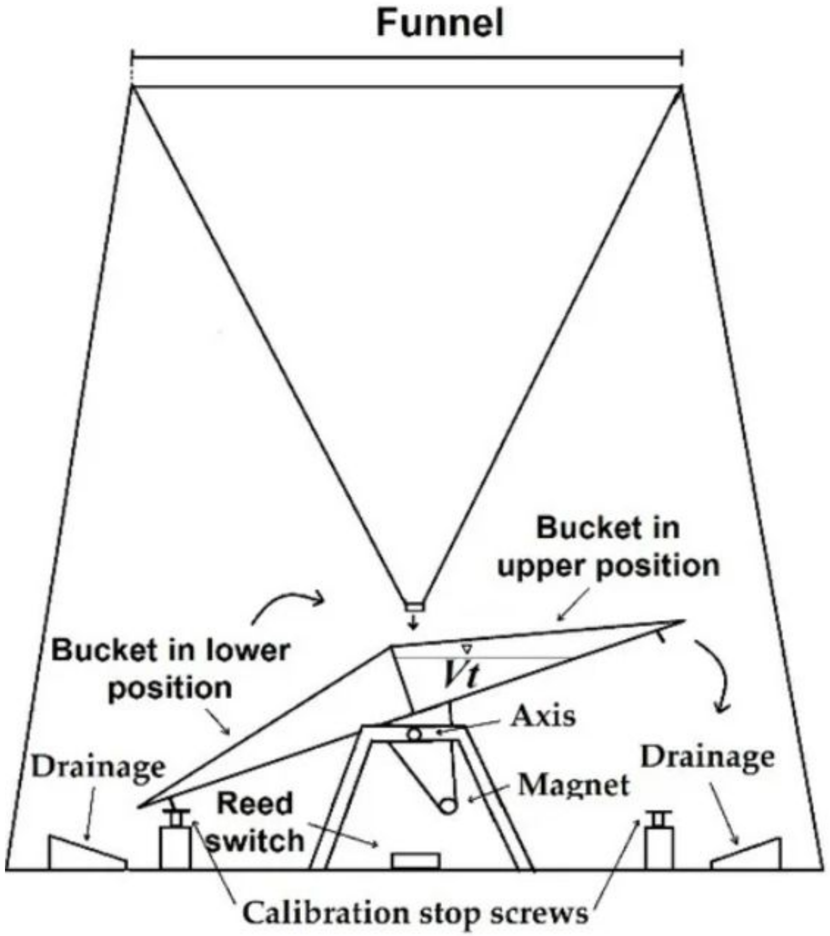
Overview and comparison of three quality control algorithms for rainfall data from personal automated weather stations (PWS)

Dr Tess O'Hara
ntg41@ncl.ac.uk

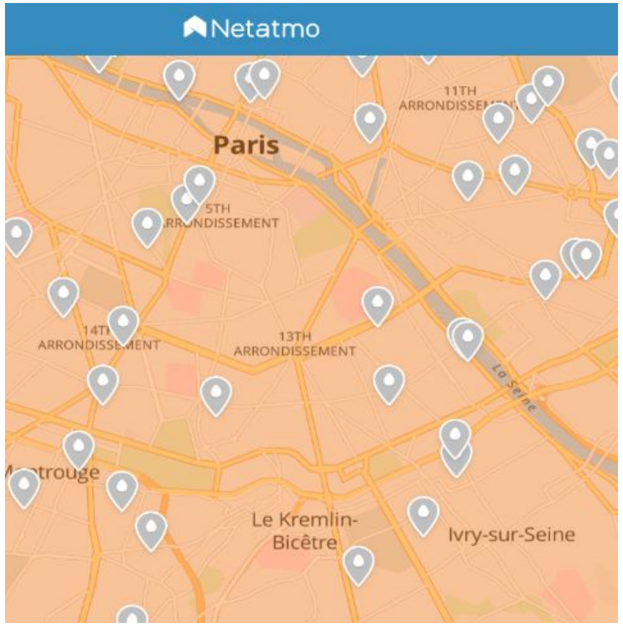
What is a tipping bucket rain gauge?

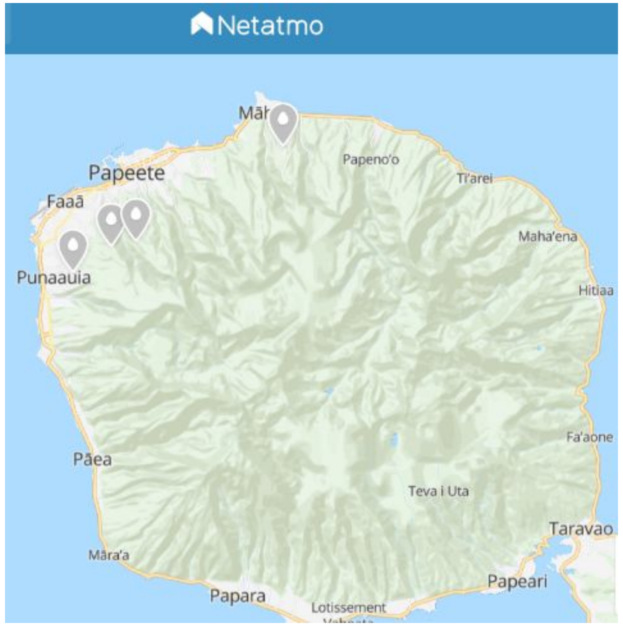
The tiniest “bucket” in the world!



Where are these tiny buckets?

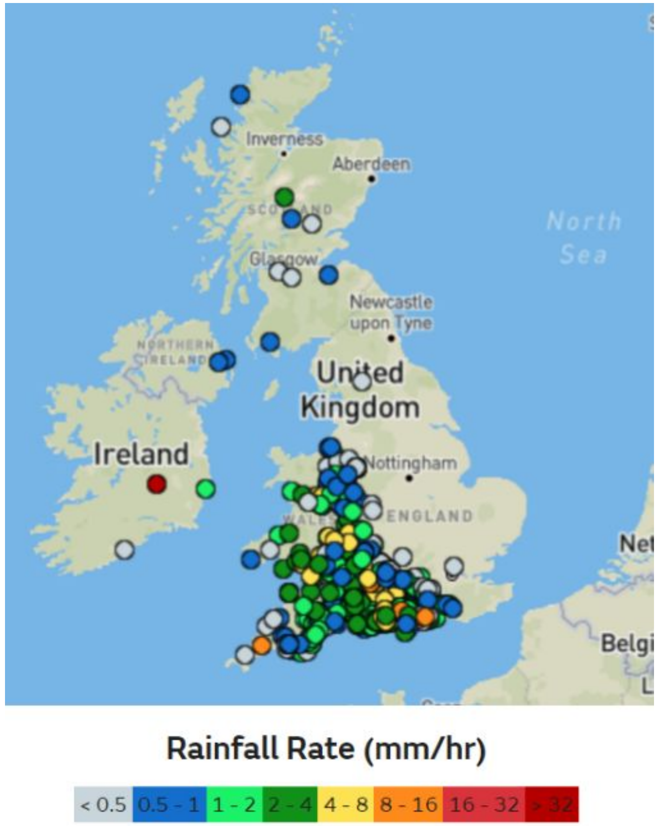
They get everywhere, from Paris to Papeete…

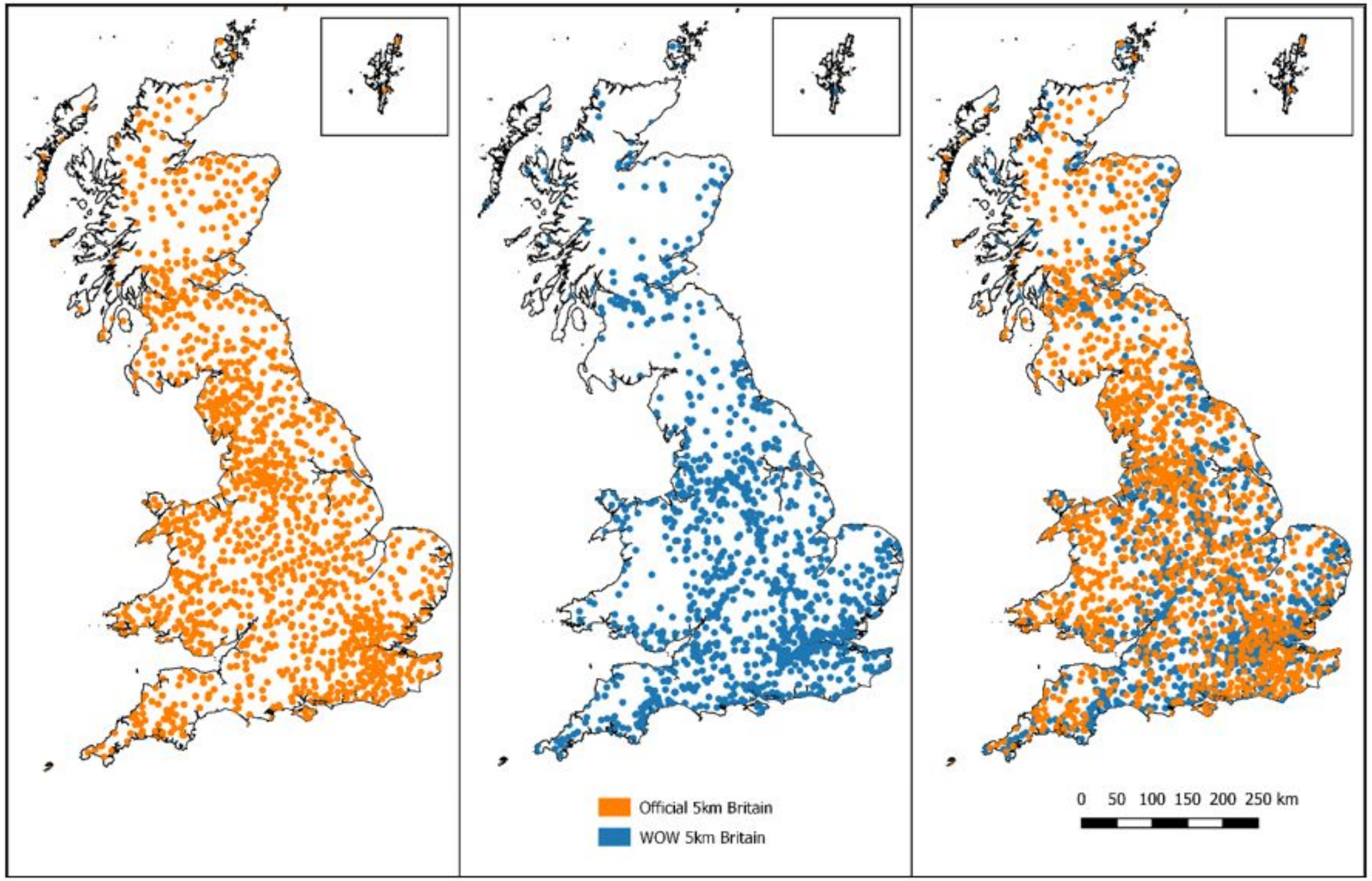




https://weathermap.netatmo.com/

PWS offer an amazing opportunity to obtain a higher spatial resolution rainfall data set.





But are the data any good?

Yes, and no.

Unlike official monitoring, the rain gauges ***may*** not have been installed following WMO Guidelines.

But they can be operated by people who take monitoring very seriously indeed.

Schoolboy train-spotters at Newcastle Station, August 1950.
© National Railway Museum / SSPL

https://blog.railwaymuseum.org.uk/researching-the-trainspotters/
https://wow.metoffice.gov.uk/support/siteratings#locationAttributesRainfall

So, we need quality control?

- Statistical quality control is focused on the output data.
- After meeting via the OpenSense network researchers from Germany, the Netherlands and the UK decided to compare notes on QC methods.
- We wrote a ***Technical Note: A guide to using three open-source quality control algorithms for rainfall data from personal weather stations.***
- We used rainfall data from Amsterdam as a case study.

What does the QC do?

1. PWSQC (Netherlands) - Uses a series of neighbour checks to filter and bias correct rainfall.
2. PWS-pyQC (Germany) – Correlates against reference network then bias corrects and filters erroneous rainfall.
3. GSDR-QC (UK) – Flags suspicious observations against user defined thresholds and neighbour and/or climate indices checks then removes erroneous rainfall.

Quality Control of PAWS Rainfall Data

Newcastle University

	PWSQC	PWS-pyQC	GSDR-QC Local
QC modules	1. Neighbour selection 2. Faulty Zeroes & High Influx filter 3. Station Outlier filter & bias correction factor determination	1. Indicator based filter 2. Bias correction 3. Event based filter	1. Flagging of suspicious observations using defined rule base 2. Filtering of suspicious observations not meeting QC criteria
Reference dataset required	No, but optional part of initialization of bias correction factor determination	Yes, required for 1, 2 and 3	Yes, ETCCDI data plus user defined maximum daily and hourly thresholds
Programming language	R	Python	Python
Ground truth used in method	Median values from neighbouring PWS	PWS should fit in space-time dependence structure of reference data	Neighbouring gauges are compared to each other and optionally against a reference dataset
Level of QC-allocation	- Per measurement	- Per full PWS time series - Event based	- Per individual measurement - Dynamic nature is suitable for longer time series
Output after running QC method	- Original PWS dataset - 3 flag files conveying flag attribution to individual observations for all three QC - 1 file with bias correction factors generated for each observation - Bias adjusted PWS dataset with only reliable observations	- Set of trustworthy PWS - Individual bias correction for each time series - Implausible time intervals removed for each time series	- Flag file for each gauge showing individual test results - Output file with reliable observations

Which algorithm is the best?

That depends, they each have their strengths…

https://www.goodfon.com/sports/wallpaper-sbornaya-germanii-chempiony-4875.html ,
https://www.nytimes.com/2019/07/03/sports/soccer/us-netherlands-world-cup.html,
https://www.standard.co.uk/sport/football/england-1966-world-cup-winners-now-sir-bobby-charlton-hurst-a4495456.html

Quality Control of PAWS Rainfall Data

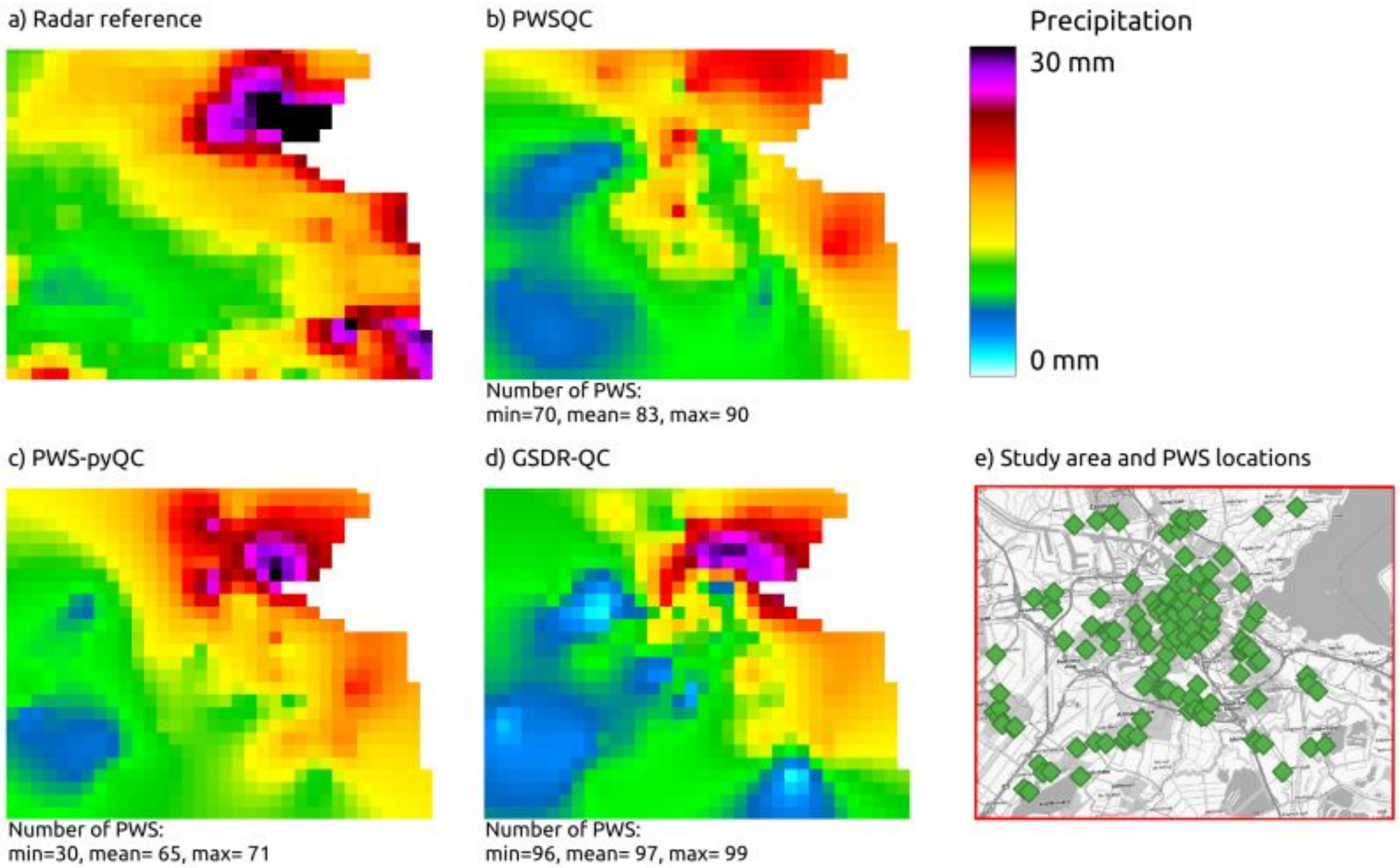
Applicability regarding	PWSQC	PWS-pyQC	GSDR-QC Local
Temporal scale	HI-filter has no lead-up time, but (with default parameters) FZ filter requires 30 min and SO-filter and bias correction require $\geq$2 weeks of data with >100 nonzero intervals. Most suitable for long periods of continuous data.	Time series should be long enough to include significant number of rain events, which is dependent on the climatic region and temporal resolution.	Where neighbouring PWS are available within 50 km there is a minimum requirement of 1 year of overlapping data. Otherwise, where climate indices are available 1 month minimum of data is required.
Spatial scale	Network can span large areas, provided that neighbour PWS values are a good proxy of the ground truth throughout the network. Neighbours are defined by a range around a station which assumes climatological agreement with neighbours in all directions	Due to need for reference set, PWS network has to overlap with reference network. For the indicator filter, the data from the reference network needs to represent the local spatial and temporal rainfall variability, but a temporal overlap is not necessary.	There are no limitations to the spatial scale. Consideration must be given that the same daily and hourly maximum rainfall thresholds are applied on the whole area.
Temporal resolution	5 minute timesteps (or longer)	1 hour timesteps (or longer)	1 hour timesteps (or longer)
Spatial resolution	Due to neighbour checks, most suitable for dense networks	Applicable for both dense and sparse networks	Applicable for both dense and sparse networks
Operational potential	Current version of code works only on static dataset, but theory applies for operational application	Current version of code works only on static dataset, but theory applies for operational application	Developed for static datasets
Approximated runtime for Amsterdam PWS dataset	1. Neighbour selection: flash 2. FZ and HI filter: lunch break 3. SO and bias correction: weekend	1. Indicator correlation: flash 2. Bias correction: lunch break 3. Event Filter: coffee time	1. Create gauge objects: flash 2. Run QC: coffee time 3. Extract QC summary: flash
Impact of PWS network scaling on run time	As whole network needs to be evaluated for each timestep, large dependency on number of stations	Calculation of distance matrices increases nonlinear with number of stations	Linear with number of PWS

Show me some results!

a) gauge adjusted radar accumulation

b), c) & d) post QC interpolated PWS accumulations

e) gauge locations



It looks amazing, can I try it?

The gauge-adjusted radar product from the Royal Netherlands Meteorological Institute (KNMI): https://dataplatform.knmi.nl/dataset/rad-nl25-rac-mfbs-5min-netcdf4-2-0

PWS dataset: https://doi.org/10.1029/2019GL083731

All QC software is open source and can be accessed in the OpenSense sandbox: https://github.com/OpenSenseAction/OPENSENSE_sandbox

And who will help me?

Abbas El Hachem
Jochen Seidel - jochen.seidel@iws.uni-stuttgart.de
Tess O'Hara – ntg41@ncl.ac.uk
Roberto Villalobos Herrera
Aart Overeem
Remko Uijlenhoet
András Bárdossy
Lotte de Vos - lotte.de.vos@knmi.nl




Is there anything else I should know?

Probably too much to mention here... but happy to answer any questions.