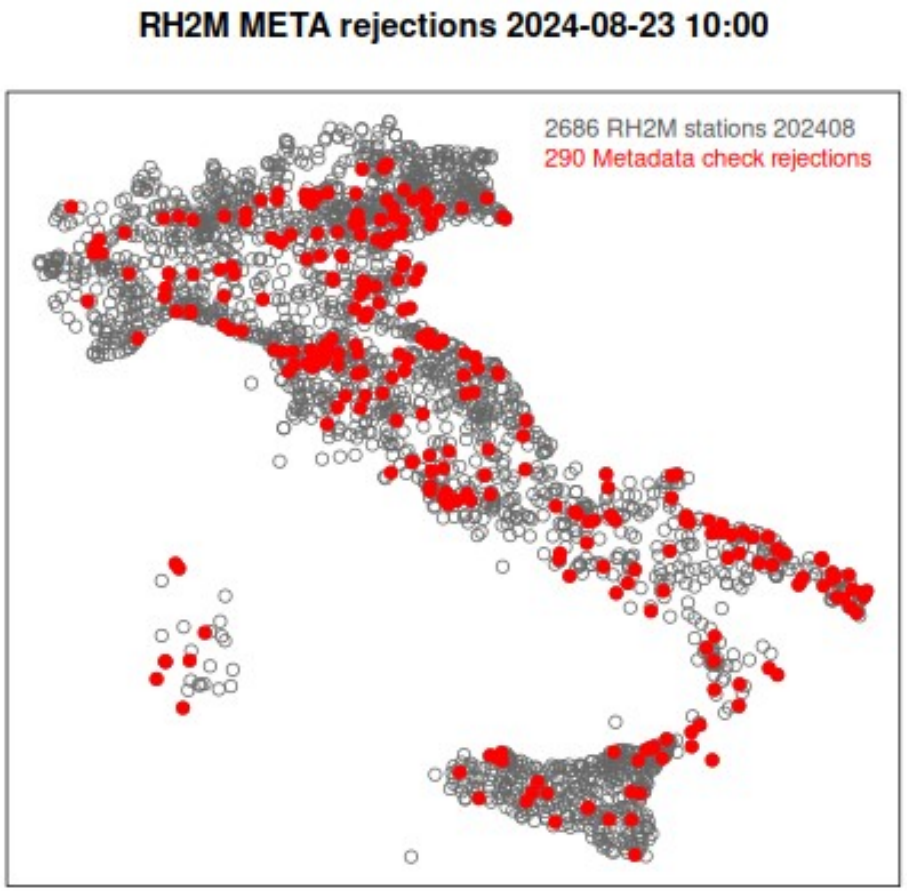
T2M META rejections 2024-08-23 10:00

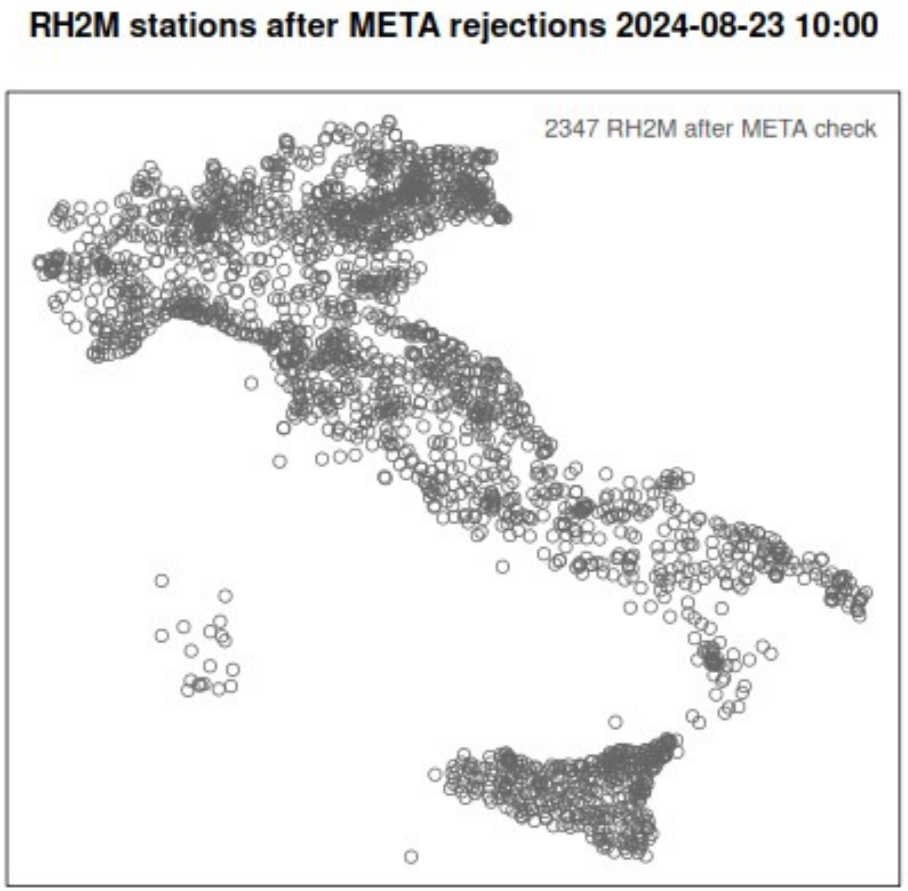


T2M stations after META rejections 2024-08-23 10:00



RH2M Metadata check rejections on 2024-08-23 at 10:00





Quality Control checks for **2-m temperature** and for **2-m relative humidity**

The main Quality Control check is the **Spatial Consistency Test** (SCT), introduced in:

Lussana, Uboldi and Salvati, 2010. A spatial consistency test for surface observations from mesoscale meteorological networks. *Q.J.R. Meteorol. Soc.*, **136**: 1075-1088, DOI:10.1002/qj.622

A formulation of the SCT is included in the open-source library **Titan** *https://github.com/metno/titanlib*

For 2-m temperature and relative humidity at CIMA the SCT was completely recoded in R scripting language to analyse the detailed behaviour of the checks.

The SCT is based on Optimal Interpolation (OI), which requires a Background Field (BF): a background estimate for QC checks is locally built from the data themselves.

The background estimate is also used as a preliminary comparison before the (finer) SCT: the **BF check**.

Quality Control checks for **2-m relative humidity**

- With few exceptions, all stations measuring relative humidity also measure temperature
- When a temperature observation is not available at the same station, the relative humidity observation is rejected (flagged)
- The **dew-point temperature** (T_d) is calculated from the observations of relative humidity and temperature

$$\mathrm{RH} = 100\frac{e_s(T_d)}{e_s(T)} \quad \Rightarrow \quad T_d = e_s^{-1}\left(\frac{\mathrm{RH}}{100}e_s(T)\right)$$

- The dew-point temperature “observation” undergoes the same QC checks as the temperature observation
- Parameter values of QC checks for dew-point temperature are different from values for temperature to account for the different spatial distribution of relative humidity observations:

  T: $D_h = 15$ km $D_v = 200$ m $\varepsilon^2 = 0.1$

  T_d: $D_h = 25$ km $D_v = 400$ m $\varepsilon^2 = 0.4$

Quality Control checks: **BF** – Background Field comparison

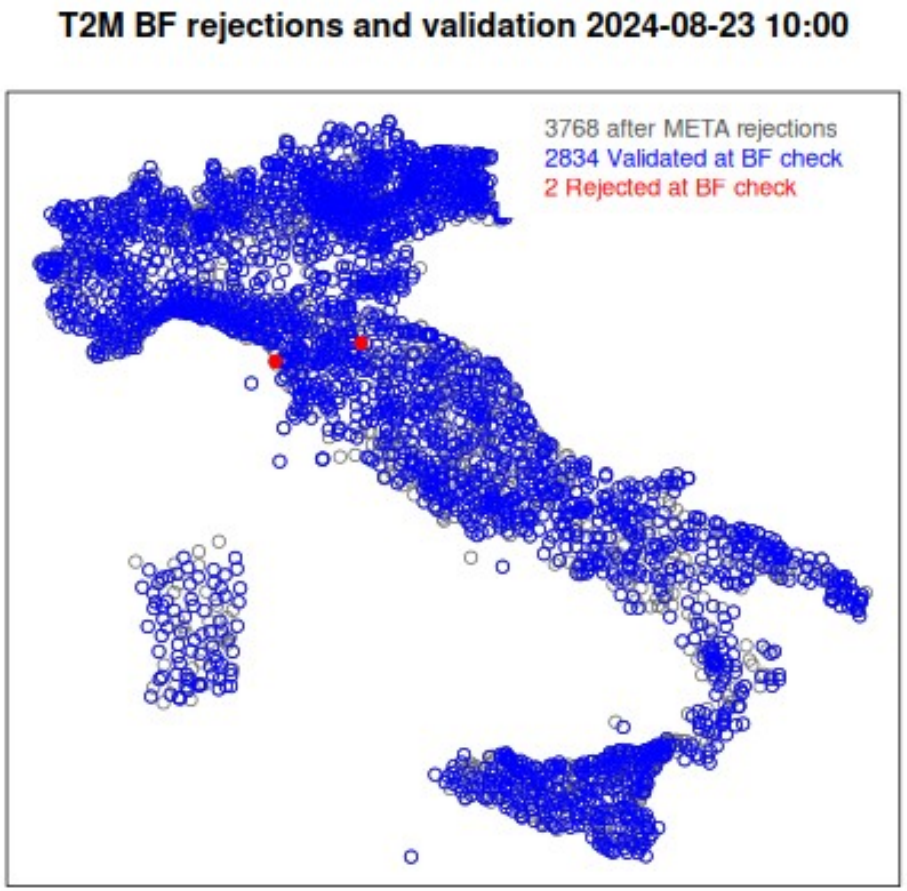
The Background Field is calculated at every obs location using neighbouring observations up to a maximum distance (30 km)

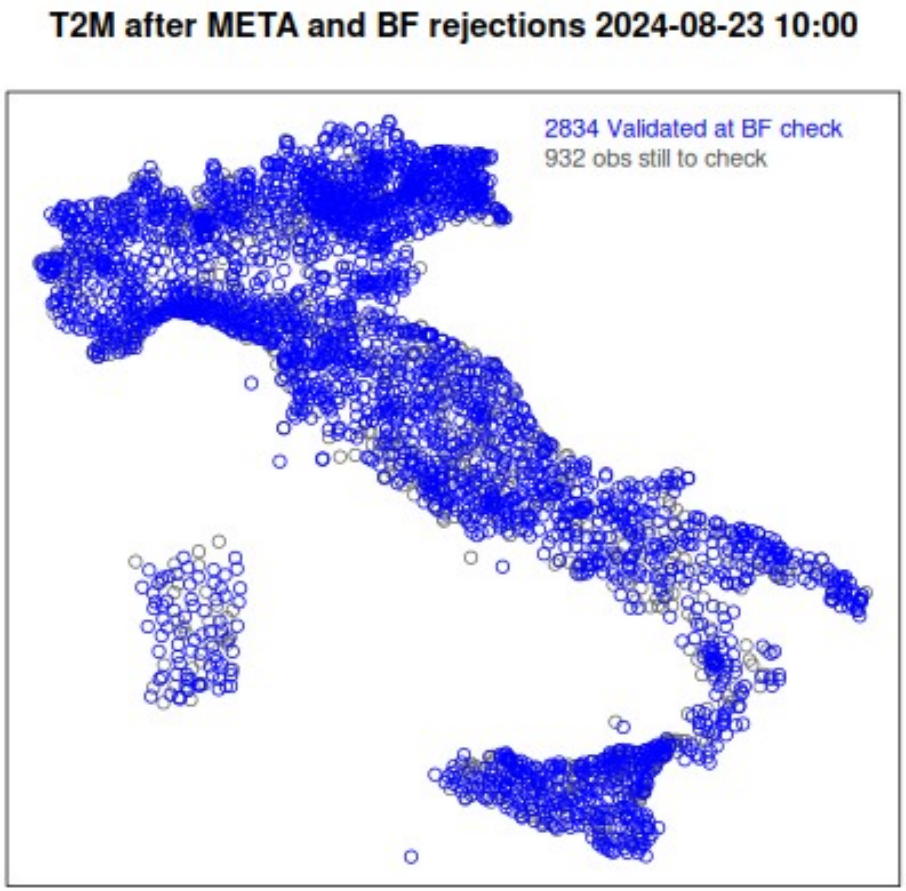
1) The **current observation** is undergoing the quality check and therefore **is not used** to compute the BF
2) A local **lapse rate** is least-square-estimated and used to "bring" neighbouring observed values to the current observation elevation
3) The **median** of these values (all referred to the same elevation) is taken as the background estimate

At the BF check **two thresholds** are used, a **large** one (T: 6°C, T_d: 12°C) and a **small** one (T,T_d: 1°C):

- Observation values **differing** from the BF value for more than the **large** threshold are **rejected**
- Observation values **close** to their BF value for less than the **small** threshold are **directly validated without undergoing the SCT**.

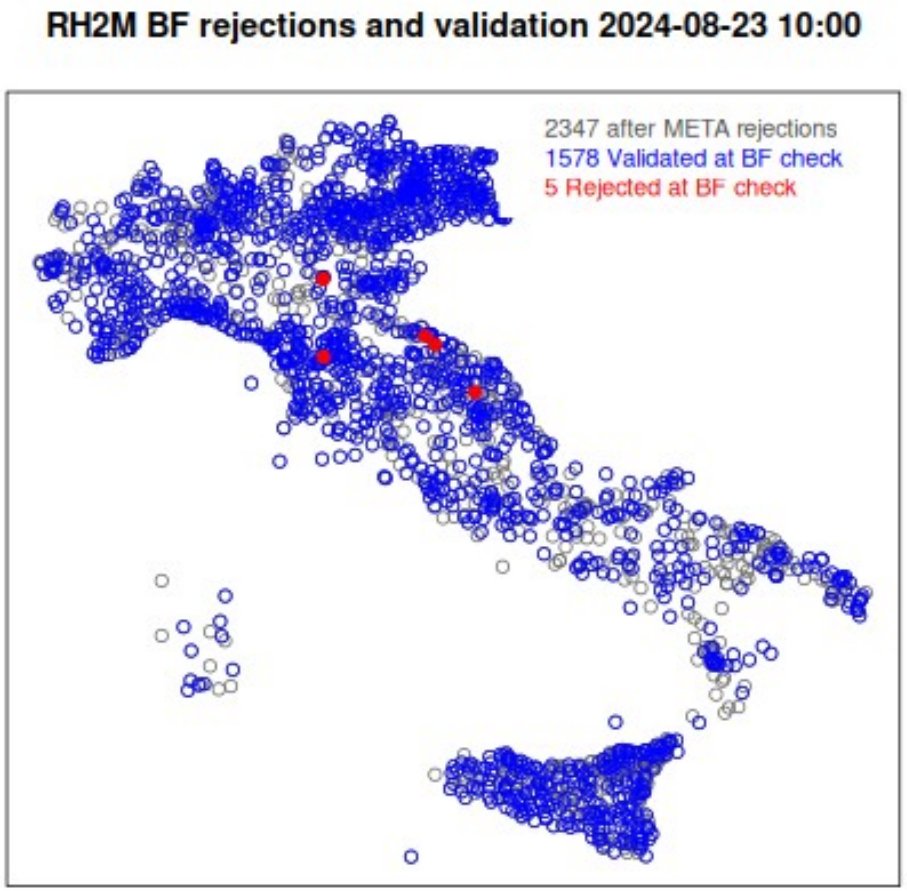
T2M BF check

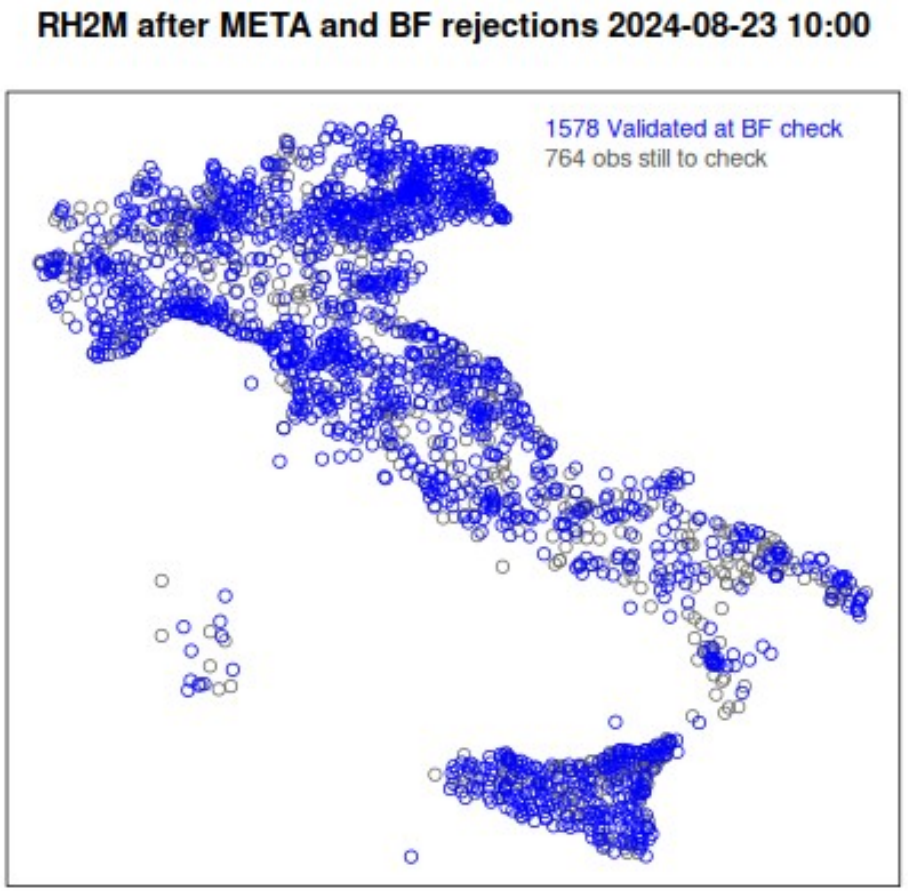




BF rejections and direct validation

RH2M BF check

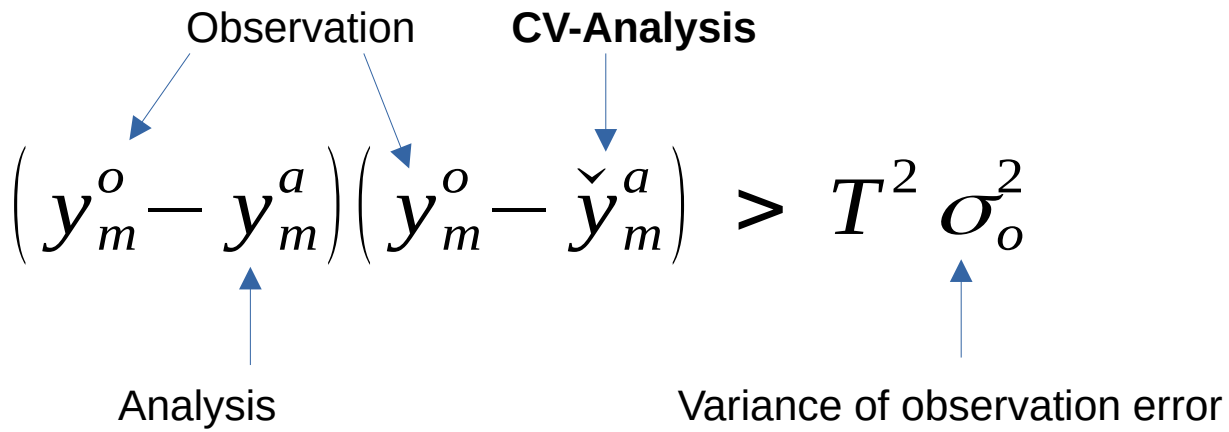




BF rejections and direct validation

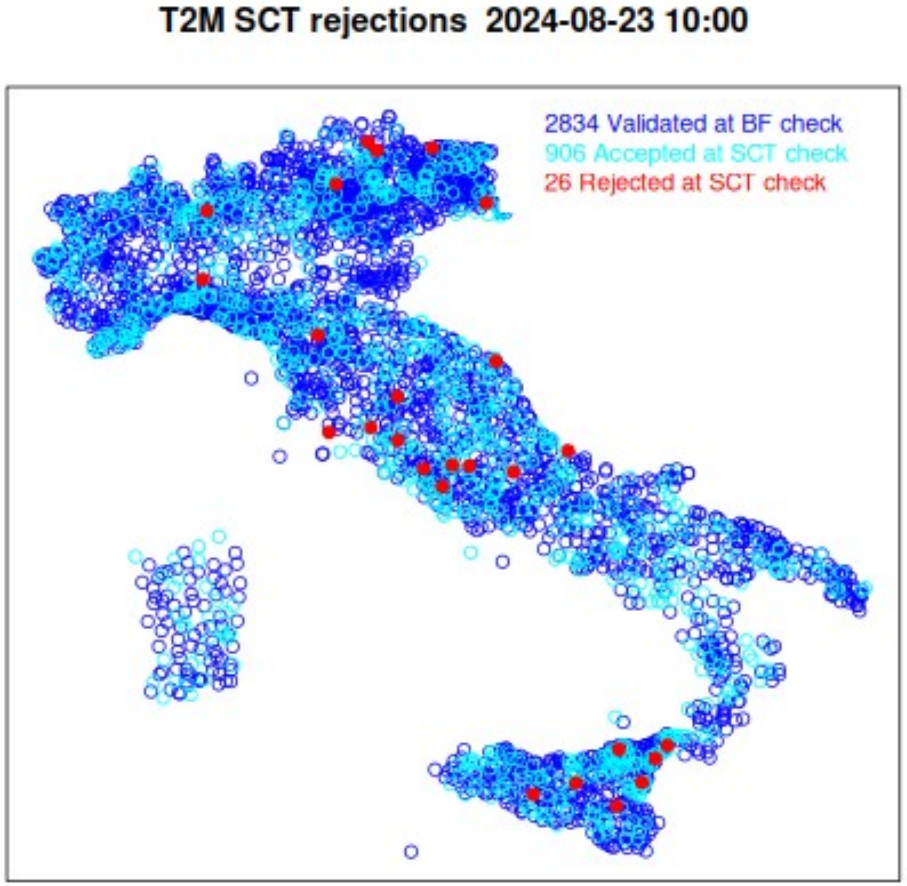
Quality Control checks: **SCT: Spatial Consistency Test**

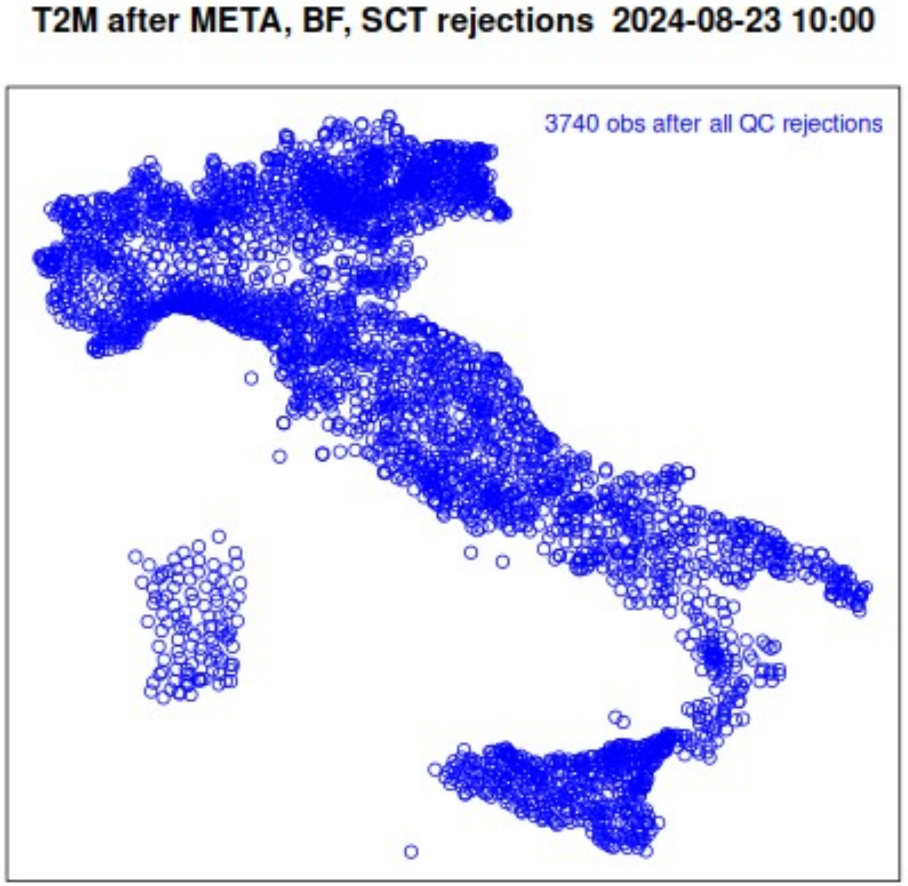
- Based on Optimal Interpolation (or 3D-Var)
- The **leave-one-out Cross Validation (CV) analysis** is the analysis estimated for the observation undergoing the quality check without using it, i.e. using all other observations only.
- The CV-analysis is a byproduct of an OI / 3D-Var analysis when the observational error covariance matrix is diagonal, i. e.: **the analysis calculations do not have to be repeated**



Lussana, Uboldi and Salvati, 2010. A spatial consistency test for surface observations from mesoscale meteorological networks. *Q.J.R. Meteorol. Soc.*, **136**: 1075-1088, DOI:10.1002/qj.622

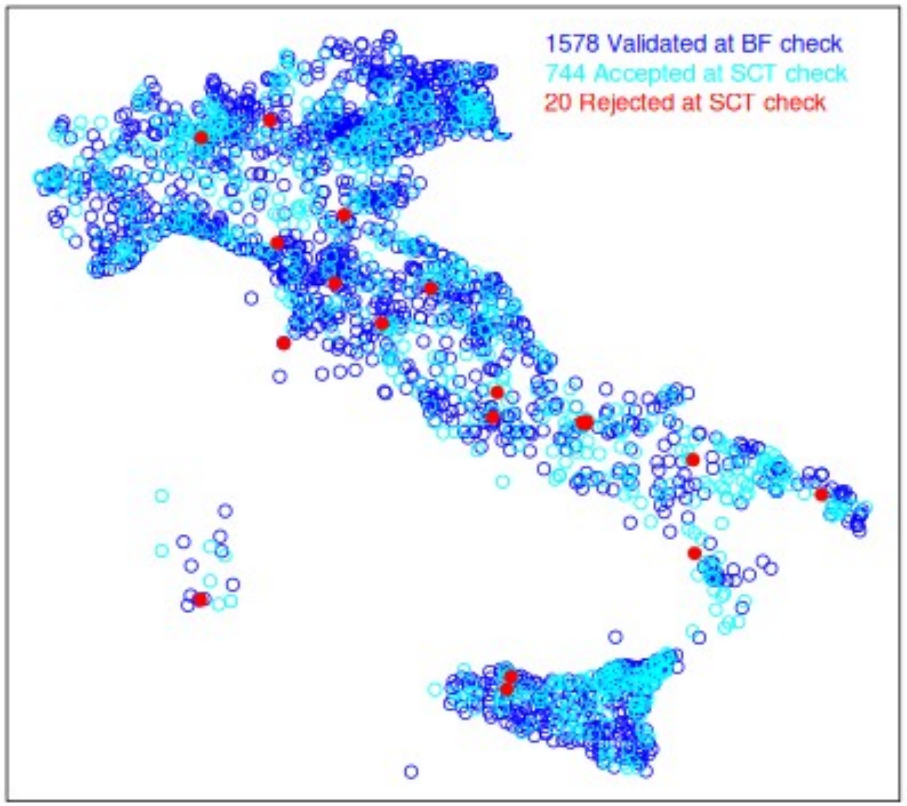
T2M SCT


T2M SCT rejections 2024-08-23 10:00
2834 Validated at BF check
906 Accepted at SCT check
26 Rejected at SCT check

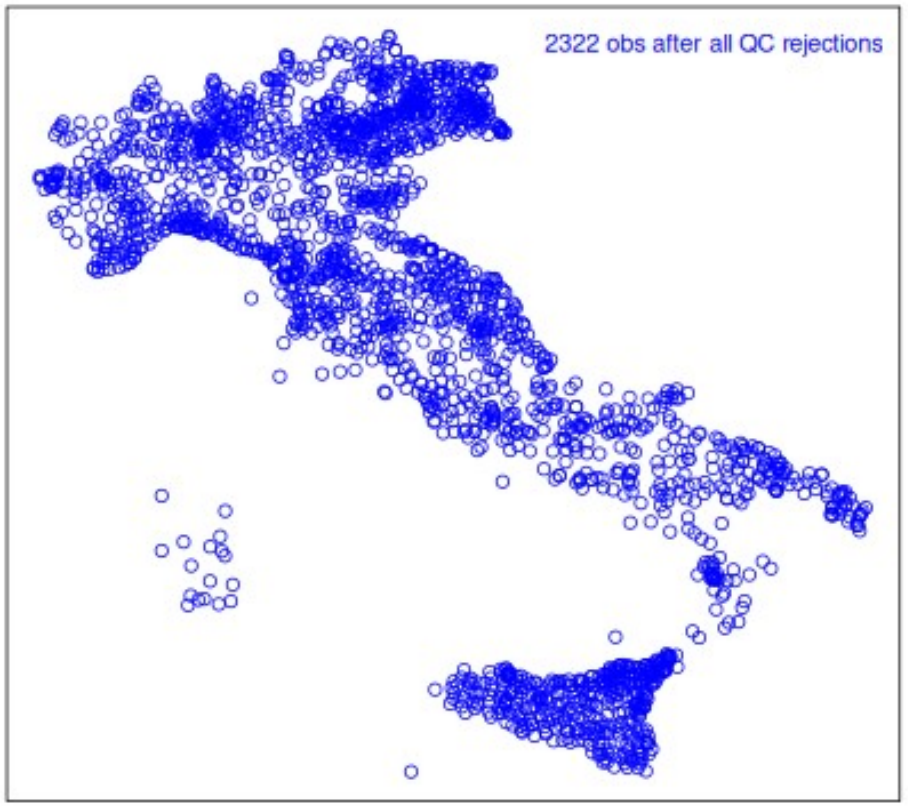

T2M after META, BF, SCT rejections 2024-08-23 10:00
3740 obs after all QC rejections

RH2M SCT

RH2M SCT rejections 2024-08-23 10:00

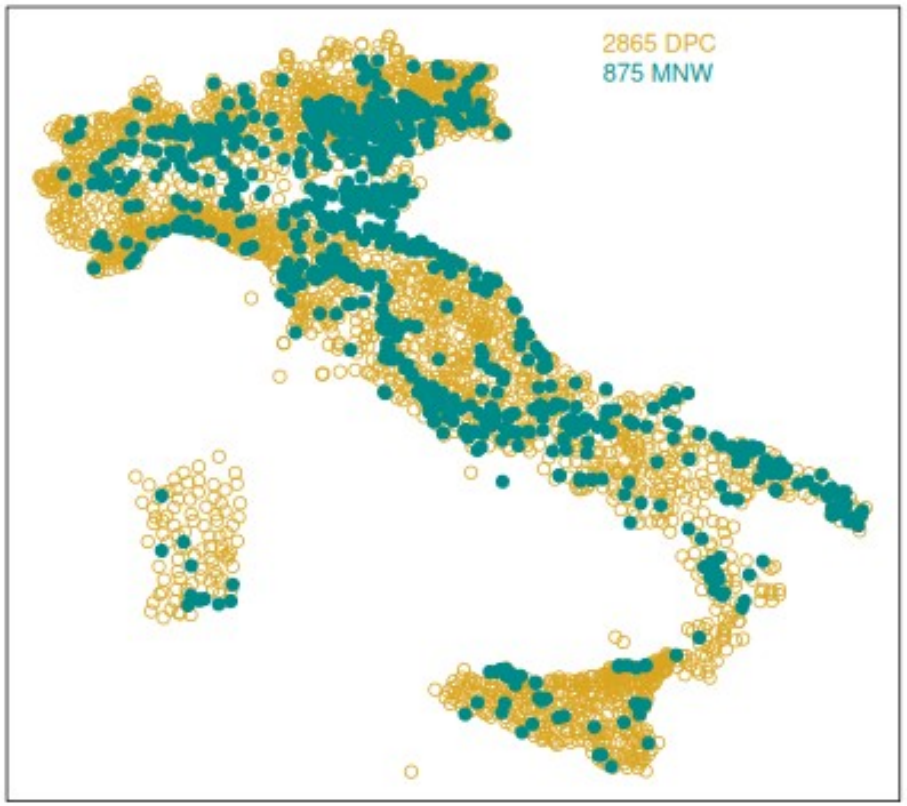


RH2M after META, BF, SCT rejections 2024-08-23 10:00

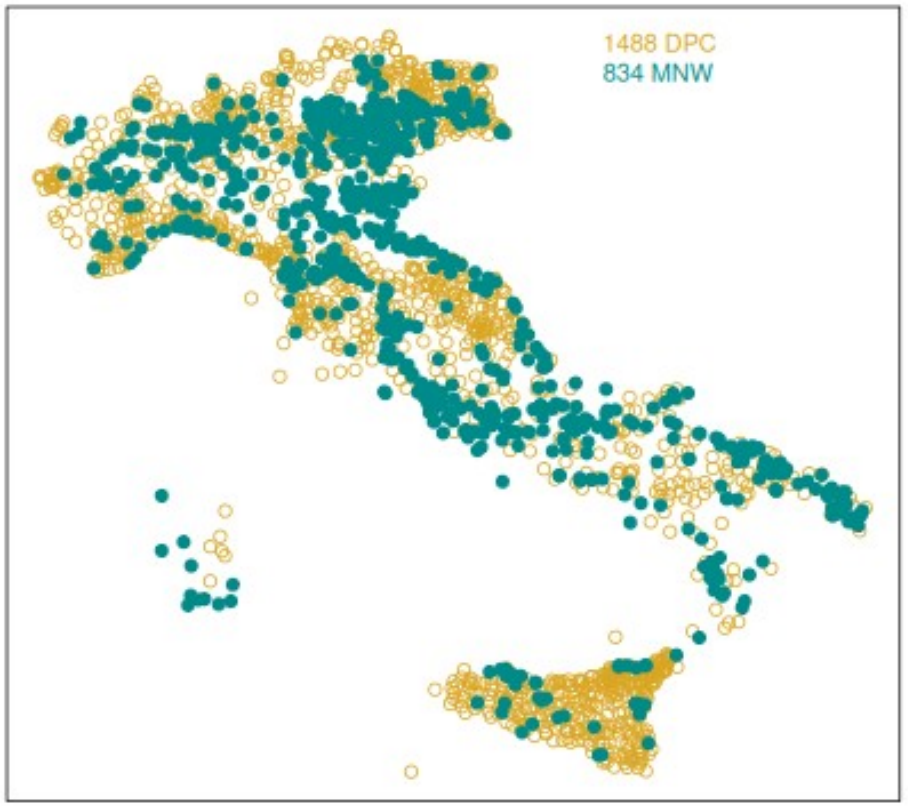


DPC + MNW after QC checks

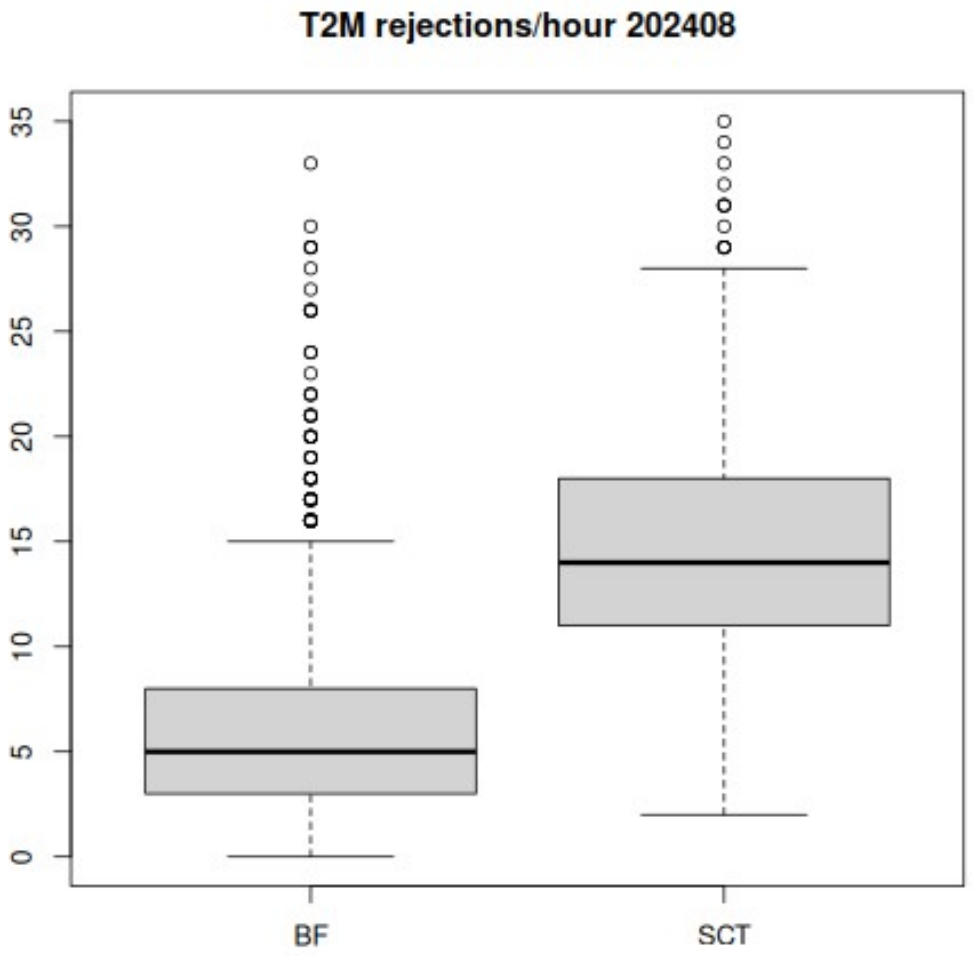
T2M after all QC rejections DPC + MNW 2024-08-23 10:00

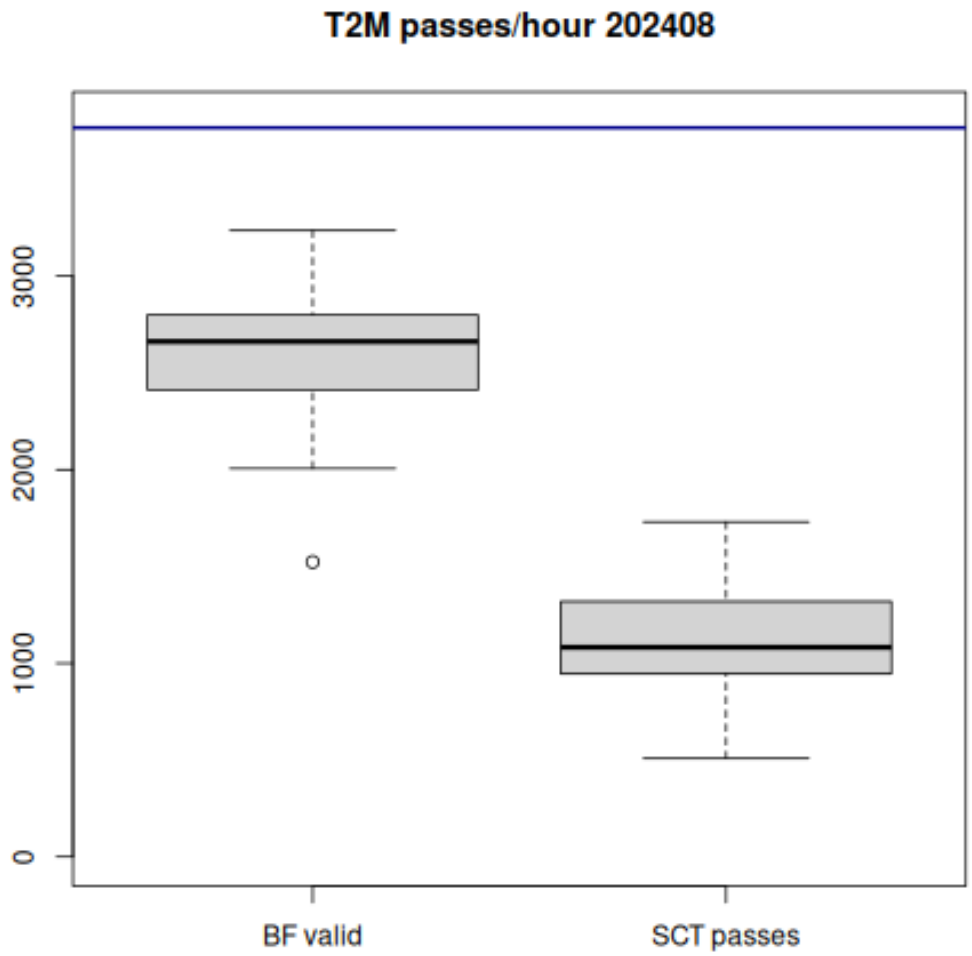


RH2M after all QC rejections DPC + MNW 2024-08-23 10:00

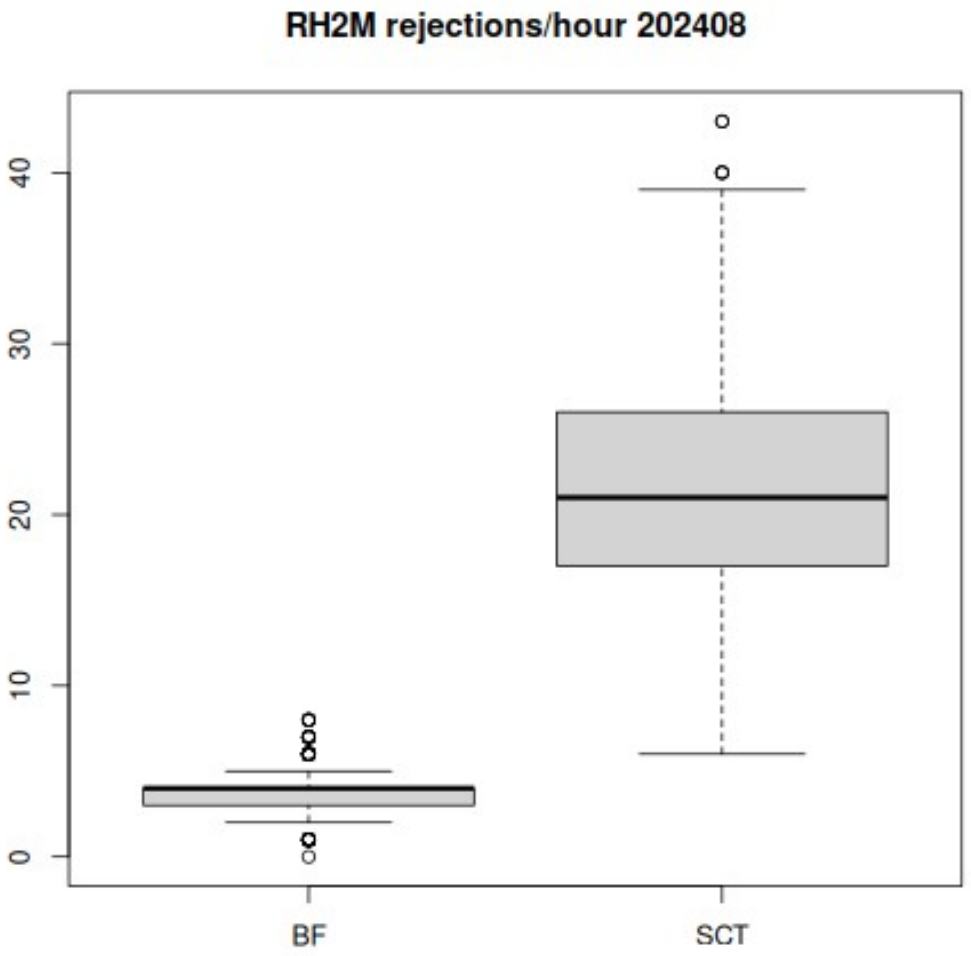


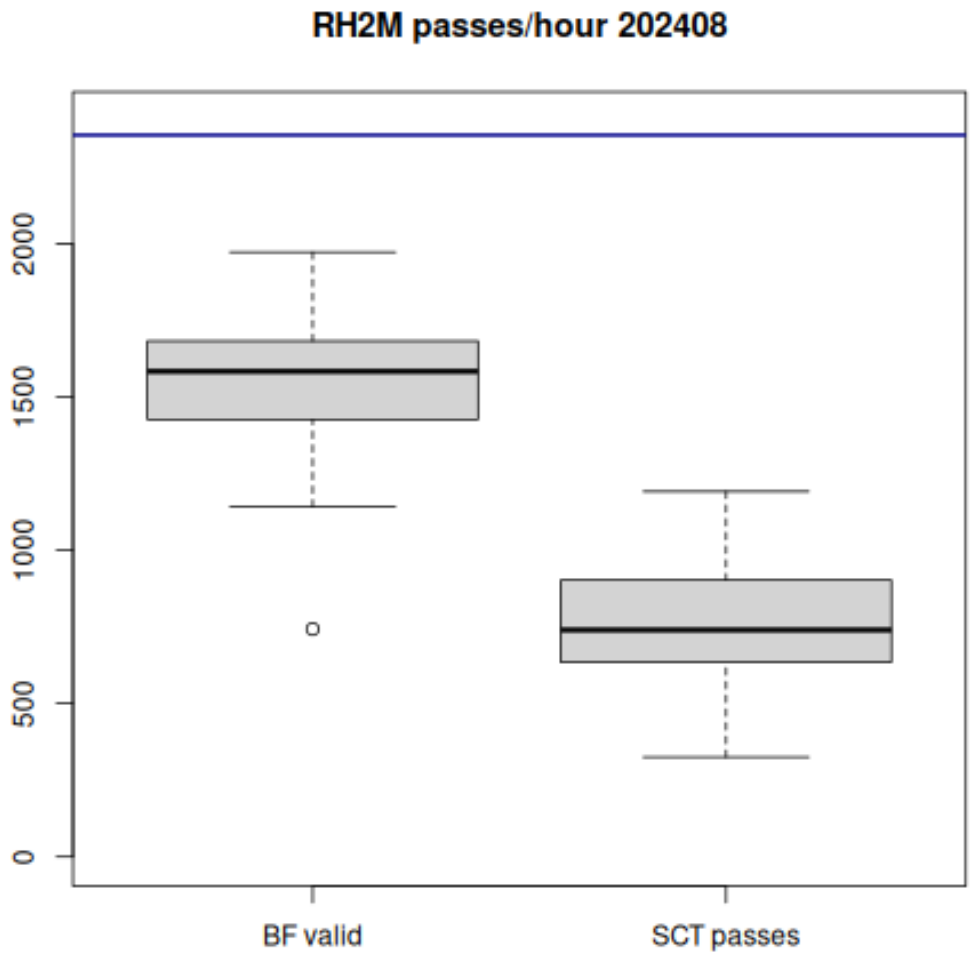
Number of observations rejected and accepted per hour. whole month **2024-08 T2M DPC + MNW**



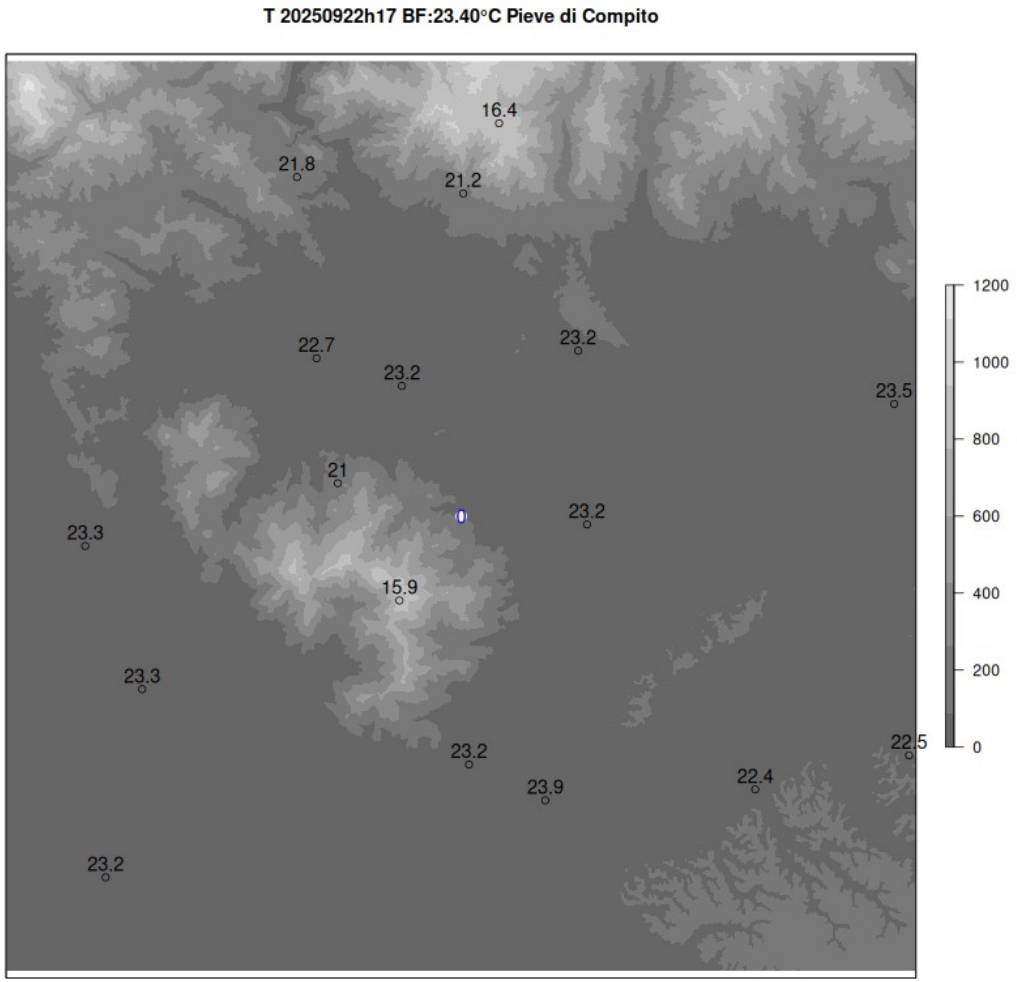


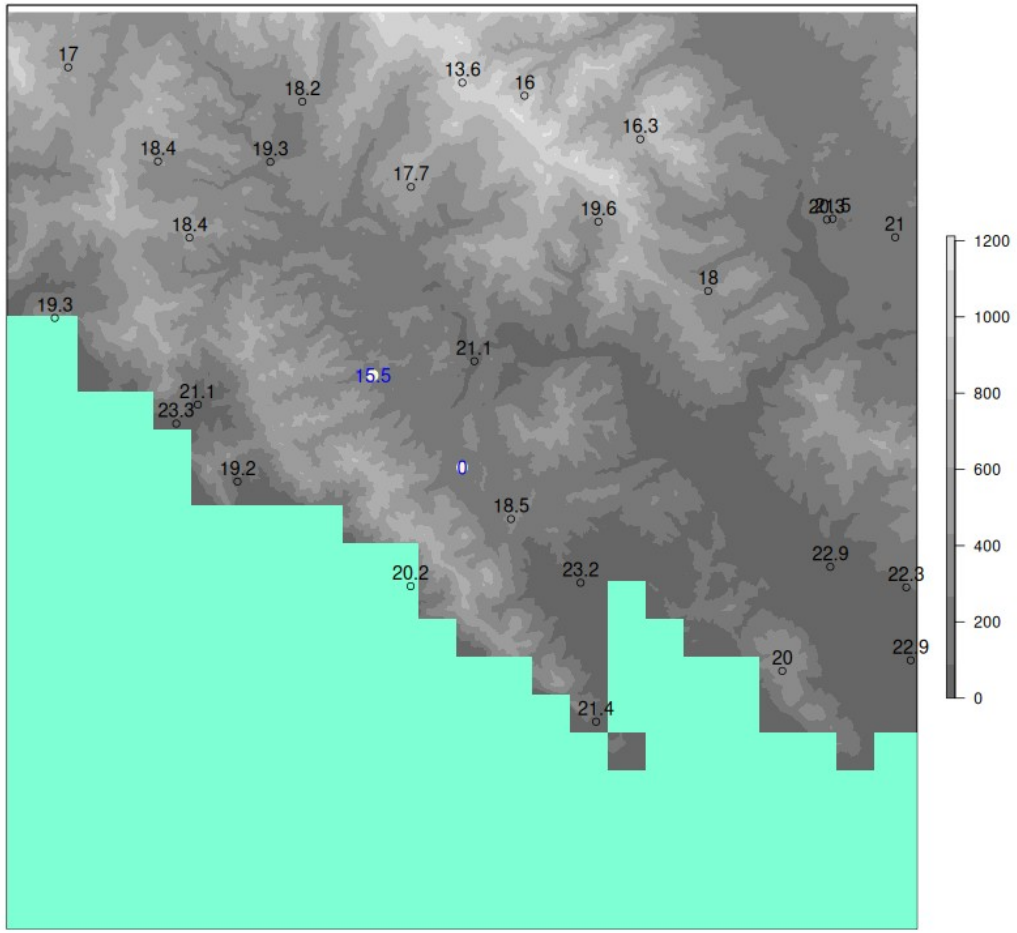
Number of observations rejected and accepted per hour. whole month **2024-08 RH2M DPC + MNW**



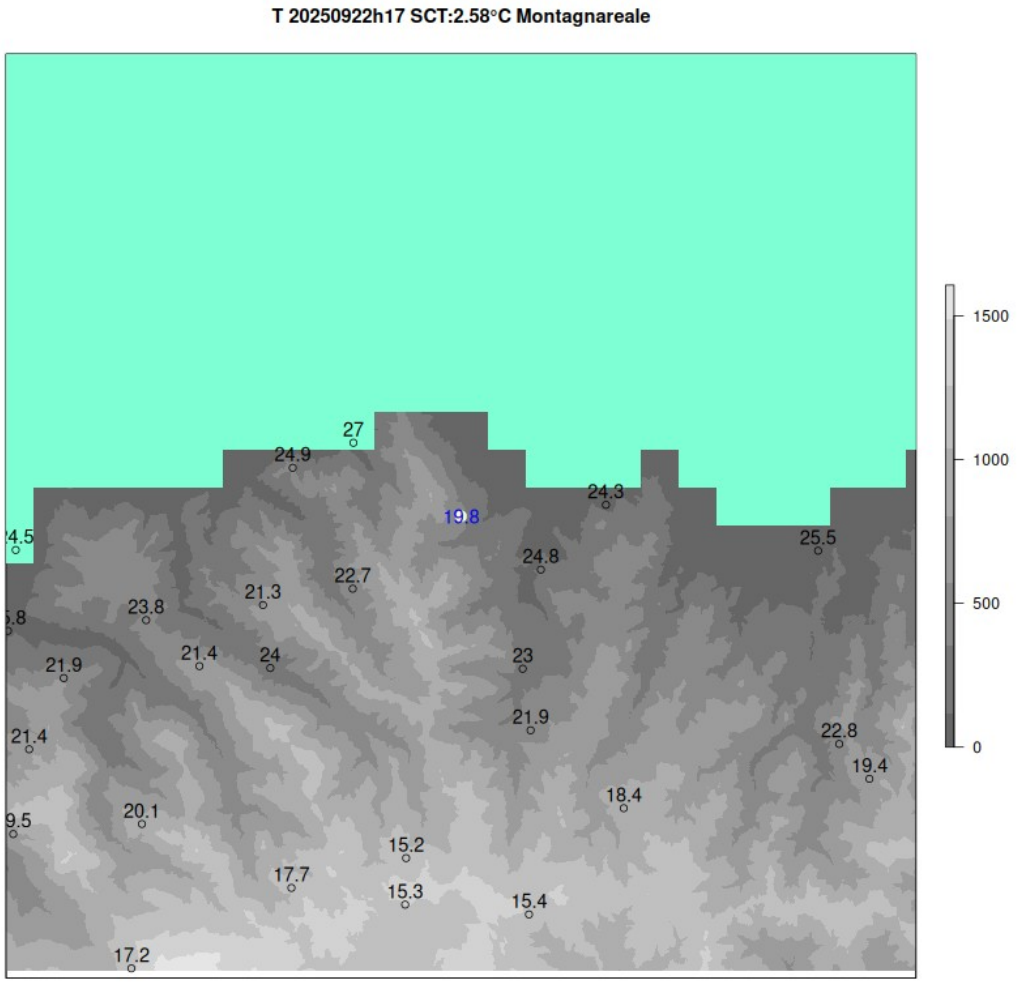


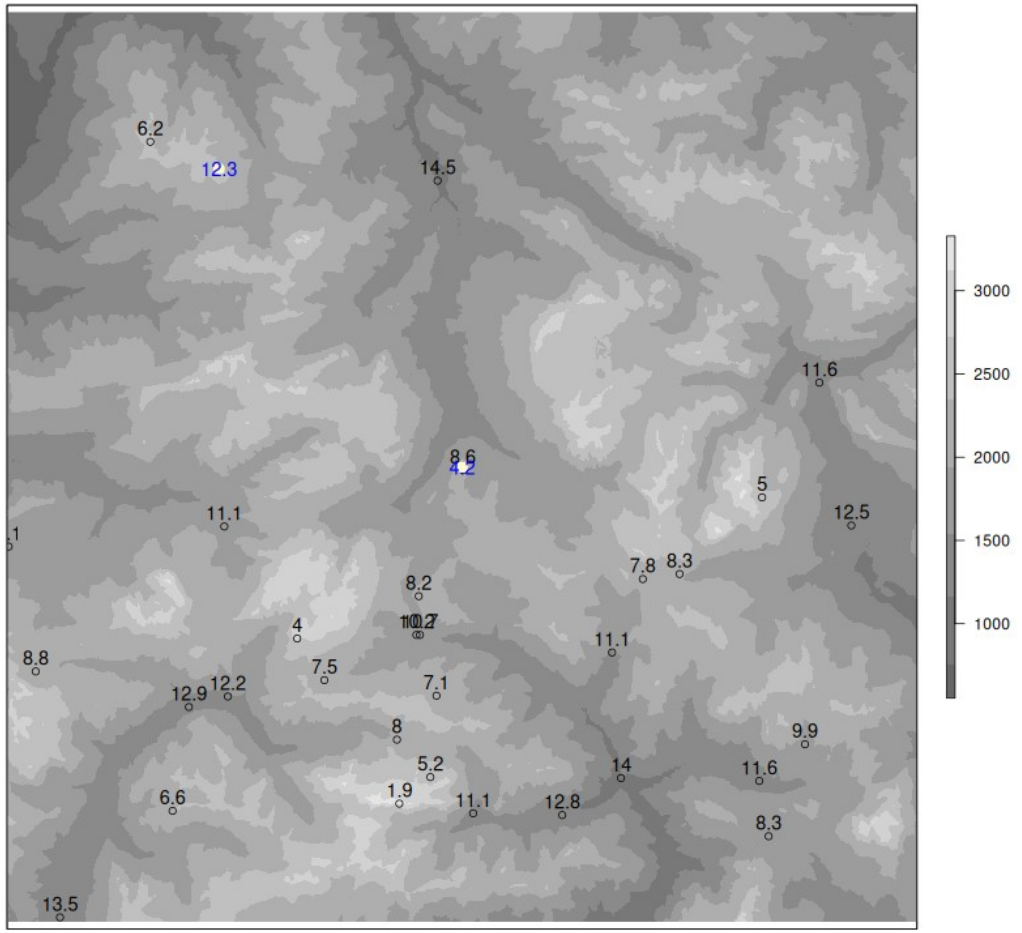
BF check: examples September 2025 DPC T2M



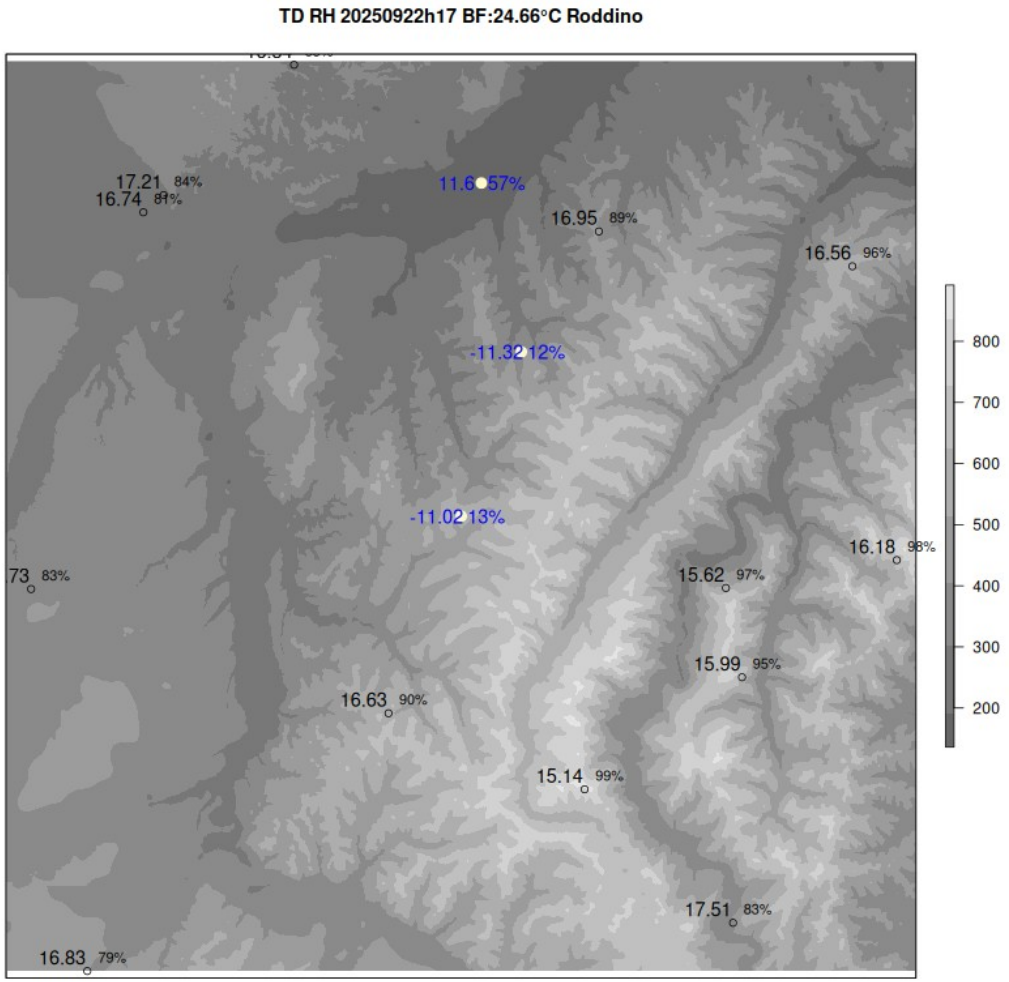


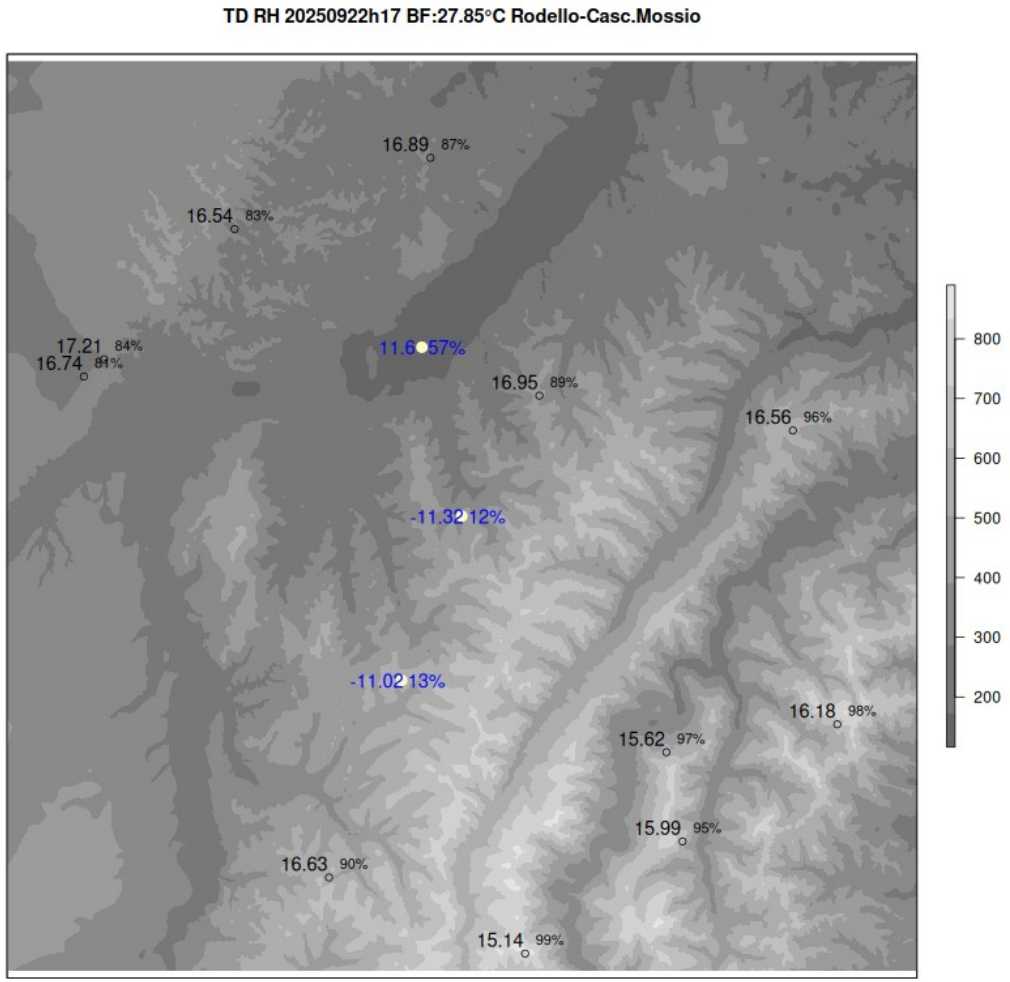
SCT: examples September 2025 DPC T2M



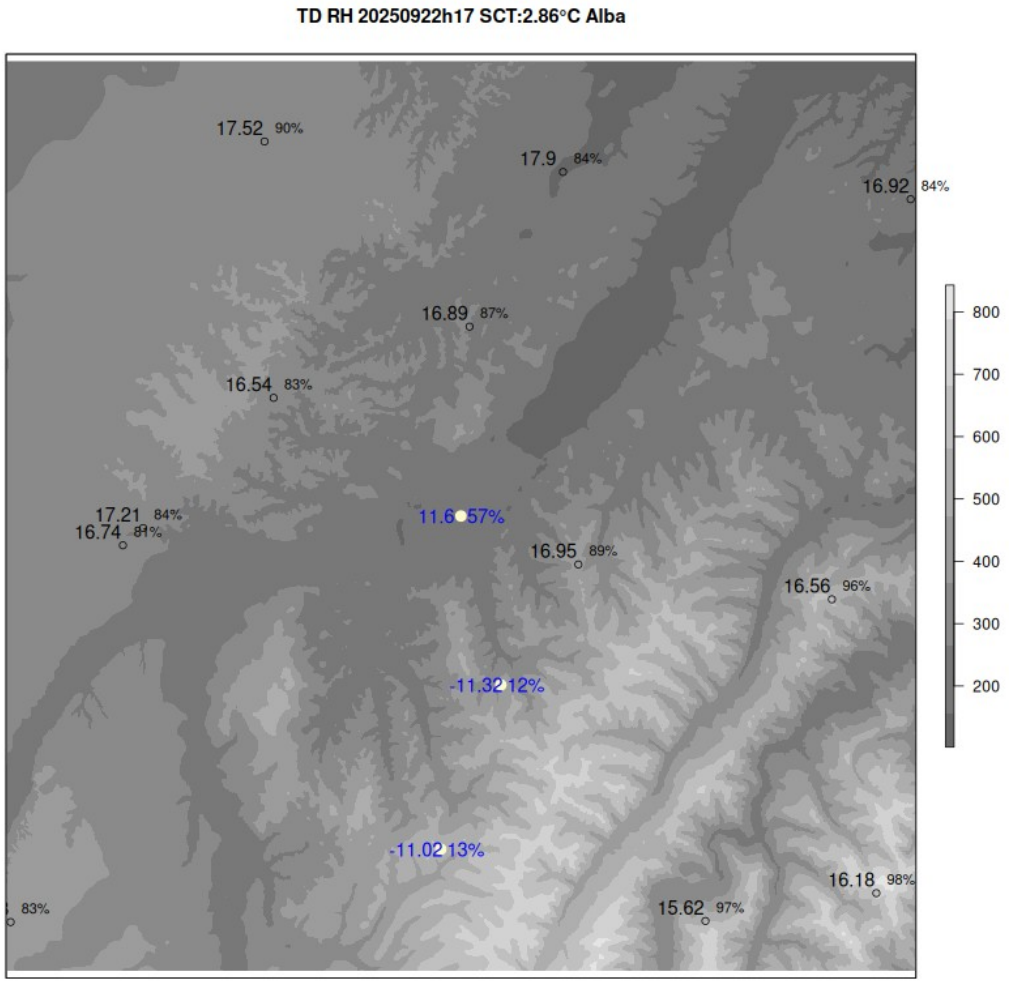


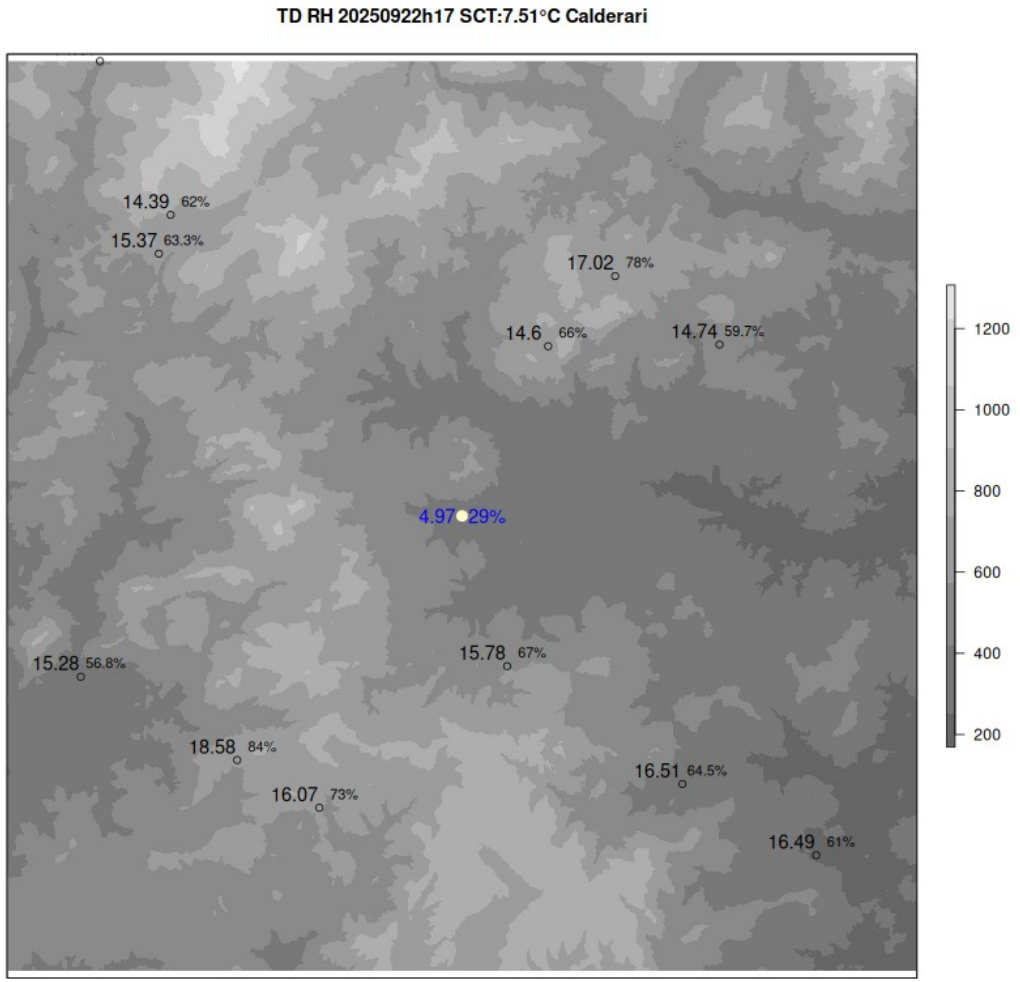
BF check: examples September 2025 DPC RH2M





SCT: examples September 2025 DPC RH2M





Assimilation experiment

- RAISE project context
- Whole month of August 2024 (project requirement)
- WRF convective-scale nowcasting system in a operational configuration: 2.5 km, including WRFDA 3d-Var assimilation of radar reflectivity and lightning, every 3 hours in a Rapid Update Cycle. 2 or 3 assimilations before 12-h forecast: 8 runs per day, 4 forecasts available every hour (mini-ensemble).
- Assimilate **DPC**+**MNW** 2-m temperature observations with WRFDA 3D-Var
- Control run CTRL: operational run real-time August 2024
- NOQC: T2M obs assimilated without any preliminary QC-check. Some QC-check are automatically performed by WRFDA: orography difference, background (forecast) state
- QC: BF+SCT based QC checks and Thinning are performed before the assimilation (WRFDA checks are also performed)
- **Thinning** is based on **IDI** and **CV-IDI**

Thinning based on IDI and CV-IDI

IDI: Integral Data Influence (Uboldi *et al.*, 2008): analysis obtained when all observed values are set to 1 and all background values are set to 0

$$y_m^{\mathrm{IDI}} = \sum_k \frac{\partial y_m^a}{\partial y_k^o}$$

IDI is about 0 at locations far from all observations, about 1 in densely observed areas

CV-IDI: $\check{y}_m^{\mathrm{IDI}}$ leave-one-out cross validation value of IDI.

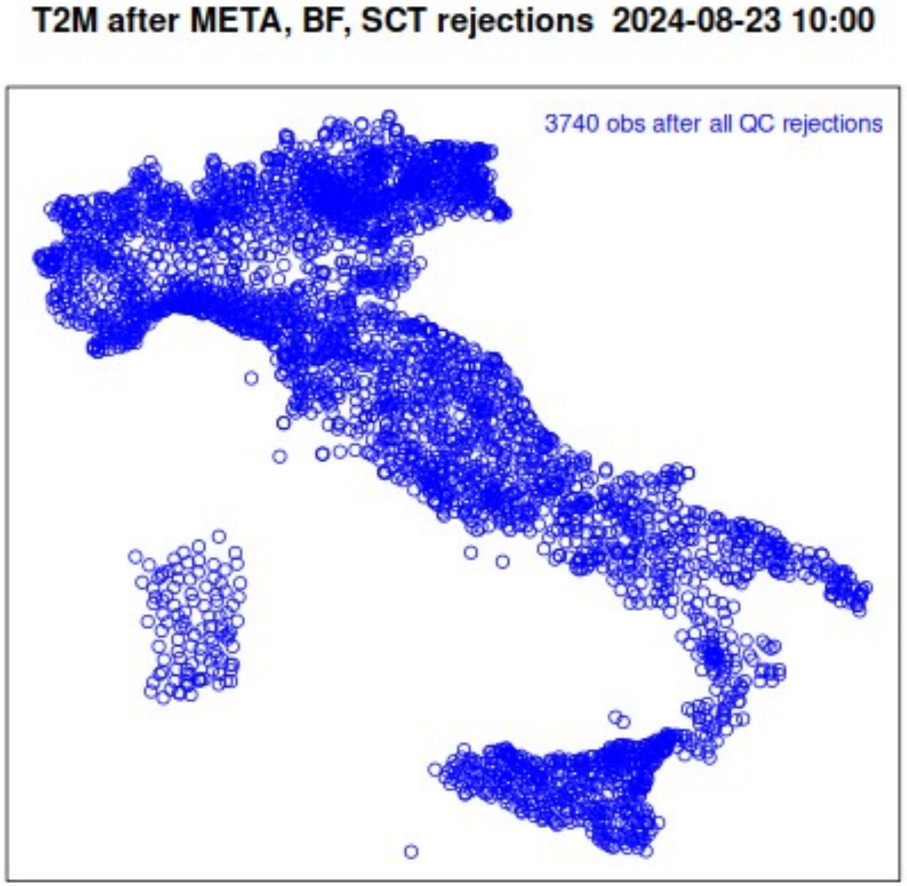
When (exactly) CV-IDI = IDI that station is totally redundant: CV-analysis = analysis regardless of observed values.

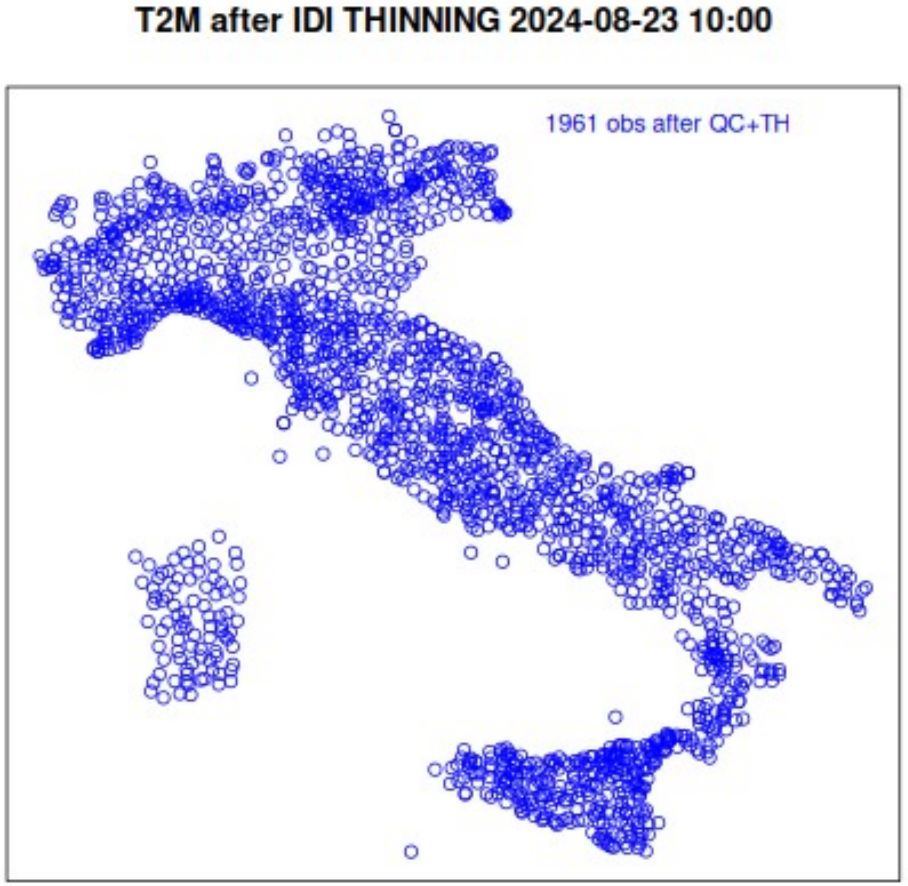
When $\frac{|y_m^{\mathrm{IDI}} - \check{y}_m^{\mathrm{IDI}}|}{y_m^{\mathrm{IDI}}} > \rho_{MIN}^{\mathrm{IDI}}$ the observation is considered “redundant” and discarded.

Only one rejection in each 5x5 (model grid) macro-box before re-calculating IDIs.

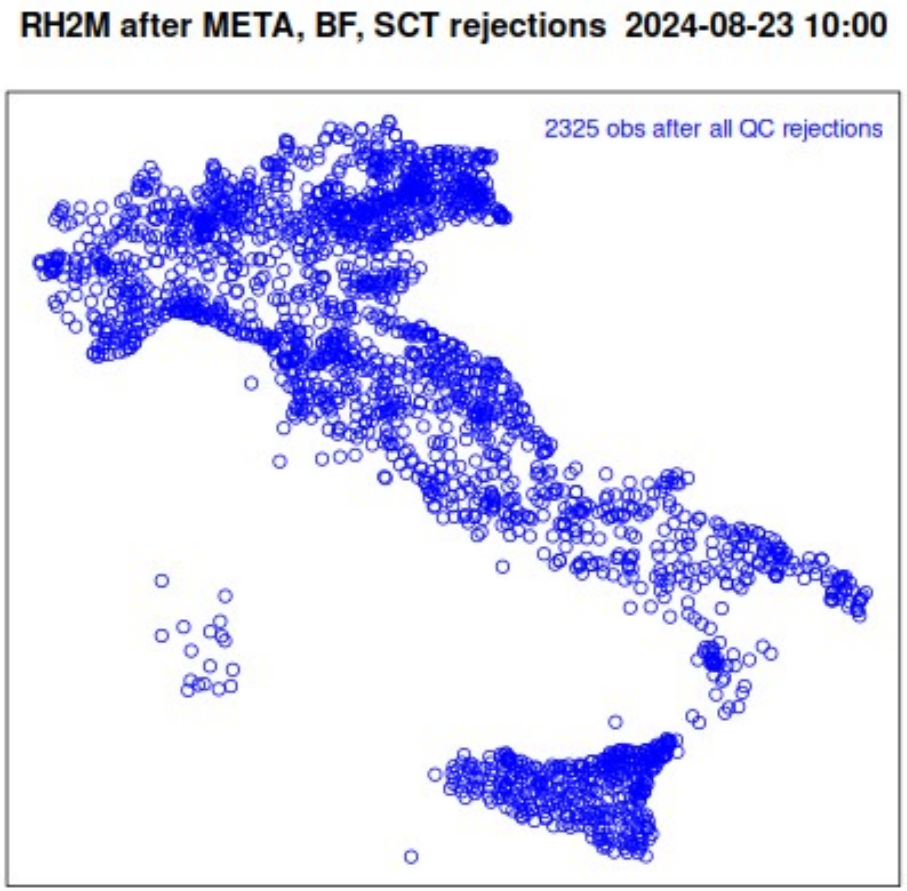
Uboldi, Lussana and Salvati, 2008. Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks. *Meteorological Applications*, **15**: 331-345, DOI:10.1002/met.76 (and DOI:10.1002/met.108)

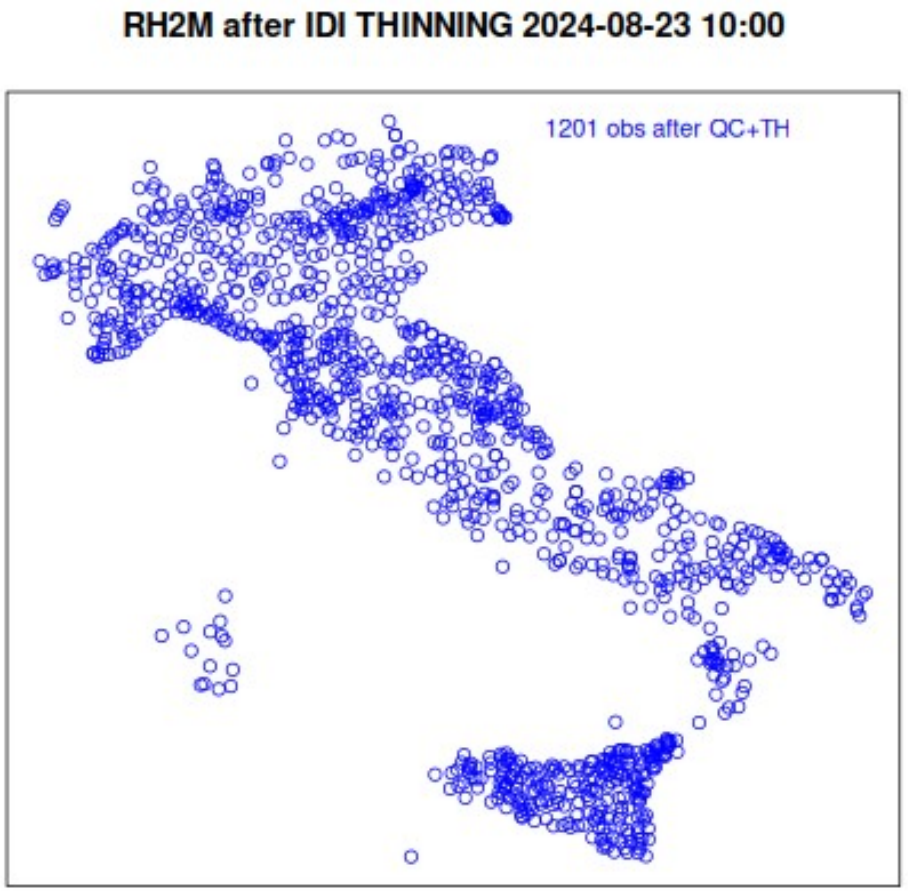
T2M IDI based thinning



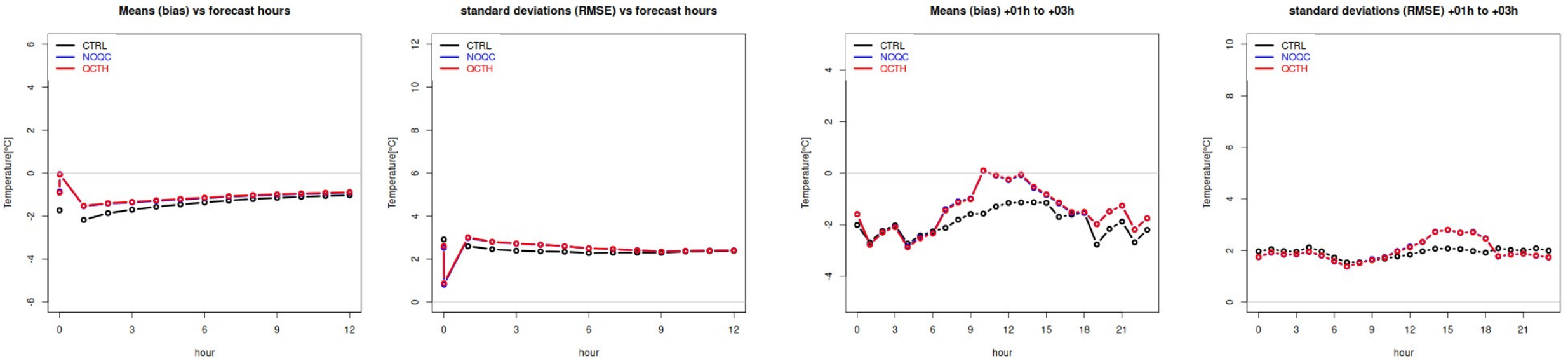


RH2M IDI based thinning (not used in assimilation)





Assimilation results



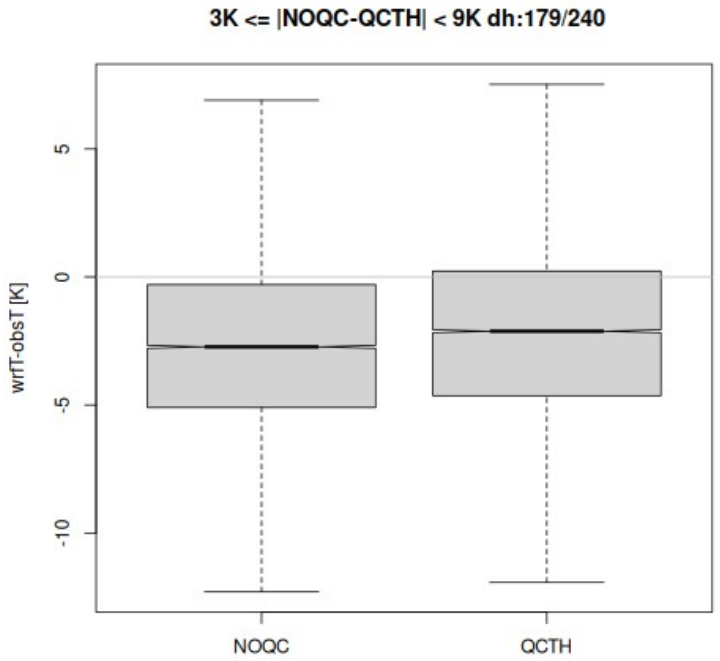
- Assimilation of T2M *in situ* obs does have significant impact, lasting several hours into forecast time
- General bias reduction (both NOQC and QCTH)
- Both NOQC and QCTH worsen (increase) the RMSE in afternoon hours (summer convective period)
- Differences between NOQC and QCTH do not significantly appear in monthly statistics
- NOQC/QCTH differences are localized in space and time, due to different data rejections

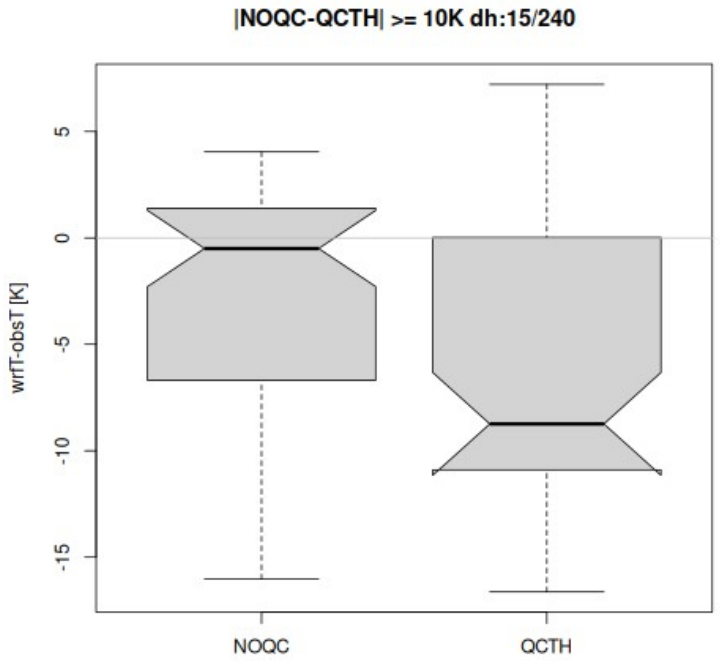
Assimilation results

N. of runs = N. of I.C. with max(|ΔT|) > τ , τ = 1,2,...15 K

N. of available runs/CI: 240 / 248 = 31x8

dT [K]	NofCI NOQC-CTRL	NofCI QCTH-CTRL	NofCI QCTH-NOQC
1	246	238	240
2	246	238	211
3	238	231	179
4	208	198	131
5	163	158	107
6	131	122	87
7	103	103	64
8	84	79	35
9	60	55	19
10	42	37	15
11	23	25	7
12	12	15	2
13	5	7	0
14	2	1	0
15	0	0	0





- NOQC/QCTH differences are localized in space and time, due to different data rejections
- QCTH appears to perform better when the overall errors are significant but not extremely large
- But in few cases of very large errors QCTH appears to even enhance local errors!

  Data filtered with QC checks provide a self-consistent signal near the surface, which sometimes is propagated by vertical correlations into upper levels and to different state variables in a deleterious way

Conclusions

- Errors in metadata, in particular on station locations, represent a major limitation in making use of dense observational networks.
- Professional DPC and amateur MNW networks are merged
- The BF check is efficient in validating most of observations and in rejecting large discrepancies
- The Spatial Consistency Test is a reliable finer check
- Representativeness errors depend on intended use and may determine spatial inconsistency
- Automatic QC is used every month at CIMA, prior to model forecast verification
- 2-m temperature and 2-m relative humidity
- Innovative data thinning based on IDI and CV-IDI
- Assimilation of 2-m temperature observations into WRF model forecasts in operational configuration has been tested for the whole month of August 2024: further study on assimilation choices is needed

Thank you for your attention!

email: *francesco.uboldi@cimafoundation.org*

info@cimafoundation.org

www.cimafoundation.org