EUMETNET Study A1.05 – Final report and experiences learned at KNMI

AUTHORS: Irene Garcia-Marti (garciamarti@knmi.nl)
Aart Overeem (aart.overeem@knmi.nl)
Kirien Whan (kirien.whan@knmi.nl)
Jan Willem Noteboom (janwillem.noteboom@knmi.nl)

DATE: 1st December 2022

Overview

The objective of this project was to make an initial comparison of the quality controls (QC) developed by the participating national meteorological services (NMS) on two large, European-wide, third-party (3PD) data collections (i.e. WOW, Netatmo) from 2020. In the past years, KNMI researchers have developed three QC procedures to assess the quality of the measurements (i.e. temperature, rainfall, wind speed) provided by these networks. These QCs were originally devised to be applied to the Dutch scope and are implemented using Python or R. Currently, these QCs are applied in "research mode", which implies that they are not yet available for operational purposes. The 3 available QC procedures were applied to the 3PD datasets; therefore this document describes the process carried out and the obtained results.TThe work carried out required a substantial amount of time for data wrangling (i.e. cleaning, re-structuring datasets) given the big data nature of both datasets, which reduced the time for analysis. Nevertheless, we aimed at applying the QCs to the largest possible portion of the data available and report on that. The main hurdles and limitations encountered are mostly computational (e.g. insufficient memory or hard drive resources on local machines), but also scaling up the QC procedures from one country to the whole European extent (e.g. temporal cost of running the algorithms). The current results show that a high number of crowdsourced observations pass the QC procedures for three selected weather variables, which implies that these inexpensive devices conforming high-density networks are a promising complement to the official monitoring networks.

In our view, 3PD collections are here to stay and expand well-consolidated workflows in meteorology and climate sciences. Thus, we think NMS should be well-prepared for these changes. Looking to the future, we think 3PD has the potential of becoming a game-changer when it comes to carrying out impact-based analysis and issuing high-resolution warnings and events, particularly in urban areas. Hence, we recommend that NMS increases their efforts in the evaluation of how 3PD can contribute to their operational services and, simultaneously, be ready for big new data engineering and fusion challenges (Garcia-Marti et al., 2022).

1 – Brief summary of the existing quality controls

In the past years, KNMI researchers have developed or adapted QCs for three key weather variables: air temperature, rainfall, and wind speed. The QCs for air temperature and wind speed were developed using data from the WOW network (i.e. WOW for the Netherlands, WOW-NL), and the rainfall QC was developed using Netatmo data. Figure 1 shows an overview of the three “research mode” QCs. shows an overview of the three “research mode” QCs.

a) Air temperature

The current air temperature QC is a modification of (Napoly et al., 2018) (i.e. researchers external to KNMI). The original research established four mandatory quality filters and three optional ones. During the implementation in Python of this QC for WOW-NL data, researchers thought it pertinent to modify this order, so that the air temperature time-series is not reconstructed, and the optional levels become mandatory. In this way, it is possible to apply the QC to time slices independently. The resulting QC applies intra-station checks (e.g. height-corrected temperature, outlier detection), hence does not need large reference datasets. At the end, each individual observation is labeled with a quality flag (i.e. M0 is the lowest quality level, M4 is the highest).

b) Wind speed

The wind speed QC is described in (Chen et al., 2021) and the R source code is available in Github (Chen et al., 2021b). This QC is an interval-based filter that operates with time intervals of 10 minutes. It contains a time standardization function that maps the timestamp of each observation to a 10-min interval, if the input datasets are not aggregated to this temporal resolution. The QC consists of four stages in which different filters are applied. These stages contain intra-station (e.g. range, step, and persistence tests) and inter-station (e.g. comparison with neighbors) filters aimed at quantifying how much a time interval of each station deviates from the neighbors and from previous measurements. In addition, this QC uses KNMI data as reference to apply a bias correction factor. As a result, each 10-min time interval has flags for each QC test and is labeled as “1” or “0”, depending on whether it passes all the stages. In this work, only the intra-station tests were applied.

c) Rainfall

The rainfall QC is described in de Vos et al. (2019). The original R code applied inter-station filters to assess whether an observation deviates from its neighbors, and a bias correction factor using local climate data as default, but able to dynamically update employing the 3PD data. This QC was later ported to Python and transformed into a radar version, which depends on unadjusted KNMI radar data (MSc thesis: van Andel, 2021), but in this project we are using the original version. This QC operates at the time-series level, that is, checks each observation belonging to a time-series to verify whether it is good enough for subsequent processing and applications.

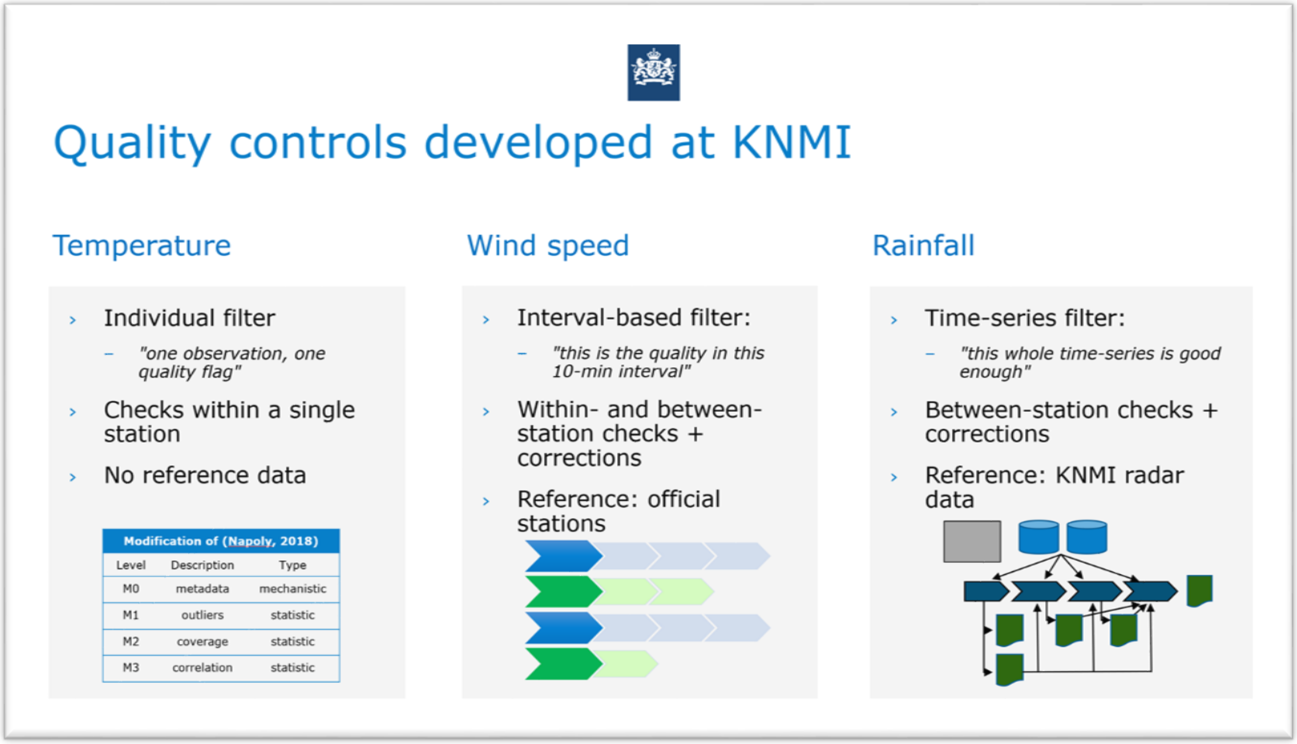


Figure 1. An overview of the three quality controls developed or adapted by KNMI researchers for third-party data.

2 – Data processing

The European-wide WOW and Netatmo collections for 2020 were downloaded via FTP from the EUMETNET Sandbox available in the CEDA repository (https://data.ceda.ac.uk/badc/eumetnet-sandbox). Both collections were stored in a workstation in its original form. In collaboration with Met Norway, a Netatmo rain gauge dataset from the period September 2019 through August 2020 was obtained, which largely overlaps with the Sandbox dataset. In addition, KNMI developed a new dataset EUropean RADar CLIMatology within the internal EURADCLIM project, which is a publicly available climatological dataset of 1-h and 24-h precipitation accumulations at a 2-km grid, which is used as a reference for the rain gauge data (Overeem et al., 2022). EURADCLIM is a combination of EUMETNET OPERA radar precipitation data and European Climate Assessment and Dataset (ECA&D) daily rain gauge data. Within the project EURADCLIM, an exploratory analysis was performed on the potential of crowdsourced Netatmo rain gauge data for improving OPERA radar precipitation accumulations. Here, only the quality of the Netatmo gauge data is studied by comparing to the EURADCLIM dataset.. EURADCLIM is a combination of EUMETNET OPERA radar precipitation data and European Climate Assessment and Dataset (ECA&D) daily rain gauge data. Within the project EURADCLIM, an exploratory analysis was performed on the potential of crowdsourced Netatmo rain gauge data for improving OPERA radar precipitation accumulations. Here, only the quality of the Netatmo gauge data is studied by comparing to the EURADCLIM dataset.

2.1 – Data extraction

WOW observations are grouped on a monthly basis and stored at a European scale, which implies there are 12 large CSV files (i.e. roughly 4 GB each) for the processing and 12 MD5 checksum files. The column names of these files is different than the ones downloaded by KNMI from the WOW repositories, hence, it was necessary to map one these new column names to well-known ones to pass the QCs.

Netatmo has a different structure, in which data is not organized following the spatial and/or temporal dimensions but following a station-based organization. Due to this and the large number of stations in this network, the data extraction yields approximately 6 million files. On the first level, data is organized in country folders, and inside each country folder there are 12 subfolders, one per month. Each of these subfolders contains a variable number of files. The files are organized in the following way: each station has associated a `XXXX.metadata.json` file, containing a description of the station. Then, based on the type of instrument monitoring the weather it will have one or more CSV files associated to this metadata file. Thus, a station monitoring all variables would also provide `XXXX.outdoor.historic.csv`, `XXXX.pressure.historic.csv`, `XXXX.rain.historic.csv`, and `XXXX.wind.historic.csv`. Each of these smaller CSV files contains the time-series for the whole month. This organization creates a vast data structure depending on the size of Netatmo data in each country. For example, for the Netherlands (~2,500 stations) in November 2020, this network reported observations packed in 9,088 files. However, for France (~25,000 stations) in November 2020, the observations produced were grouped in 71,900 files. For the Netatmo rain gauge data as obtained in the project EURADCLIM, the structure is similar, except that data from all countries are in one folder per month (~10 GB per month) and only the rain historic files are made available.

After this step, both data collections were iterated to create a list of the available stations for each network, regardless of the country where they are placed. In total, there are 5,828 stations in WOW and 34,062 stations in the Netatmo collection. Note that for Netatmo, not all stations report all weather parameters. Once these lists are prepared, it is possible to create basic visualizations to better explore the distribution of stations across Europe. Figure 2 shows a histogram of the number of stations per country for each of the two networks. Figure 3 shows the same histogram but turned into a map. Note that to ease the visualization between the two networks, the color scales span across different ranges.

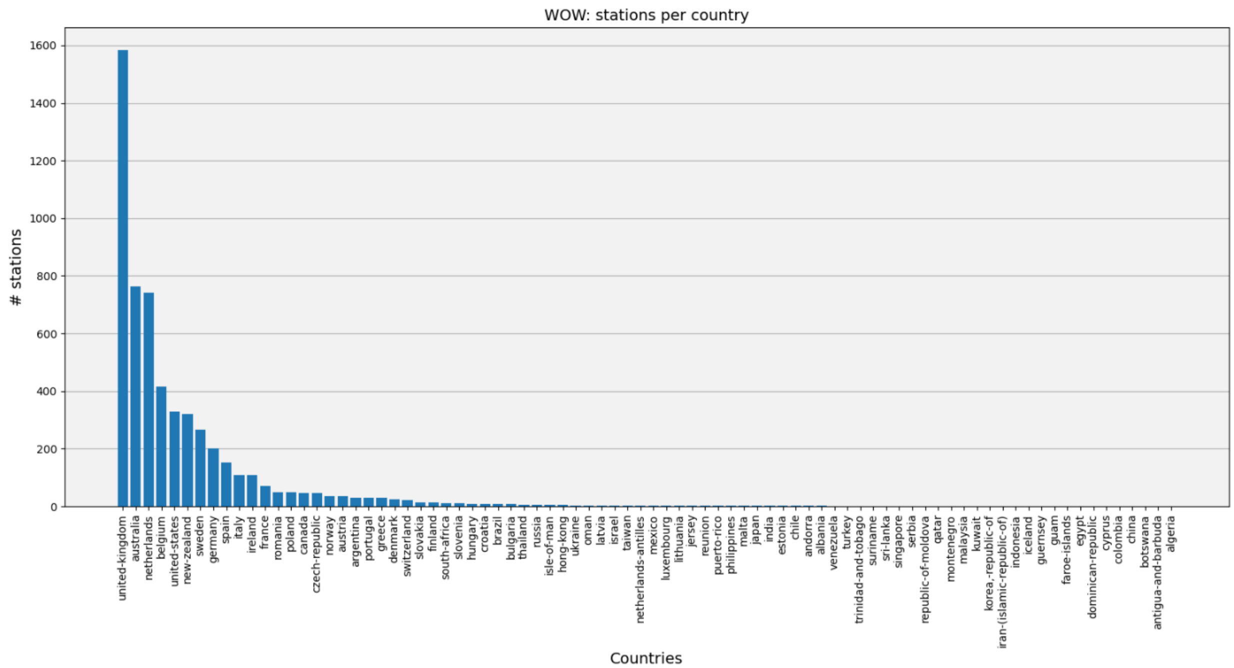


Figure 2. Histogram showing the number of stations per country and network (WOW, above; Netatmo; below)

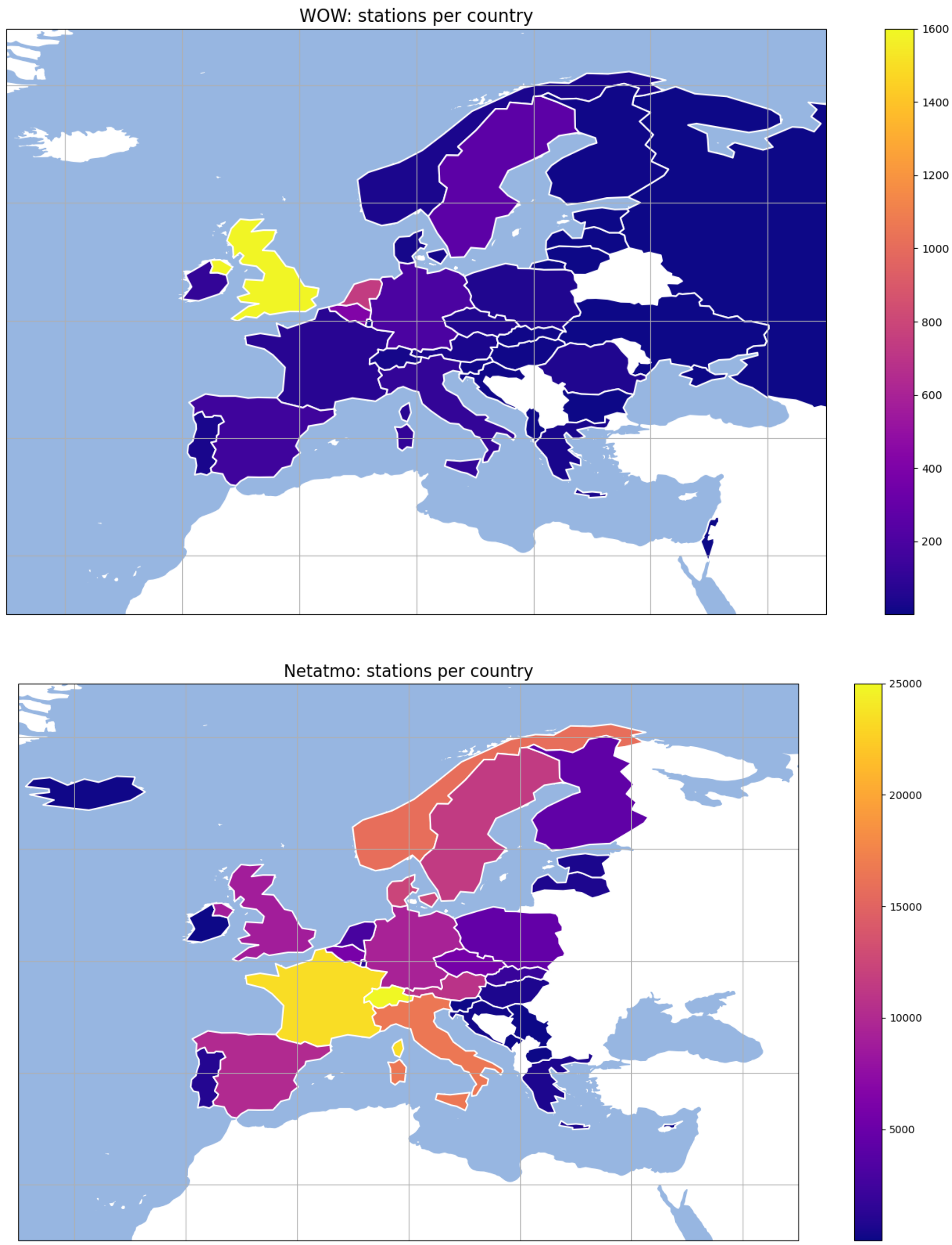


Figure 3. Maps showing the number of stations per country and network (WOW, above; Netatmo, below)

2.2 – Data aggregation, organization and compacting

The available QCs for 3PD operate at different temporal resolutions. The temperature QC can work with observations at the timestamp resolution, but the wind speed QC operates at 10-min intervals. The type of temporal aggregation that QCs need at the input is a good reason to aggregate data, but not the only one. Given the sheer number of files in the Netatmo collection, it seemed appropriate to combine the separate files per variable, so that there is only one “metadata” file and one “data” file (instead of up to 4). Hence, the raw data collections were re-organized, aggregated and/or compacted to facilitate the process of applying the quality controls.

For this purpose, the different Python workflows created in this project apply these operations at various stages. For example, a dedicated script aggregates the observations to an X-min interval that can be later used to write smaller CSV files easier to process by the QCs. Hence, WOW data was re-organized to a 10-min resolution (but not averaged on the time-interval). We applied this process to Netatmo as well, but this collection needed compacting too. Hence, Netatmo per-country file structure was traversed to aggregate this collection to 5-min and 10-min, so that if any station reports more than one observation in this time interval it is averaged. Also, all partial observational files from Netatmo we grouped, so that each station only has one “metadata” and one “data” file.

These type of operations on large datasets tend to have a high temporal cost. The original workflows described in this sub section were originally developed for WOW, but the size of the Netatmo collection, prompted us to use basic Python threading to speed up these data wrangling operations (i.e. in a local high-end laptop computer). In this way, the wrangling operations for Netatmo decrease from 18h to 13h, which suggest that the availability of larger machines or cloud infrastructures could prove relevant to handle such data collections. Both collections are stored in CSV format.

For the EURADCLIM’s Netatmo rain gauge dataset, data are preprocessed per month to construct a Netatmo rain gauge dataset at regular 5-min intervals with one data and metadata file per month in binary format (R Object). This is based on openly available code written in language R (https://github.com/LottedeVos/PWSQC), developed by de Vos et al. (2019), with some modifications to work with the data on an European level, and not including any computations with the reference dataset. The aggregation and compacting takes approximately 90 hours for the 1-year Europe-wide dataset when run on one single core of a high-end desktop computer. The locations of Netatmo rain gauges are shown in Figure 4, displaying a generally large network density. Some countries have a much lower network density, but this density will often still exceed that of official gauge networks from which data are available, such as those in the European Climate Assessment and Dataset (https://www.ecad.eu/). In general, the network density is expected to be ten times as large as that of the gauges available in ECA&D.

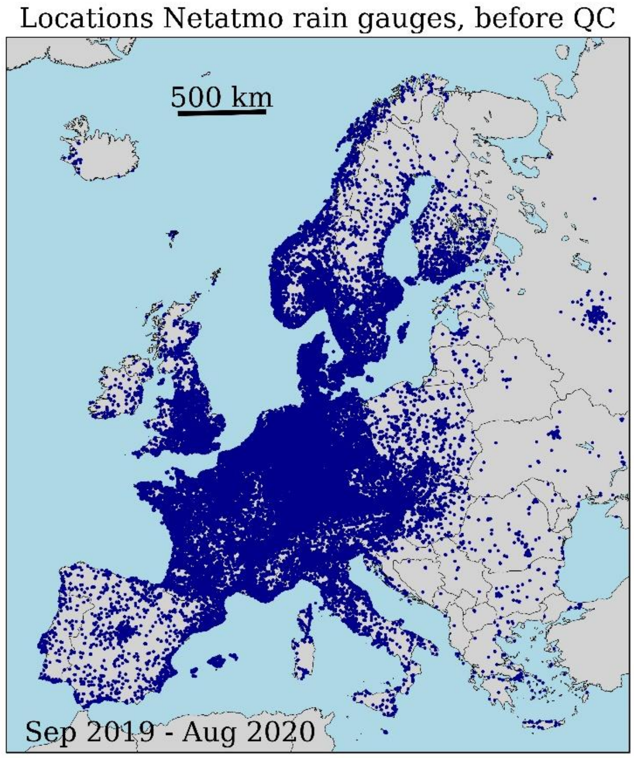


Figure 4: Locations of Netatmo rain gauges over Europe over the period September 2019 through August 2020, before applying any quality control.

2.3 – Preparing data stores and reference data for quality controls

The application of QCs on WOW and Netatmo requires assembling reference datasets and/or creating databases to store the quality flags. This section describes the requirements to apply each QC to the data collections.

a) Temperature

The air temperature QC relies on a Postgres database to, first, store the raw observations in separate tables per country and, second, to store the results of the QC in a matching table. This data organization was chosen (back in 2019) keeping in mind that once this QC enters the KNMI Operations pipeline, it is more likely that the observations will be stored in a database server (possibly in KNMI's digital infrastructure in the cloud), hence feeding other services from there. Thus, in this project we decided to keep the original principle for the air temperature QC with European-wide 3PD collections.

The initial database model was extended, to include other countries than the Netherlands. Python scripts and SQL queries were created to create an empty structure capable of holding the observations. This database organization is also convenient to quickly retrieve slices of data and represent observations in different visualizations. Therefore, the *"eumetnet_wow"* database contains 34 tables to store the observations of each country and 34 more to store the (partial) results of the QC. This process was adapted (e.g. new SQL queries, field names, testing) and repeated for the *"eumetnet_netatmo"* database, which contains 32 tables for the observations and 32 for the QC results. The WOW database was populated with the raw observations and the Netatmo database with the observations aggregated at 5-min. Note that this database is in a local laptop and was not accessible to the rest of the researchers for the team.

Regarding additional reference data collections, the air temperature QC does not need much to run. One of the procedures of the original QC (Napoly et al., 2018) is using an elevation file to correct for the atmospheric lapse rate. In this project, we have used GTOPO30 global elevation map from the USGS. The QC also requires counting the number of observations that a station has reported in each day and month. This is necessary to identify outliers in the data. For this purpose, Python scripts using SQL queries were developed, so that two boards (in CSV) are written containing these counts and made available before the execution of the QC. Note that this is a temporally costly operation that takes several hours to complete.

b) Wind speed

The wind speed QC uses the 10-min aggregation CSV files described in Section 2.2. The wind QC was applied to both WOW and Netatmo data. The 10-min European-wide files have the same format as the ones used in (Chen et al., 2021). Out of the three stages of the original QC (i.e. within-station, between-station and bias correction), we performed the within-station tests described in this publication. The within-station stage consists of the following steps: 1) Isolation test: remove stations with insufficient observations; 2) Range test: flag implausible wind speeds (e.g. greater than a category 4 hurricane); 3) Step test: flag wind speeds where there is a large jump (i.e. using KNMI data for the Netherlands as reference); 4) Persistence test for zero and non-zero values: flag wind speeds where the variability is too small. Using these within-station filters, no additional reference data is needed, but note that the application of the entire QC would require European climatology data.

c) Rainfall

The rainfall QC (de Vos et al., 2019; https://github.com/LottedeVos/PWSQC) uses the 5-min aggregated rain values in binary R Object files. To run the rainfall QC it is necessary to accumulate these 5-min files into 1-hour accumulations. This process will be carried out if the availability of the 5-min files is at least 83.3%. Also, only Netatmo gauge accumulations larger than 0.25 mm in 1 hour are kept. Moreover, only Netatmo gauge accumulations are selected when an OPERA-based 1-hour radar precipitationi dataset has more than 0.25mm at the respective gauge location. In this way, erroneous zeros or low gauge values are effectively not evaluated. The selected Netatmo 1-h gauge accumulations are compared to those from EURADCLIM (Overeem et al., 2022), a merged OPERA radar and ECA&D rain gauge product, which is used for an independent verification

To speed up the computations of the rainfall QC, it is possible to create additional auxiliary files. For each Netatmo rain gauge a neighborhood list with the 20 closest neighbouring istation within a 10 km radius is created. The idea behind this step is to reduce computational time in case the QC is temporally costly. For example, running one process per month on a high-end desktop computer with 8 cores takes 12h for the 1-year European-wide dataset.

2.4 - Comparing EUMETNET's Netatmo and EURADCLIM project Netatmo raw datasets

In section 2.1 we briefly described a new dataset developed by KNMI researchers: EURADCLIM. During this project's development, a Netatmo dataset spanning Sep 2019 to Aug 2020 was made available. To speed up the production of insights, we decided to incorporate the results of EURADCLIM into this Study A1.05, since the QC of Netatmo data was an important part of the project. However, we prepared a couple of visualizations to assess whether both Netatmo collections are similar.

For this purpose, we developed Python scripts traversing the two Netatmo file structures and counting the number of encountered stations and the associated observations during the two overlapping periods. With these count data, we prepared two heatmap visualizations intended to show how similar both collections are. Figure 5 (top) shows the number of stations per month in both datasets, whereas Figure 5 (bottom) shows the total number of observations per month reported by both Netatmos. The heatmaps resemble each other but are not the same. This was in part expected, since both Netatmos only partially overlap. However, both panels show that the heatmaps do not become the same for the overlapping period either. This might be an important detail to keep in mind, since ensuring that researchers across organizations receive the same datasets in different moments of time seems important for the reproducibility of future experiments using 3PD. For this project's purpose, we believe both datasets are similar enough to include the results of EURADCLIM in this EUMETNET Study A1.05. Hence, in this document, the results for the rainfall QC come from EURADCLIM's Netatmo.

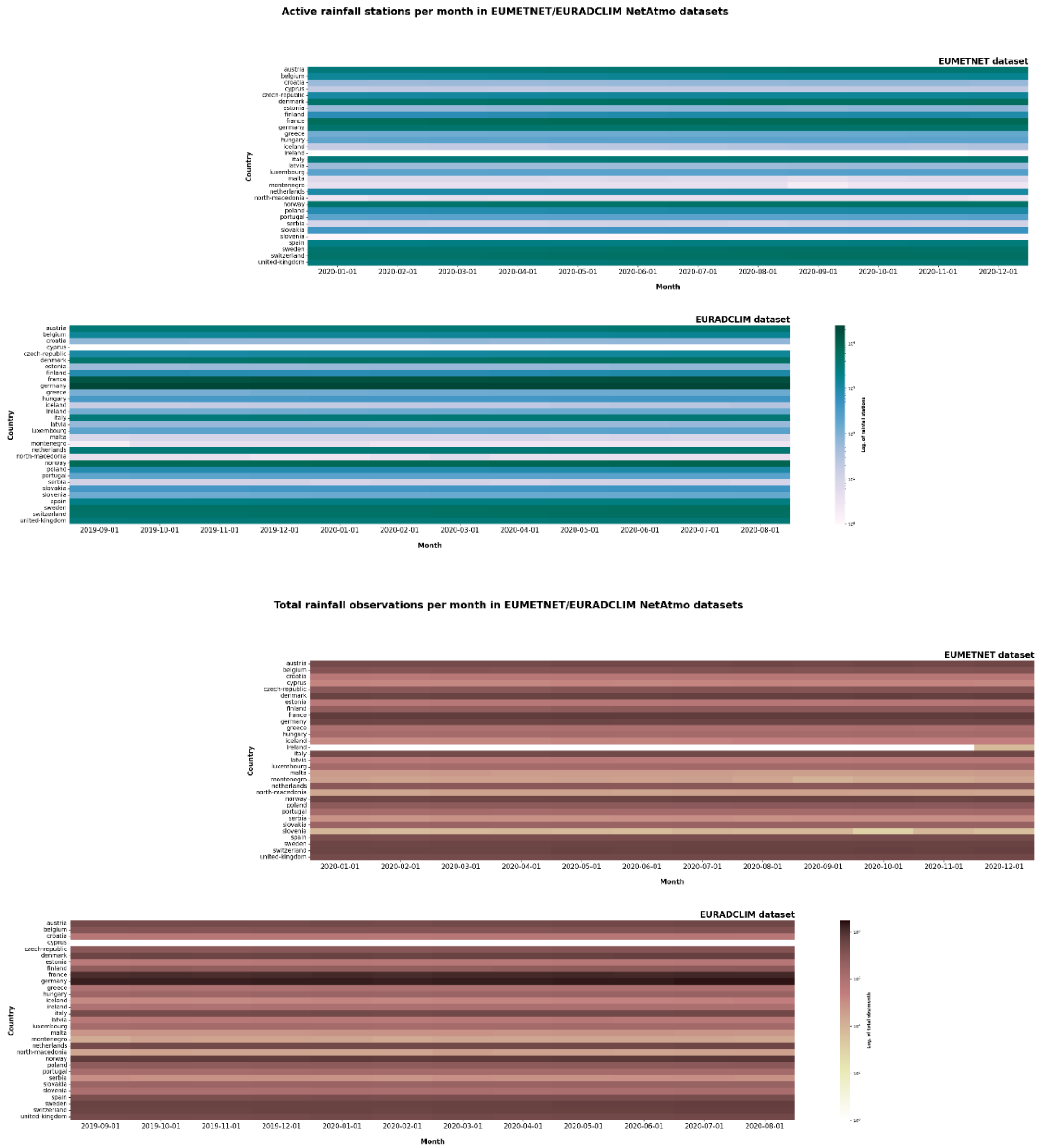


Figure 5. Heat maps showing the number of monthly contributing stations (top) and the number of monthly observations (bottom) for the two available Netatmos (ie. EUMETNET and EURADCLIM)

3 – Applying quality controls: description of experiments and results obtained

After the data processing described in Section 2, we applied the available QCs to the largest possibly portion of data. Table 1 summarizes these efforts.

	Variable	WOW	Netatmo	Comments
1	Air temperature	Yes (3.1)	Yes (3.2)	WOW: applied to whole data collection Netatmo: Disk space exceeded, QC only available for 4 countries
2	Wind speed	Yes (3.3)	Yes	WOW/Netatmo: Applied to 12K out of 28K stations (as per 30/11/22)
3	Rainfall	No	Yes (3.4)	Netatmo: EURADCLIM project, excessive memory and computational usage required simplifying QC

Table 1. Summary of the QCs applied in this project

3.1 – WOW + Temperature QC

The application of the air temperature QC to WOW data was tackled in two phases to identify potential problems. First, Python scripts were developed for Belgium, to adapt and test the original scripts for a region that is not the Netherlands and resolve minor problems. Second, it was extended to all countries found in the dataset. The air temperature QC is structured in two parts: mechanistic filters and statistical filters. The mechanistic filters first check whether a station is one of the invalid ones, then checks whether the metadata are incorrect (at this point is verifying whether lat/lon are the same), and then checks whether the daily/monthly coverage is sufficient, following the heuristics from the paper. As seen this party relies on the reference tables created in section 2.3a above. The statistical checks first correct for the temperature lapse rate using the elevation map described above. Then, the Z-score of the observations is computed and checked with a robust Qn estimator that would determine whether an observation is an outlier. Finally, a Pearson correlation coefficient is computed between the median of the inspected station on that hour and the median of the other stations reporting observations in that hour.

After this step, a quality flag is calculated based on a simple if-else structure and attributed to an observation. Once a slice of time is QC'd, the quality flags are inserted back into the database, as described in section 2.3a. The quality flags are: M0 (ie. insufficient metadata), M1 (ie. Outliers), M2 (ie. Insufficient daily/monthly coverage), M3 (ie. Insufficient correlation), M4 (ie. OK). Hence, observations reaching level M4 are the ones with the highest quality, according to this QC.

Figure 6 shows a summary of the results of the air temperature QC applied to the whole European extent. The histograms (above) show each country placed along the X-axis, whereas the Y-axis shows the frequency of the quality labels (ie. note that the scale is different between plot rows). The main highlights of the figure is that the WOW network does not seem to have too many problems with metadata, outliers, or coverage, since their associated color bars are small. Some countries present uncorrelated values, but in general this percentage is low too. The plot on the right shows the same information in tabular form. The column “percent” shows the percentage of observations per country that reach level M4. As seen, half of the countries provided at least 70% of the observations with the M4 quality level.

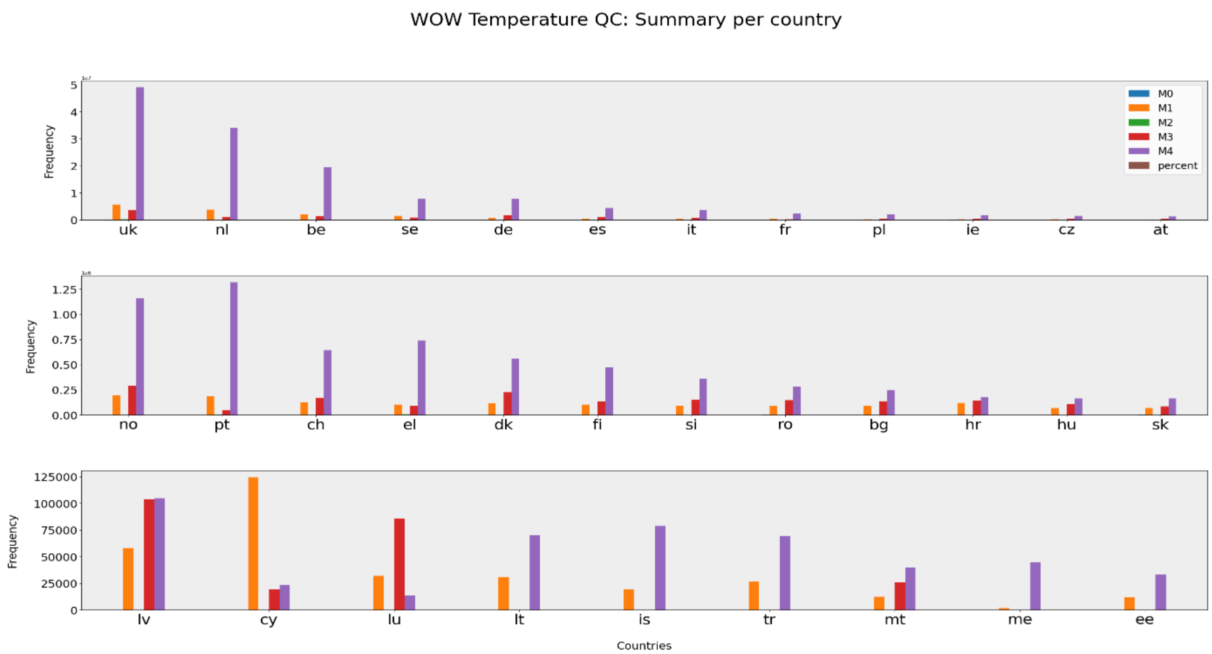


Figure 6. Summary per country of the air temperature QC applied to each country of the dataset. The histograms (above) show each country along the X-axis and the number of quality flags (note the change of scale between plot rows) in the Y-axis. Each color in the legend represents the quality levels that the observations reach. Level M4 (ie. In purple) is the one grouping the ones with the highest quality. The right plot shows the same information but in tabular form. The last column “percent” shows the percentage of observations per country that reached the highest quality. As seen, half of the countries provided at least 70% of the observations with the M4 quality level.

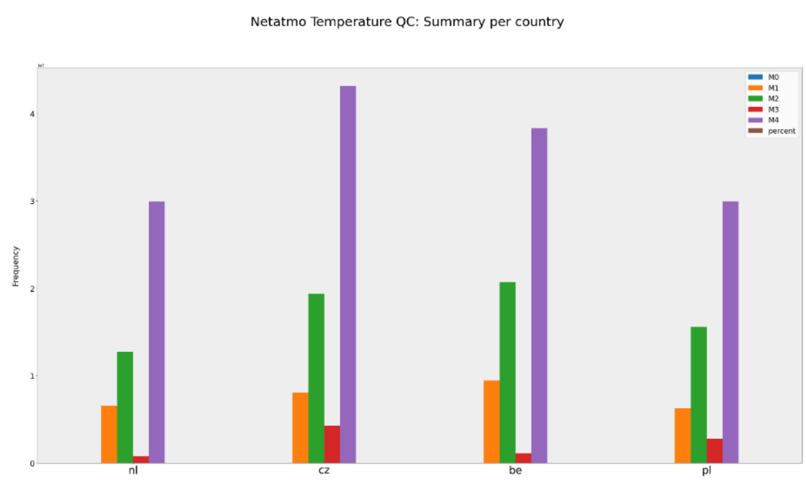
	M0	M1	M2	M3	M4	percent
me	0	1596	2	0	44404	97.0
nl	19563	3809807	47381	1015555	34123250	87.0
be	12987	2044421	14348	1413080	19601196	85.0
pt	3	186997	1978	48339	1318585	85.0
uk	75979	5751821	191215	3557576	49064772	84.0
is	0	18988	15	0	78677	81.0
el	643	104016	5310	91701	737988	79.0
fr	1639	337390	6076	299459	2408296	79.0
se	12983	1531414	17305	936486	7781688	76.0
es	651	500509	11740	989920	4435465	75.0
de	8697	764339	9542	1750387	7711293	75.0
ee	3	11963	0	6	33374	74.0
it	3784	523802	4915	752852	3721099	74.0
pl	21	281040	1794	440566	1991037	73.0
tr	0	26767	0	0	69113	72.0
cz	2989	207667	3970	408822	1459400	70.0
lt	4	30843	0	0	70412	70.0
no	1723	194170	1546	291274	1159430	70.0
ch	1134	122606	1470	171869	643941	68.0
at	6073	177909	807	417174	1261117	68.0
fi	11	103760	630	132564	476538	67.0
ie	86408	237249	2598	524934	1723840	67.0
al	0	3649	180	12326	27396	63.0
dk	864	117707	85	226565	555995	62.0
si	1142	87732	34	154390	359636	60.0
ro	2037	89019	32	145696	279315	54.0
bg	0	93858	816	135153	245105	52.0
mt	0	12271	82	25873	39640	51.0
sk	1855	69942	155	86605	162966	51.0
hu	1182	68893	500	107632	164672	48.0
hr	0	119926	505	139320	175493	40.0
lv	1	58191	0	103823	104757	39.0
cy	0	124446	44	19060	23323	14.0
lu	0	31852	0	85879	13738	10.0

3.2 – Netatmo + Temperature QC

The application of the air temperature QC to Netatmo data was tackled as described in section 3.1, first for Belgium, then extended to all countries in the dataset. The QC code is also the same as the one used in the previous section.

While applying the QC to the Netatmo dataset we found out a new serious limitation for this project: the local laptop holding the database ran out of disk space. A closer inspection revealed that the Netatmo database occupied 530GB of disk space, thus consuming all the free space in the hard drive of the local laptop. Hence, despite preparing all workflows to apply the QC to Netatmo data, we could only apply it to four selected countries: the Netherlands, Czech Republic, Belgium, and Poland.

Figure 7 shows a summary of the QC labels produced for the four selected countries. As seen (above), the QC identifies a substantial number of observations flagged as either outlier (ie. M1) or with insufficient daily/monthly coverage (ie. M2). The table (below) shows the percentage of observations per country reaching level M4. In here, the country with the highest number of M4 observations is the Netherlands (60%) followed by the Czech Republic (58%). These results are quite different than the use case with WOW data, which might suggest that Netatmo stations acquire observations with less quality. Nevertheless, since we could only apply this to four countries, the results are not complete enough yet to make a sound conclusion.



	M0	M1	M2	M3	M4	percent
nl	2618	6564895	12734732	761343	29924963	60.0
cz	16598	8096666	19350554	4261808	43123738	58.0
be	15034	9452832	20702747	1105406	38323372	55.0
pl	0	6229056	15556000	2795907	29936740	55.0

Figure 7. Summary per country of the air temperature QC applied to four selected countries in the Netatmo dataset (I.e. the NL, CZ, BE, PL). The histograms (top) show each country along the X-axis and the number of quality flags in the Y-axis. Each color in the legend represents the quality levels that the observations reach. Level M4 (purple) is the one grouping the ones with the highest quality. The bottom plot shows the same information but in tabular form. The last column “percent” shows the percentage of observations per country that reached the highest quality. As seen, the Netherlands is the country producing the most observations with the highest quality level, with a total of 60%, followed by the Czech Republic with 58%.

We put together the results obtained in Section 3.1 and 3.2 for the four common countries with quality labels. Figure 8 contains four sub panels (one per country) comparing the histograms obtained after applying the QC to WOW and Netatmo. The sub panels show a similar pattern: the number of level M4 observations in the Netatmo network is higher than in WOW (except for the Netherlands), but the Netatmo network seems more prone to generate outliers (ie. M1) or observations with insufficient daily/monthly coverage (ie. M2) than WOW. More investigation is required to find out whether this pattern also occurs for more European countries.

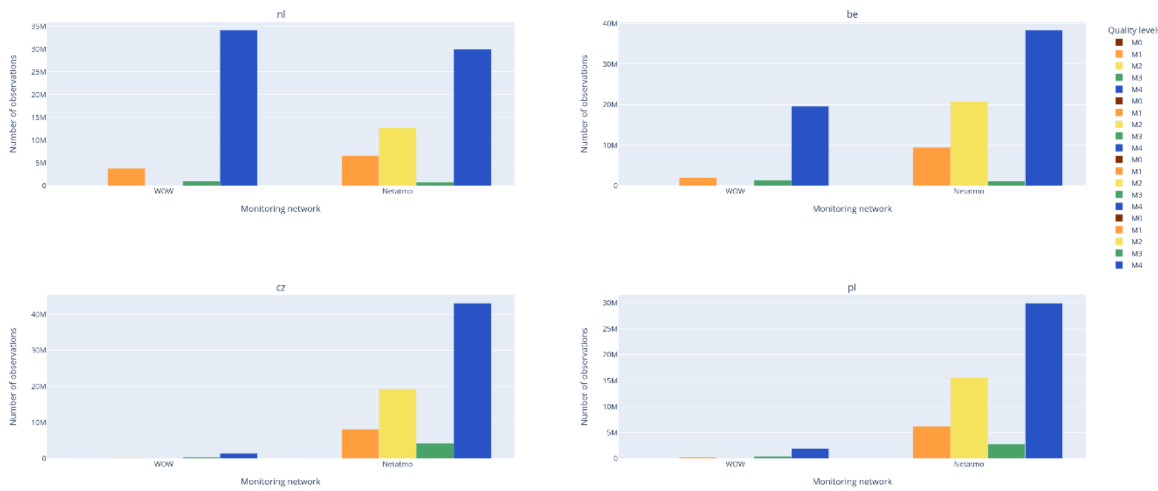


Figure 8. Comparison of air temperature quality labels for WOW and Netatmo for four selected countries. The current results might suggest that the WOW network produces air temperature observations with a higher quality than Netatmo, but more research is required.

3.3 – WOW + Wind speed QC

The within-station wind quality filters were applied to the WOW and Netatmo collections, but in this sub section only results with WOW are shown. Out of all the available stations in both datasets, 37,422 provide wind speed measurements. We applied the isolation test and 28,399 stations were sent to the step, range, and persistence filters. After this process, roughly 14,000 stations have been QC'd with the within-station tests (i.e. as per 30/11/22). We discard stations that have too many flags (>95% of the data), which leaves a final set of roughly 7000 stations. We show below some examples of the effects of the quality control in a few stations from Greece and Italy.

Figure 9 shows a station located in Greece (i.e. EL_646). The central panel depicts the original time-series and the colored dots show observations in the time-series that are suspicious of not having a good quality. In this case, some of them report >1500km/h, and some others a too steep increase in the wind speed. In the right panel, these observations have been removed, and the reader can see that the time-series looks better now. A similar pattern is found in Figure 10, in which a WOW station located in Italy has produced abnormal wind speed values. However after the QC becomes more credible. Figure 11 shows another station located in Greece (i.e. EL_648). As seen, in here the QC does not label too many observations as suspicious, so that when these are removed (right panel) the time-series is practically the same as before. This could be an example of a “good station”, at least concerning the within-station filters.

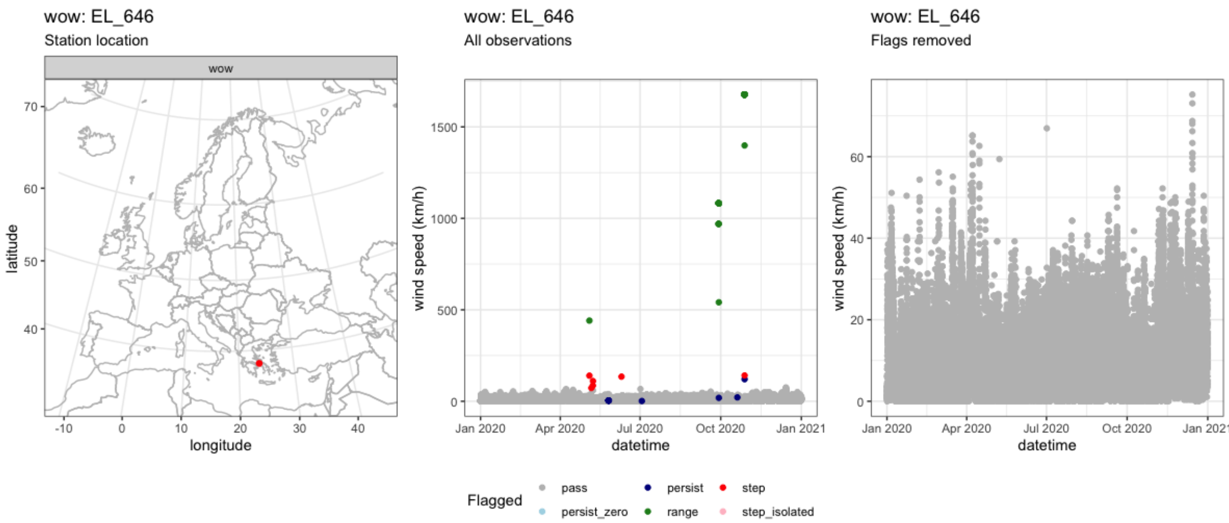


Figure 9. A WOW station (EL_646) located in Greece before the QC (center) and after the QC (right)

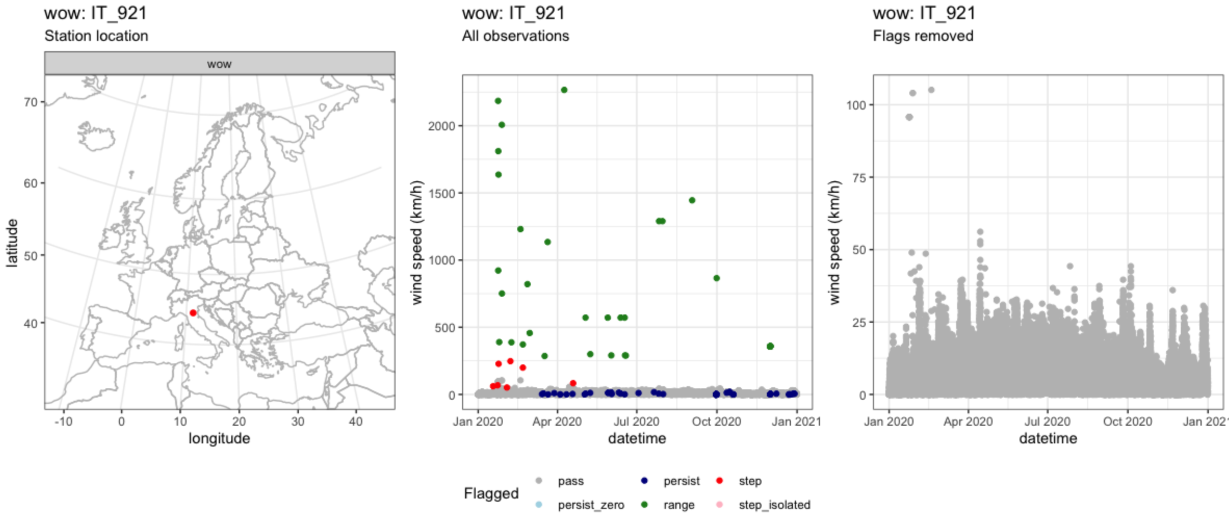


Figure 10. A WOW station (IT_921) located in Italy before the QC (center) and after the QC (right)

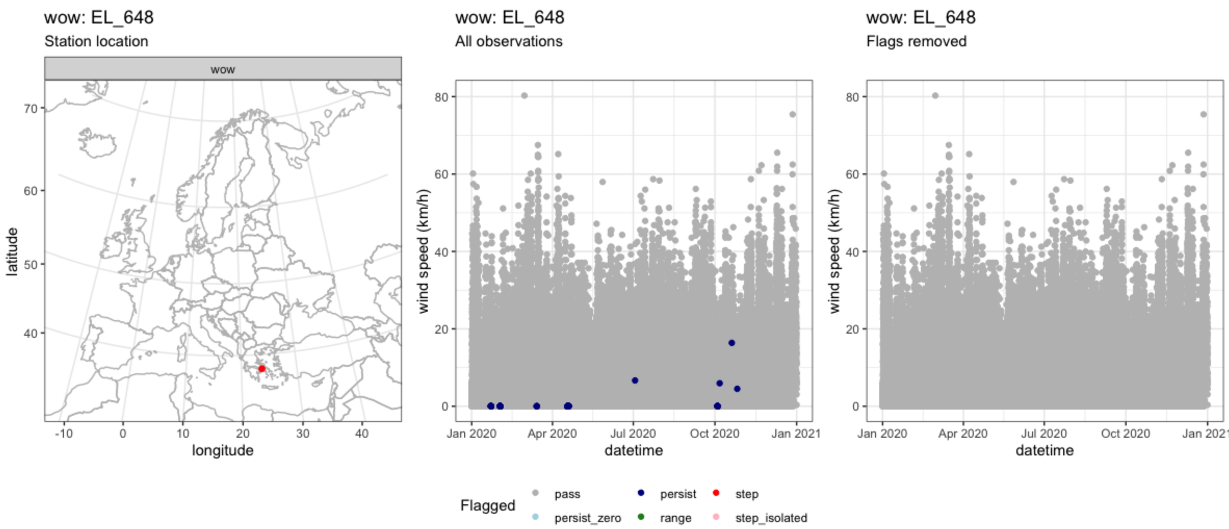


Figure 11. A WOW station (EL_646) located in Greece before the QC (center) and after the QC (right)

3.4 – Netatmo + Rainfall QC

We applied the faulty zeroes and high influx filters from the original rainfall QC described in (de Vos et al., 2019) to the 5-min Netatmo data (with some technical modifications). With this setup, running one process per month on a high-end desktop computer with 8 cores takes 1.5 hours for the 1-year Europe-wide dataset.

Figure 12 provides scatter density plots of Netatmo 1-h rainfall accumulations against those from the EURADCLIM dataset over the 1-year European-wide dataset. In total, more than 28 million 1-h Netatmo accumulations are compared to EURADCLIM accumulations. The spread is quite large, which is partly due to representativeness errors. EURADCLIM is based on radar data with 4 km^2 grid cells, whereas the measurement volume of a Netatmo gauge is only a fraction of the grid cell size. Moreover, radars measure aloft leading to timing differences and the possibility that precipitation actually reaches the Earth's surface at another grid cell. Apparent are the group of large Netatmo accumulations for low EURADCLIM values. The metrics in the graphs indicate the relative bias in the mean, which is quite close to 0, the coefficient of variation of the residuals (CV), a measure of spread, and the coefficient of determination (squared Pearson correlation coefficient). Note, however, that the relative bias varies in space (not shown), will likely also vary in time and depend on the applied thresholds. Moreover, the group of large Netatmo accumulations may conceal underestimations. Application of a dynamically updated bias correction factor per station (de Vos et al., 2019) may help to address this, but may also require more development for this large dataset containing many different weather conditions. A residual is defined as the Netatmo accumulation minus the EURADCLIM accumulation. Application of the quality control leads to a much lower value for CV and the coefficient of determination also improves from 0.09 to 0.30. Some initial results, not shown, where comparisons with unadjusted OPERA radar data are used to remove suspicious Netatmo values, further improve these metrics (CV of 1.05 and coefficient of determination of 0.38), and would remove the group of large Netatmo accumulations Drawback of that approach is its dependence on OPERA radar data.

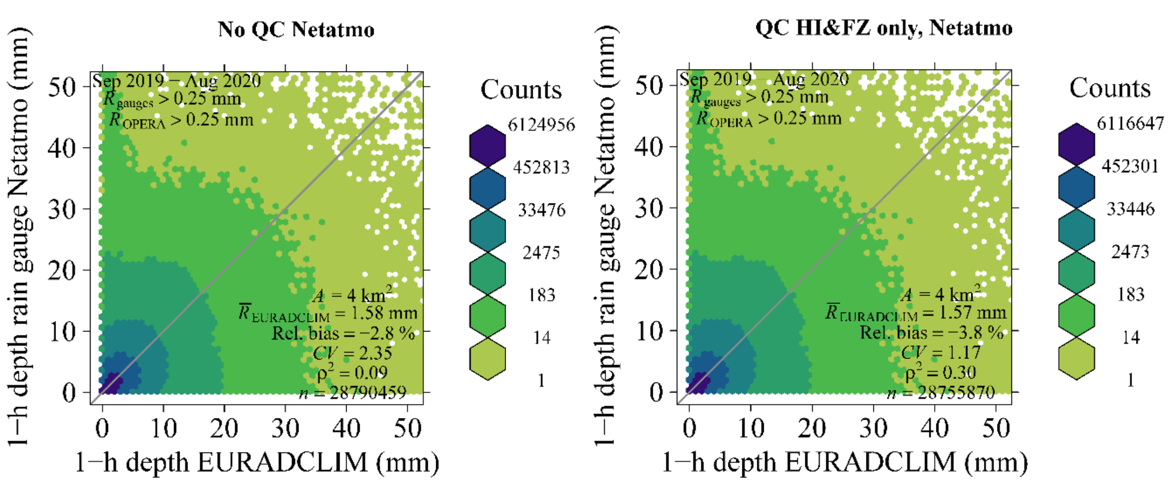


Figure 12. Scatter density plots of hourly Netatmo rain gauge accumulations against EURADCLIM gauge-adjusted radar accumulations over the period September 2019 – August 2020. Results are shown for the

Netatmo dataset without quality control (left) and for a Netatmo dataset which has undergone the high influx & faulty zeroes filter. In both plots, only values are shown when both the Netatmo gauge and the unadjusted OPERA radar dataset (not EURADCLIM) are larger than 0.25 mm.

4 - Challenges encountered during the process

The challenges encountered during this project have been predominantly within the (big) data engineering category. The original QCs were devised for the Netherlands, which is small (but data rich) country. However, at the time of scaling up the original QCs to a larger extent, we started bumping into problems that required shifting the original goals of this project. Perhaps these problems could be mitigated or even non-existent had we used high-performance computers or a cloud-based infrastructure, but these were not available resources in the context of this project. Hence, assuming that the processing and application of the QCs is done in normal to high-end laptop or desktop computers, these are the most limiting challenges:

1. Insufficient hard drive resources: The application of the air temperature QC to Netatmo data revealed that the local database (laptop computer) had exhausted the free space in the hard drive (i.e. used at least 530GB of space), which hampered the application of the QC beyond the selected 4 countries.

2. Insufficient RAM memory: The application of the rainfall QC has a couple of demanding steps, specially during the bias correction and the station outlier filter. The processing would take over 300 hours due to memory limitations (64GB). Since this processing step has a spin-up time of at least two weeks, applying it per month limits its added value. This would imply that the processing is effectively not applied to the first two weeks of each month. Ideally, data would be processed over several months or even the entire year, but this is currently not possible on a desktop computer due to memory limitations.

3. Reproducibility: Having two Netatmo collections available (i.e. EUMETNET and EURADCLIM project) we realized that these collections do not seem to have the same number of stations and therefore the number of observations. We acknowledge that exhaustive tests have not been conducted, but in case that other organizations face a similar situation, this might pose questions when it comes to the reproducibility and transferability of the scientific results.

Other minor issues found during the development of the project are:

1. Data types: Some WOW stations report the wind speed parameter as a string. The casting to float64 fails, since the string contains a poorly formatted float (e.g. “36,612.3” m/s). This required continuous patching of the original Python scripts for QC.

2. Headers name: The received WOW collection has header names different than the ones KNMI researchers have been using since 2019.

3. Implausible values: During the processing or QC we encountered physically implausible values for some measurements (e.g. negative rainfall, wind speeds higher than 60,000 km/h).

4. Consistency of the measurements: Wind speeds are binned for some stations and for parts of some stations.

5 – Recommendations for the future

Our recommendations for the future usage of 3PD by other organizations aims at solving these (big) data engineering problems. For example, teams of software developers, data architects and data engineers could help at organizing these datasets in a cloud-based environment, so that researchers do not need to dedicate substantial amounts of time to the data processing phase. Also, the deployment of these large 3PD collections in the cloud would help researchers to have available larger machines with increased computational resources (e.g. memory, disk space, faster CPUs) and hand, thus allowing them to focus in the analytical part. In addition, such system would enhance the transparency and would make easier to share results at different stages of the development of projects. For a more complete description of our view on the future usage of 3PD at NMS, we refer the reader to Section 5 of our recent paper (Garcia-Marti et al., 2022).

Optimizing code and perhaps translating it to a more efficient programming language may provide a solution. The use of auxiliary data, such as OPERA radar data, could also help to improve quality control on top of the already applied faulty zeroes & high influx filter. And the radar version of the quality by de Vos et al. (2019), as developed by van Andel (2021), could be modified to work with the European Netatmo and radar data. Note that van Andel (2021) found that for dense Netatmo networks, the quality control by de Vos et al. (2019) gives better results than the radar-based version.

Finally, connect to the (outcome of the) COST action OpenSense where different 3PD QC and retrieval algorithms will be tested on large benchmark datasets in collaboration with different institutes (https://opensenseaction.eu/) for, e.g., precipitation from personal weather stations (PWS), such as those from Netatmo. OpenSense also includes the development of standardized data formats for, e.g., PWS data.

References

Chen, J., Saunders, K., & Whan, K. (2021). Quality control and bias adjustment of crowdsourced wind speed observations. *Quarterly Journal of the Royal Meteorological Society*, (July), 1–18. https://doi.org/10.1002/qj.4146

Chen, J., Saunders, K. and Whan, K. (2021b) QCWind: quality con-trol for citizen science wind observation. Github repository. Available at: https://github.com/jieyu97/QCwind.

de Vos, L. W., Leijnse, H., Overeem, A., & Uijlenhoet, R. (2019). Quality Control for Crowdsourced Personal Weather Stations to Enable Operational Rainfall Monitoring. *Geophysical Research Letters*, *46*(15), 8820–8829. https://doi.org/10.1029/2019GL083731

Garcia-Marti, I., Overeem, A., Noteboom, J. W., de Vos, L., de Haij, M., & Whan, K. (2022). From proof-of-concept to proof-of-value: Approaching third-party data to operational workflows of national meteorological services. *International Journal of Climatology*, 1– 18. https://doi.org/10.1002/joc.7757

Napoly, A., Grassmann, T., Meier, F., Fenner, D., & Bell, S. (2018). Development and Application of a Statistically-Based Quality Control for Crowdsourced Air Temperature Data, *6*(August), 1–16. https://doi.org/10.3389/feart.2018.00118

Overeem, A., van den Besselaar, E., van der Schrier, G., Meirink, J. F., van der Plas, E. & Leijnse, H.: EURADCLIM: The European climatological high-resolution gauge-adjusted radar precipitation dataset, *Earth Syst. Sci. Data Discuss.* [preprint], https://doi.org/10.5194/essd-2022-334, in review, 2022.

van Andel, N. van. (2021). Quality control development for near real-time rain gauge networks for operational rainfall monitoring Master Thesis. Utrecht University.