EUMETNET Study A1.05
UK contribution Examination of a ‘bias correction scheme’ as a QC tool for PWS

Test Case 16th August 2020

Jeff Norwood-Brown

Mike Molyneux

If printing double-sided you will need this blank page. If printing single sided, please delete this page

Contents

Version

Draft 0.4	24/11/22
Draft 0.6 including comments from Jacqueline Sugier	01/12/22
Draft 0.7 clean version for review in the project team	01/12/22

0 Executive Summary

EUMETNET are currently exploring if citizen submitted weather station data can be used as an additional and reliable source of observation data. There are known errors in such measurements and therefore a Quality Control (QC) tool is required. This study examines one method that might be used and considers general lessons learned from the work. Our intention was to include several case studies to examine the actions of the system, but difficulties in re-configuring the tool meant that our scope was limited to one good case study in the UK. Three areas were considered, one experienced a mesoscale rainstorm and two that did not. We show that the tool is effective when considering temperature and pressure. However, for wind information the observation rejection rate is too high to be useful and we note that further work is required on wind if a reasonable percentage of data is to be accepted.

1 Introduction

In the UK Met Office we use a 'bias correction' scheme for Personal Weather Stations (PWSs), This compares PWSs with Met Office stations (typically within 20km) and creates a timeseries of biases over 72 hours. In general terms, QC is based on filtering out values where the timeseries exceeds limits on magnitude or variability. This software was built to merge PWS values to provide greater detail alongside our managed sites. It has been shown to improve nowcasting of severe events, such as in Reference 3.

Note that in this study we are considering PWSs as belonging to members of the public who are largely contributing out of personal interest. However, there are a range of weather stations that are run by other organisations. It is most likely that these measurements can be processed by the same QC software studied here. These may have different purposes such as road weather.

We also discuss the lessons learned from this work and the wider consideration of a PWS QC tool and consider sustainability.

As a test case, UK WOW and Netatmo measurements were obtained for a thunderstorm event that occurred over Cambridgeshire close to 1500 UTC on 16th August 2020. The PWS measurements were delivered via the EUMETNET sandbox database. It is acknowledged that they were purchased by EUMETNET to support this study.

The data was processed with the bias correction scripts and the output examined.

General points on analysis

What to look for? In this kind of study, it is not entirely possible to predict what good behaviour looks like. What we anticipate is that most readings will fall within a fairly narrow distribution in good conditions. In general, we consider good conditions are consistent weather, terrain without mountains under 400m and Met Office stations typically less than 30km distant. It is also hard to predict what is desirable in terms of a 'observation rejection rate'. This is because it could change depending on the intended use. Broadly, we consider that a rate of 20% is acceptable. However, it is not enough to show acceptable performance in good conditions. We also need to include other events. To do this we choose a period where a mesoscale weather system is included. The tools have allowed us to show this for one good test case.

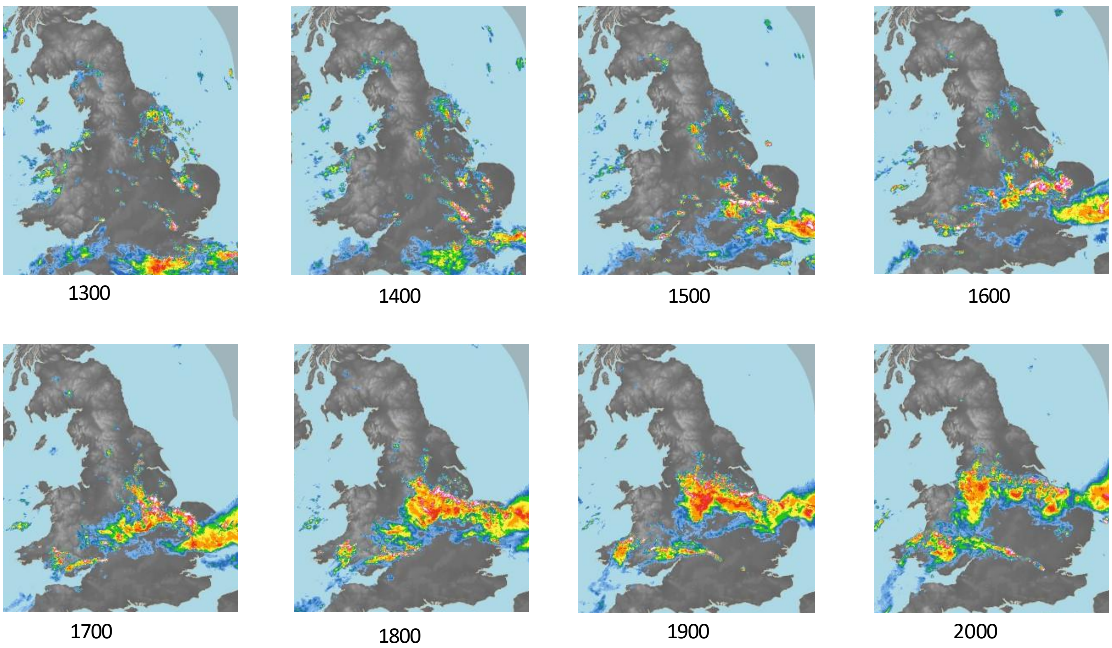
We have found two very useful methods for examination of the results. Firstly, a time series 'spaghetti plot' of the last 24 hours shows very clearly the distribution of measurements and which have been rejected. This method is particularly useful for studying short lived features. This is because all the most recent 24 hours of biases are plotted in one chart. As an example please see Fig 2a within the next section. It shows all the PWS temperature biases both pass and fail. It clearly demonstrates that most stations are closely grouped and only three or four show clear differences from the main group. The storm passes and the relative biases can change but tight grouping remains clear. This is a powerful illustration of the QC method. Most stations agree within approx. +/-1 C which is reasonable for closely located stations in flat terrain.

Secondly, we have examined the location of PWSs (especially those that are outliers) using Google Earth. This has proved to be quite useful in identifying why some stations have poor performance. Examples are given below. This is not intended to be a routine part of the system, which needs to be automatic. However, we believe it is important to try to identify why some stations have worse performance than others, especially at the study stage. This may be inferred from some images of the location. QC should be about more than just statistics!

2 Test Case

On 15th August 2020 a line of thunderstorms/squall line developed and passed overhead Cambridge between 1500UTC and 1900UTC. 24 hours' worth of output from quality-controlled data was obtained for analysis. Figure 1 shows the rainfall radar plots of the period we examine.

Fig 1 Rainfall Radar 16/8/2020



A test area and two control areas were established. With Cambridge being the area affected by the thunderstorm, the two control areas (Edinburgh and Liverpool) were chosen as being of similar geographical size as the Cambridge/Cambridgeshire area, a similar population and/or having a similar density of reporting citizen weather stations. Neither control areas were affected by thunderstorms on the day.

Map 1 – shows location of test areas

Cambridge box co-ordinates

Lat 52.30, Long -0.05	Lat 52.30, Long 0.22
Lat 52.10, Long -0.05	Lat 52.10, Long 0.22

Edinburgh box co-ordinates

Lat 56.0, Long -3.30	Lat 56.0, Long -3.13
Lat 55.8, Long -3.30	Lat 55.8, Long -3.13

Liverpool box co-ordinates

Lat 53.5, Long -3.07	Lat 53.5, Long -2.9
Lat 53.3, Long -3.07	Lat 53.3, Long -2.9

Temperature, pressure and wind speed & direction were examined to see how the thunderstorm squall line affected the measurements.

2 a Temperature

Temperature biases plotted using Power BI. Power BI is a convenient method of exploring the area, range, and shape of the timeseries data. The chart (fig 2a) below shows the traces of 51 stations in the Cambridgeshire test area. Each line represents the calculated temperature difference (bias) between the PWS and the closest Met Office official AWS. It has been found during development that temperature traces are best considered using biases to nearest station rather than a normal temperature. This is because diurnal differences are common in temperature and that complicates the diagram. If all the stations had agreed, then the ideal plot would be horizontal and at zero. Figure 2a shows all the stations pass or fail and in contrast Figure 2b shows only stations that passed QC. We see that stations are generally showing a small warm bias. This is reasonable and probably due to use of smaller radiation shields and locations being more domestic than our managed sites. Additionally, there appears to be a reduction of bias values towards the passing of the squall line (1500 UTC). After the event has passed, the temperatures begin to return to previous (morning) values. Further investigation is required to define why this happened, but a hypothesis is that cloud cover, associated with the thunderstorm, reduced the biases and then they increased once the cloud had passed. This could explain why different weather stations 'bounce back' differently with site and situation being a factor in the overall measurement discrepancies and the different rates of change.

Fig 2a Bias timeseries Cambridge all sites

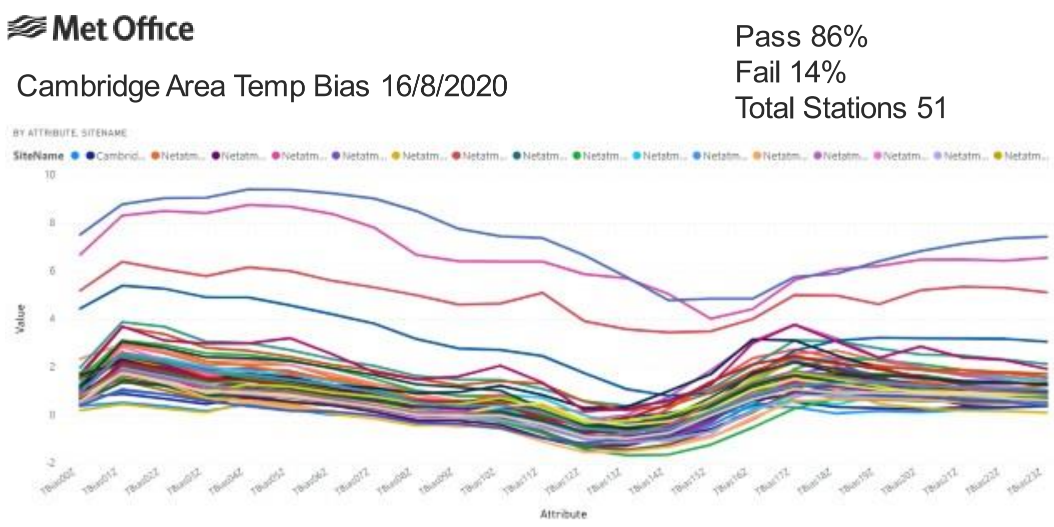
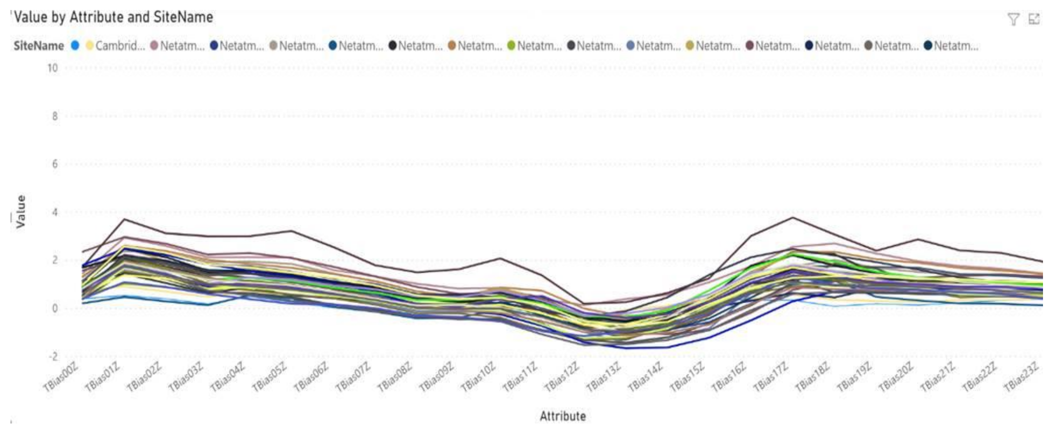


Fig 2b Bias timeseries Cambridge fails filtered out



The Edinburgh control area showed no dip/improvement during the day but did show far less cohesion during the evening. No clear explanation has been found for this (possibly low sun impact) but note it indicates that small offsets can be passed by QC. Further investigation will be required to establish the cause of this but, again, cloud at the control (SYNOP) site could be a factor. A couple of the outliers on this graph were securitised in detail (see below) but no conclusion has been established.

Fig 3a Bias timeseries Edinburgh all sites

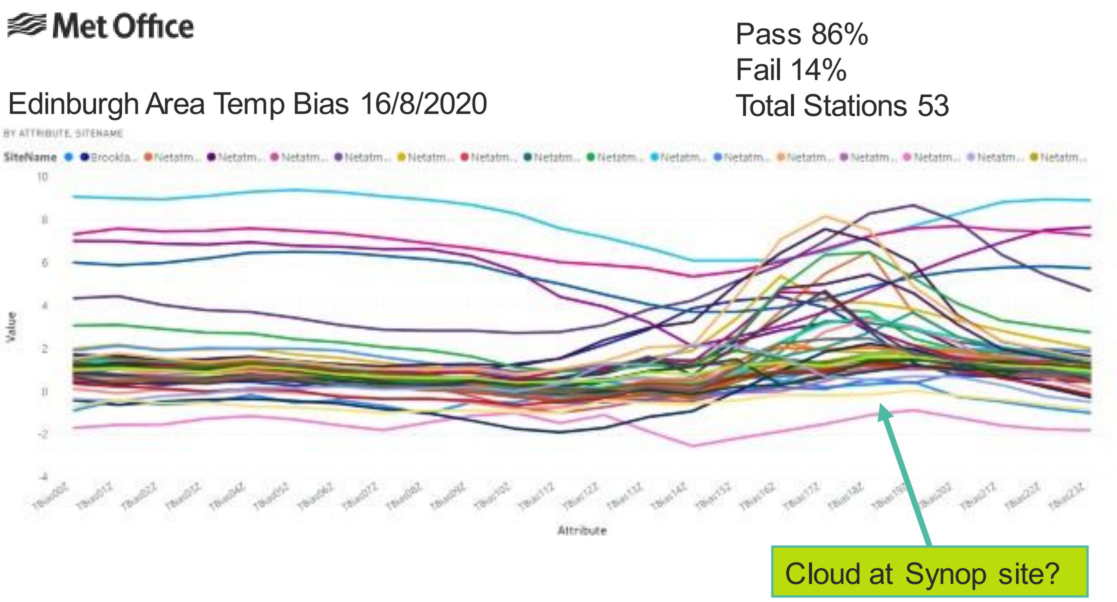
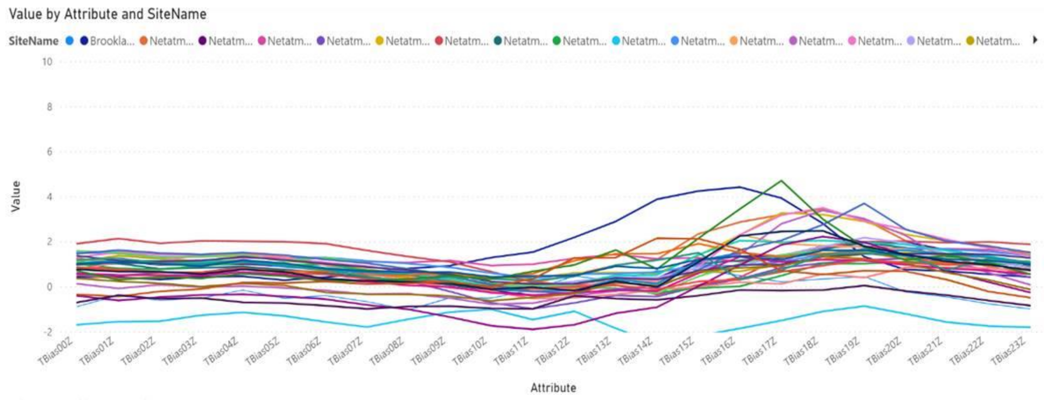


Fig 3b Bias timeseries sites that failed have been filtered out



Furthermore, the second control area, Liverpool, also showed a spreading of values during the evening. Around 0500 rain in the form of showers passed overhead Liverpool and caused a similar dip in bias values. However, these were now underreading compared to the local SYNOP stations. It has not been possible to establish why this is, the initial guess is that rain and cloud affected the SYNOP station at a later stage but, again, further investigation is required.

Note that in comparing Figures 3a and 3b it can be suggested some small errors are not being removed by the QC process. They are the outliers from the tightly grouped stations. They could be as large as +2 degC at times. We suggest that this is acceptable overall. However, we therefore suggest that measurements supplied to a user after QC should have a small explanation added (such as a terms and conditions of use relating to quality). This is important so that Users can be made aware of limitations.

Fig 4a Bias timeseries Liverpool all sites

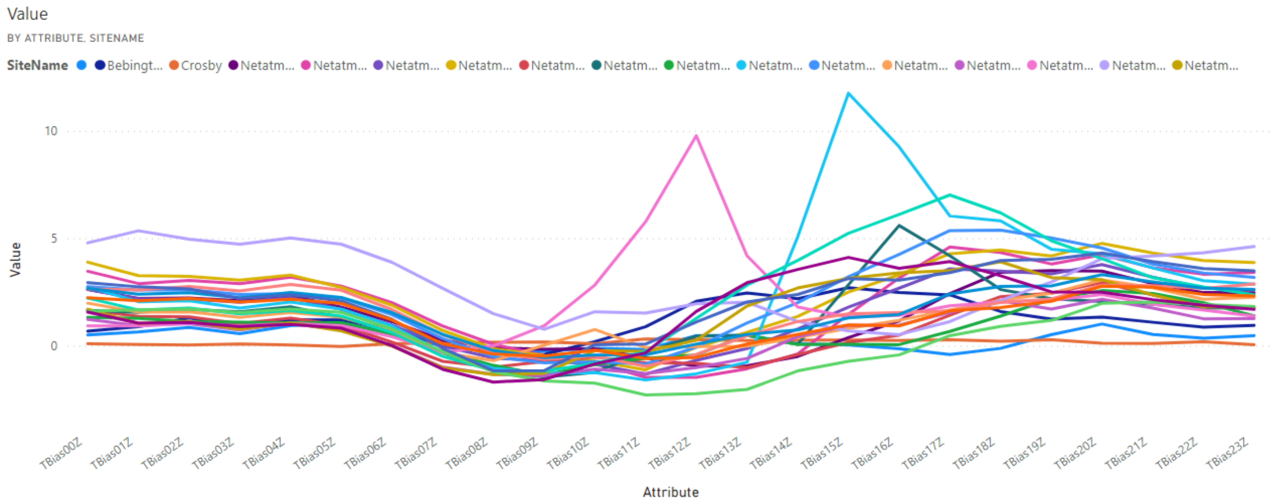


Fig 4b Bias timeseries Liverpool sites that failed have been filtered out

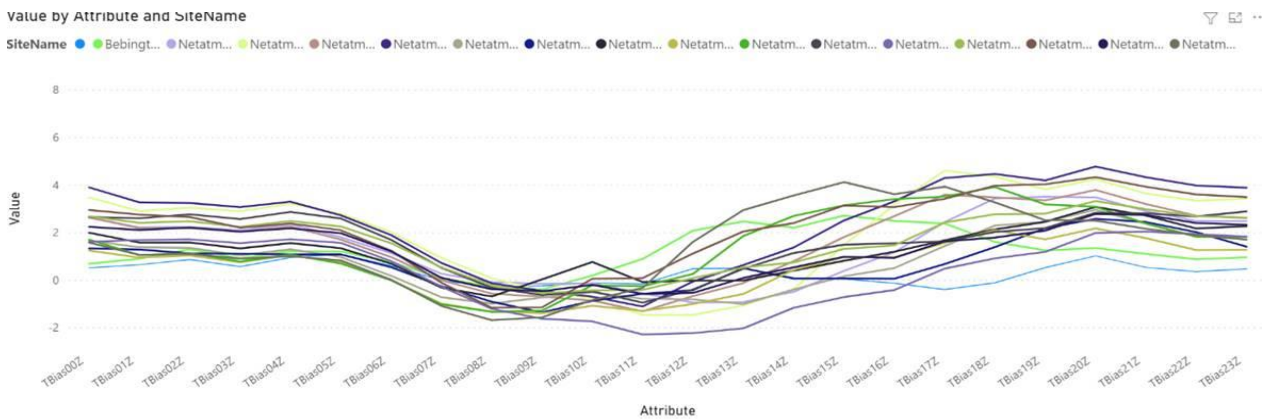
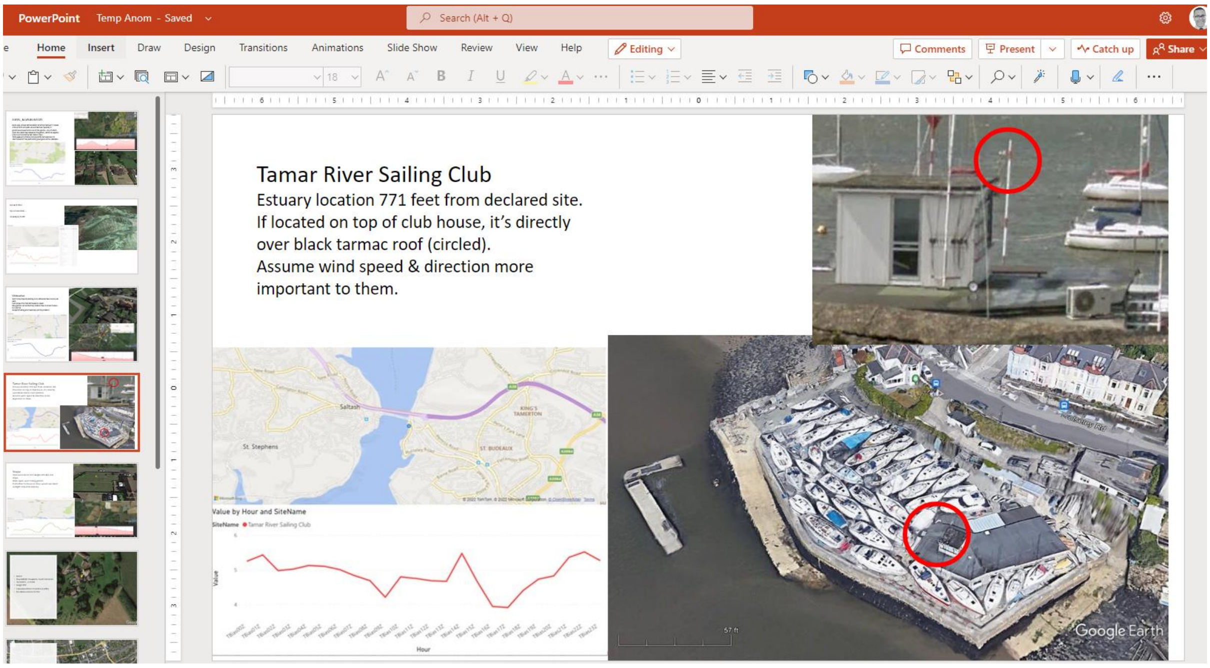
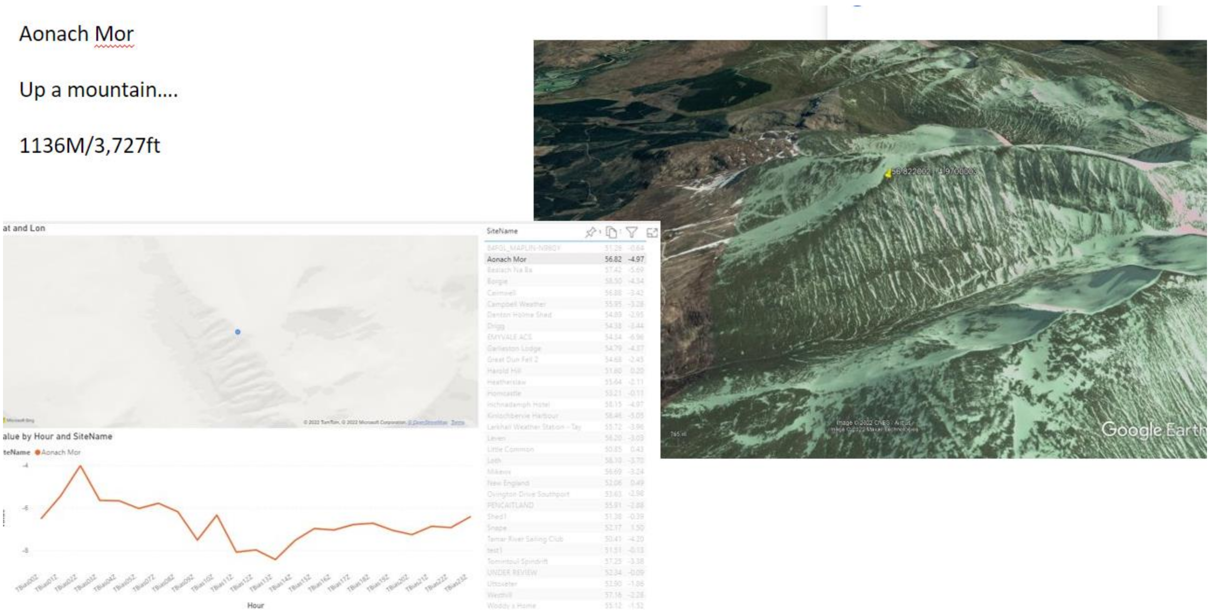


Illustration of some stations rejected for temperature

These stations were found in some initial examinations and are not in the test case. However, they do show some pitfalls with the variety of stations in use and why they could be rejected. One of the stations is managed by the UK MO but is used to measure mountain conditions. Since we use no 'geography' filter, this station was an outlier in the distributions.





Conclusions for temperature

In the test case used, we study a moving mesoscale feature where conditions change rapidly. In the times before and after, we also have conditions where the changes were less rapid. This makes an ideal test case - we can illustrate the spread of differences between stations before, during and after a period of change. We find that large percentages of stations pass QC checks in a range of conditions, and these can be seen to be tightly grouped. Some anomalies are shown but the majority of stations seem acceptable. We believe this is evidence that the system performs well in case of good conditions and during periods of high change due to small features.

2b Pressure

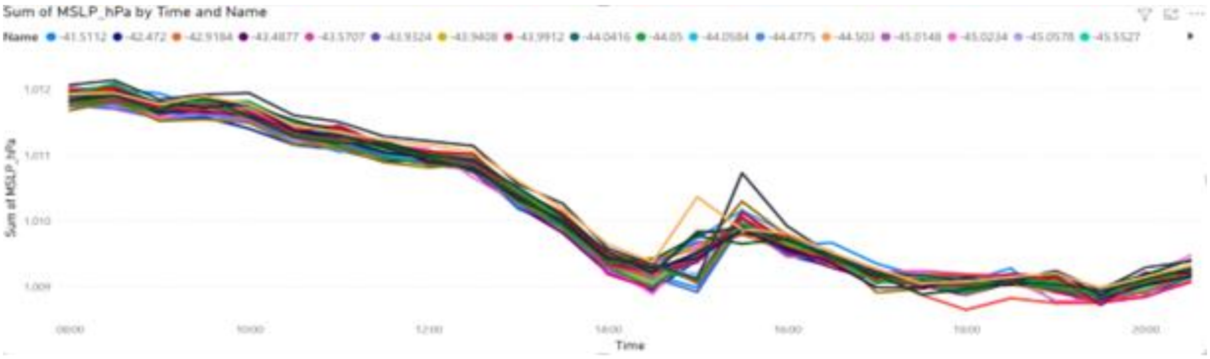
Using the output from the quality control script, 24 hours' worth of pressure data were obtained. In this case the pressure readings are actual values in hPa. Pressure values do not generally show a diurnal variation and the normal values can be plotted and easily studied. The method for obtaining these data is described below.

The resulting chart shows a steady decrease in pressure during the morning which continues after the storms have passed. Between 1230 UTC and 1430 UTC the pressure begins to fall more quickly, then sharply rises to re-establish the morning trend just before 1500 UTC. This dip occurs as the storms approach and pass overhead.

This sudden change in the rate of pressure change could well be used to detect the approach of a storm system.

Fig 5 Pressure timeseries Cambridge

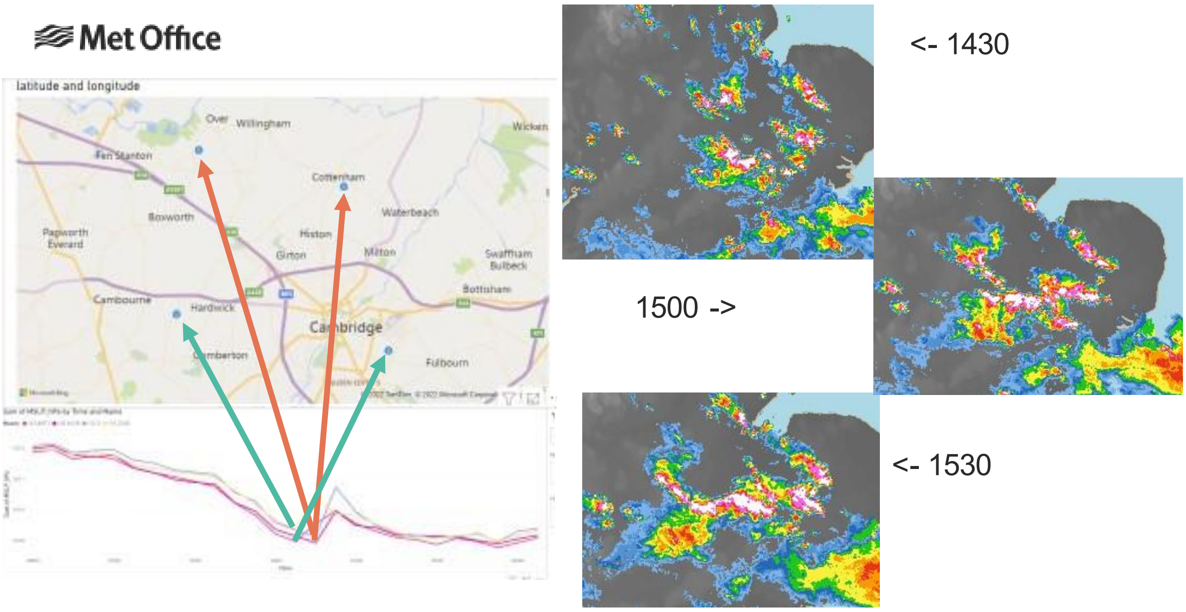
Cambridge Area MSLP 16th August 2022



It can be clearly seen that for most of the time all the stations are tightly grouped. This shows that the system can be expected to work well for most station most of the time. It is worth noting that not all traces react at the same time. It has been established that the weather stations that reacted later were to the North-West of stations that troughed first. This ties in with the storm moving in a South-East to North West direction. The total diagonal distance of the area monitored was 30 kilometres and reveals that speed and direction of the squall line can be inferred from these plots.

Fig 6 Synoptic analysis Cambridge

Image showing the location of the 'early' and 'later' reports of a pressure change.



By comparison the two control areas, Edinburgh and Liverpool, showed no such rapid increases in rate of pressure change.

Fig 7 Pressure timeseries Edinburgh

Edinburgh MSLP 16th August 2020

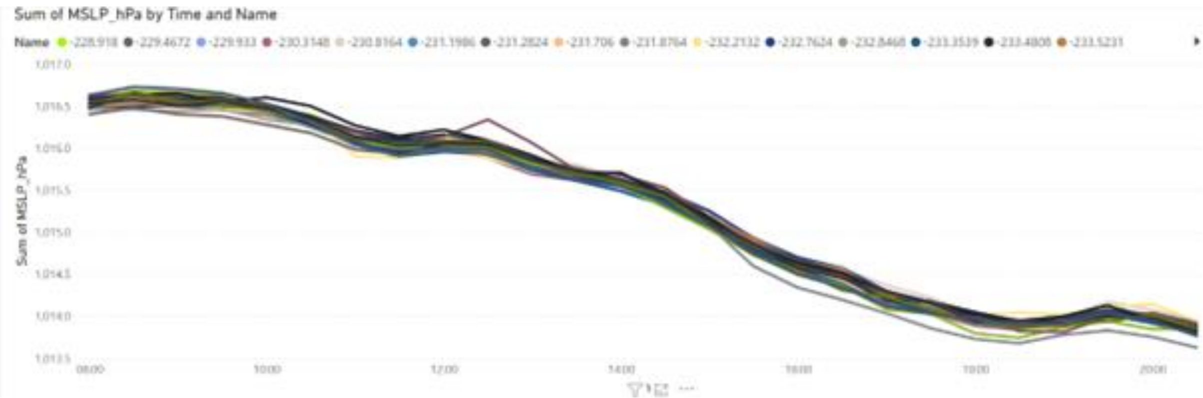
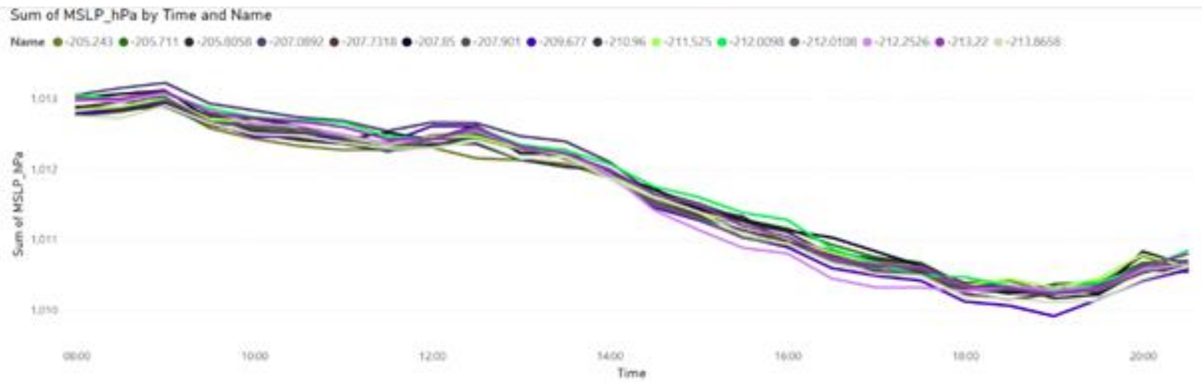


Fig 8 Pressure timeseries Liverpool

Liverpool MSLP 16th August 2020



Met Office

Site by site examinations Pressure Anomalies

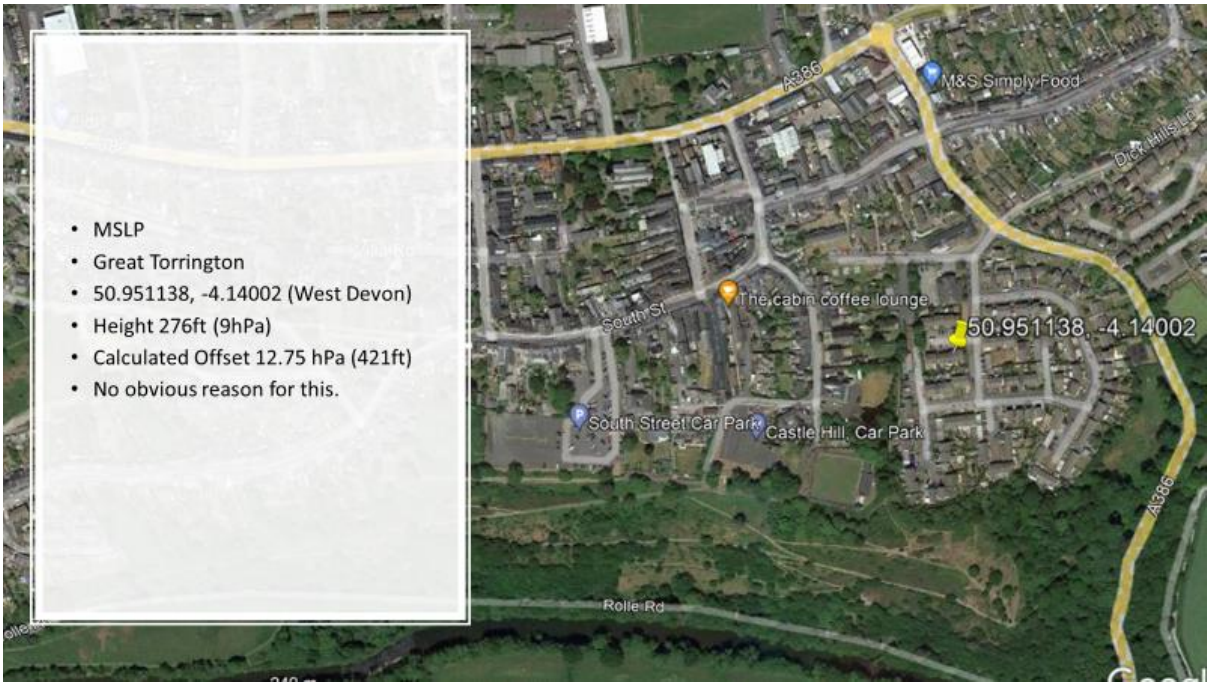
We looked at the location of rejected pressure sites.

In these figures we show the sites of pressure rejections and suggest a reason for rejection. In some cases, no reasons were found, these are likely to be poor equipment or set up. Note that QNH/QFE is a reference to pressure reduction with height methods in aviation. We use this as shorthand where we are concerned that either the wrong station height has been entered or the method used to reduce pressure to MSL is not exactly as expected. Both of these issues are relatively easy to encounter in PWSs. Additionally in these cases it is probably acceptable to correct a standing offset.

Met Office

• MSLP
• Campbell Weather (Edinburgh)
• 55.9459, -3.2776
• Height 296ft (8.9 hPa)
• Calculated Offset 10.4 hPa (343ft)
• QNH/QFE error?
Carnmuir Rd
Corstorphine Hill Ave
55.9459, -3.2776
Gordon Rd
Belgrave Rd
794 ft

• MSLP
• Ratlands (SE London)
• 51.379272, 0.092140004
• Height 252ft (7.6 hPa)
• Calculated Offset 9.47hPa (313ft)
• QNH/QFE error?
Bicknor Rd
BROOM HILL
51.379272, 0.092140004
728 ft



Conclusions for Pressure

The plots for pressure show tight grouping and good agreement in shape especially in the cases where no mesoscale features are present. This demonstrates well that the QC system can be used with confidence. The only point to note is that we believe care is needed for fast-moving small-scale features where a 'phase difference' can be seen but it was not rejected. There is some variability of stations within the general spaghetti plot. In general, this not a major concern. One reason for this could be wind impact, WMO recommend the use of a 'static pressure head' to reduce wind impact but these are unlikely to be found in

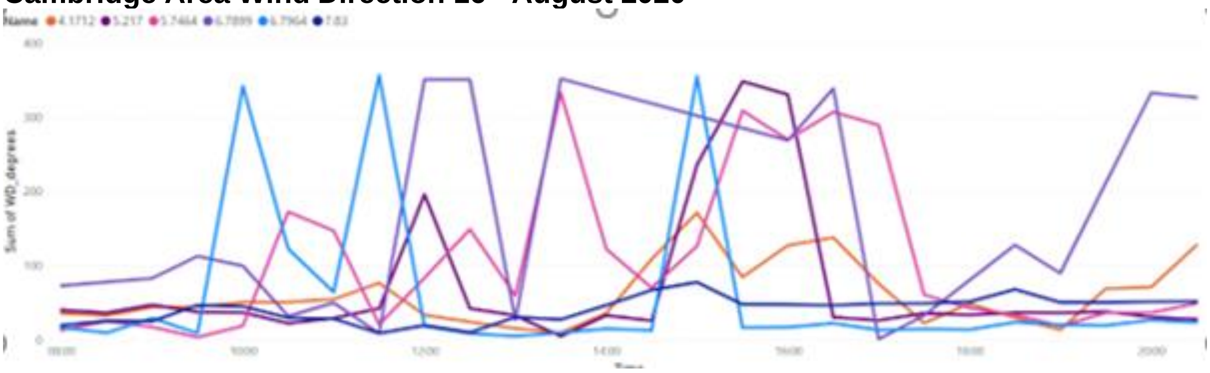
PWSs. It is worth noting that the ‘rate of change’ of pressure in the past was widely used in synoptic meteorology and that there could be additional value in looking at PWS rate of change of pressure.

2c Wind Speed and Direction

After analysis it was decided there were too few citizen anemometers to form any inference from the data. The ones that were available showed no cohesion in their data. This is almost certainly due to the difficulties of anemometer exposure. To obtain good quality wind measurements the WMO recommendation is that a 10m mast is used with few significant obstructions. In a small well-developed environment such as a private garden, it is unlike that most PWSs will be able to achieve this and therefore major errors will be introduced. The wind flow ‘bending’ and turbulence at most sites will be severe. These large and variable errors will not be easy to correct. Note there very small number of sites plotted in contrast to temperature and pressure.

Figure 9 Wind timeseries

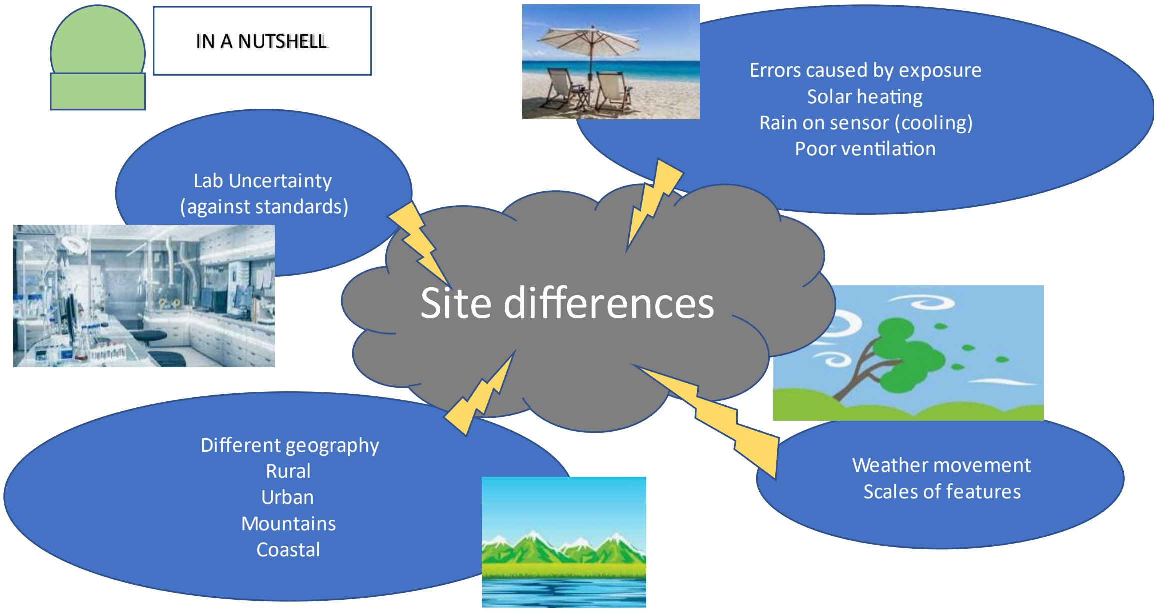
Cambridge Area Wind Direction 16th August 2020



Conclusion for Wind speed and direction

At the time of writing, we do not believe this method is very useful since good results are at a low percentage. We recommend that further work (outside of scope here) is carried to investigate this problem.

3 Discussion of the implementation – General Lessons



As a simple graphic this figure shows the 4 key sources of differences between nearby weather measurements. Allocation of these differences should be our long-term goal for measurement and QC

3a Sustainability and simplicity

We have demonstrated that the methods we have studied can be considered a good and workable version 1 of this bias tool for temperature and pressure. We have only been able to include one case study. Therefore, it may be possible the software would not perform so well in other types of events and at other locations. Additionally, we should also consider that any QC tool has to mature and be sustained for many years. This brings up some questions on what will make a QC tool both effective and at the same time easy to sustain in the long term. In order to sustain the tool, it should be as uniform as possible. In fact, the whole life of sustained science support and improvement should be considered from the outset and careful decisions made.

If we consider what other low-cost measurements might require QC in the near future, we speculate on requirements for boundary layer wind values (from drone flights) and precipitation type (from connected vehicles). As the complexity increases in this way the tool needs to use techniques that are similar. This will be important, even if the tools remain unchanged, input from new contributors will need to be tested and quality understood.

We suggest that careful consideration is given to the use of ‘O-B’ techniques in a future development. This QC method is widely used and compares O (observations) with B (background from NWP). While it may bring slightly different performance, its key advantage would be that the methods can be much more uniform as many more parameters will be available in the model background. This means that the science documentation and staff resource is likely to be cost effective in the long term.

Our system at considers a total bias from all the sources of difference (such as shown in the graphic above). It does not attempt to attribute the differences to known causes. For example, radiative heating errors in temperature are not distinguished from differences arising say from a cooler location on the coast in summer. In measurement best practice we should produce measurements and an assessment of uncertainty. This may not be easy for weather station measurements (see reference below) but could be considered a long-term goal.

If we were to calculate uncertainty components against each of the four principle causes of error in the diagram above, (by whatever method) the QC tools would then simply select the quality required by the user.

However, these are not simple issues and proper thought should be given to the design of the software the science expertise needed to extend and test it over the life of the system.

3b Some general features of a QC tool set

It should
Provide consistent data. That should include a short description of quality (see 3d)
Analyse test cases of past events (for R&D or system testing)

It is highly desirable that it includes

- Time persistent flags (example 3 rejects in a month and a station is 'out' until a system manager puts it back)
- Monitoring and reporting. (We need to be sure that the system is working, so it needs to output some stats on performance)
- Manual blacklist

It could also
Include a **'**Superstation list' some sites that have been shown to be consistently good quality

Note that several of these imply we should allocate staff effort for system management.

3c Assumptions made between 'managed' and 'third party' measurements

In most applications used in weather station QC we make assumptions about the features of the measurements we are controlling based on careful management by National Met Services so that

1. Sensors are selected for good performance and are well exposed to the atmosphere
2. Sensors are maintained in good working order
3. Sites are selected to be (mostly!) in open natural environments and the influences of other factors are minimal. For example, reflected radiation, warm walls and sources of heat like air conditioning units.

We can see that these assumptions cannot be carried forward into all our new cases and that our measurements and techniques may be challenging.

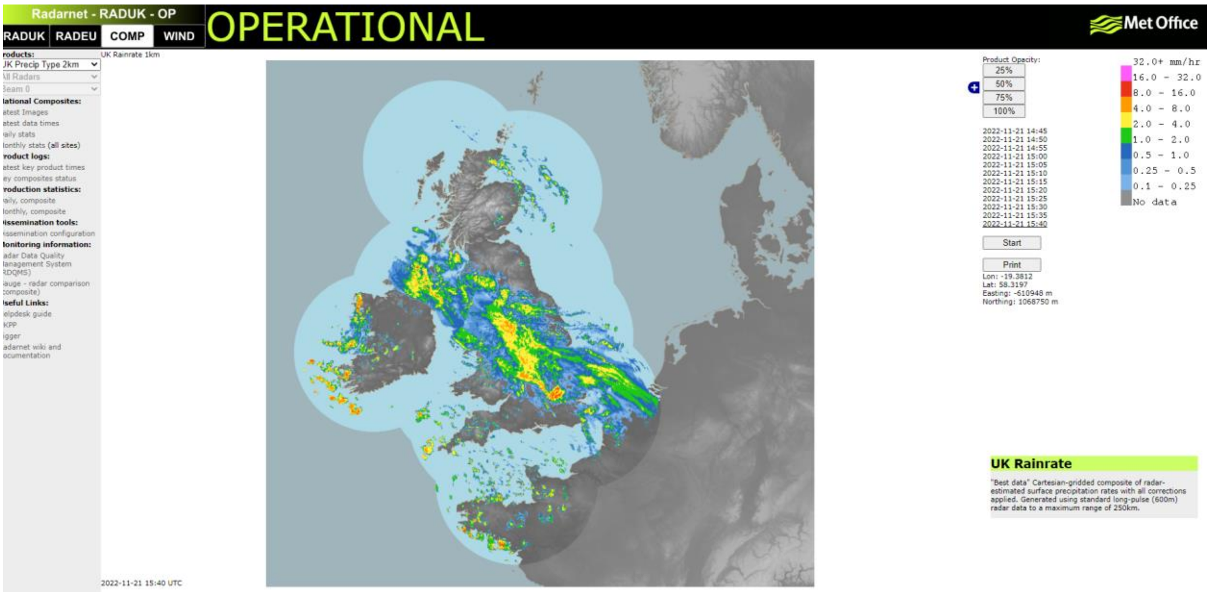
1 We cannot dictate the design of instruments and shelters used in PWS networks and this may lead to new error characteristics. We can expect QC to remove large errors, but we may find it hard to identify small offsets, especially if they are common to a network. For example, if a direct radiation error is adding 20C to air temperature it can be removed easily. However, if a night-time error is 1C due to a nearby warm wall then it is most unlikely that it can be easily separated from a naturally warmer location. Therefore, values may be passed forward including the error. In addition, some assumptions in instrument performance can be difficult when moving into an urban environment. For example, the shades on a radiation screen tend to anticipate radiation mostly from 'above', whereby in an urban environment the radiation may be often coming from below (reflections from windows or cars).

2 Sensors are unlikely to be maintained. This may not matter for temperature sensors where sensors typically stay close to specification for many years, but more commonly low-cost humidity probes may drift at 1% per year and so may be quite poor after several years of use.

3 In the UK managed sites are mostly at low levels and the topography is not very challenging compared to Alpine or Nordic countries. Our tool has not been well configured for more challenging places, for example we found early on that one station that was often rejected was one of our own high-altitude stations! It was, therefore, well managed and run, but since the nearby stations were generally much warmer, it was consistently rejected.

3d User self-service – quality or quantity

We believe that a future system should anticipate different demands by users. For example, we may consider that a user may select quality data in low quantity or the other way round, low quality in high quantity. In fact, the UK weather radar system already anticipates this in a graphic form by offering 'Best' or 'Low FAR' (False Alarm Rate) outputs. For the user to make sensible choices, it therefore requires a simple but well described rationale as shown in the examples here. Firstly, we show an overview of rainfall graphic with simple description of the output bottom right



Secondly here are descriptions of the contrasting outputs for precipitation, illustrating how users can be presented with predefined options to suit their needs in a simple short paragraph.

UK Surface Precip Type 2 km (uses NWP)

Best estimate of the type of precipitation falling at the surface. This product uses the radar-based surface rain rate estimate, as well as the NWP wet-bulb freezing level height forecast, to provide an estimate of the type of precipitation falling at ground level. Hail is indicated in high probability cases only. For more information on the algorithm used to generate this product, see the corresponding scientific documentation page.

UK Low FAR Products

The low false alarm rate (FAR) product is a custom rain rate product with very specific users. It is generated from the same data as the standard UK rain rate composite, but with stricter quality control. This minimises the breakthrough of false echoes into the radar composite, but also risks classifying some precipitation as non-meteorological.

5 Concluding thoughts

We have examined the bias calculation schemes and rejection processes developed for nowcasting as a QC tool. We have shown that for temperature and pressure most PWS values pass QC and are closely grouped even during a severe rainfall event. This shows that the tools will work well in a variety of situations. This means that PWS value can be well used as additional data after QC. However, we find that wind values do not conform so well, and the QC process may need further work. We identify some cases where users may need to be aware that the QC process is passing smaller errors, but these are probably acceptable most of the time.

We also discuss some lessons learned as a result of focussing on the tools in a more general way. We suggest that care is needed so that in future the selected QC system has the right balance of ‘fit for purpose’, long term maintainability and adaptability.

6 References

1 Meteorological Applications. Fine-scale analysis of a severe hailstorm using crowd sourced and conventional observations.
Matthew R. Clark, Jonathan D. C. Webb, Peter J. Kirk 2018 2

2 Thesis Summary: Quantifying Uncertainty in Citizen Weather Data.
Simon Joseph Bell. Available at Bell_Simon_J._2015.pdf (aston.ac.uk)

3 Towards simplified expression of uncertainty in meteorological observations
S Bell, M Molyneux, S Beardmore, M Clark
National Physical Laboratory, Internal Report

7 Annexes

Annex 7a Flowcharts of the method

Fig 10a Temperature – Bias calculation and rejection described as flow charts

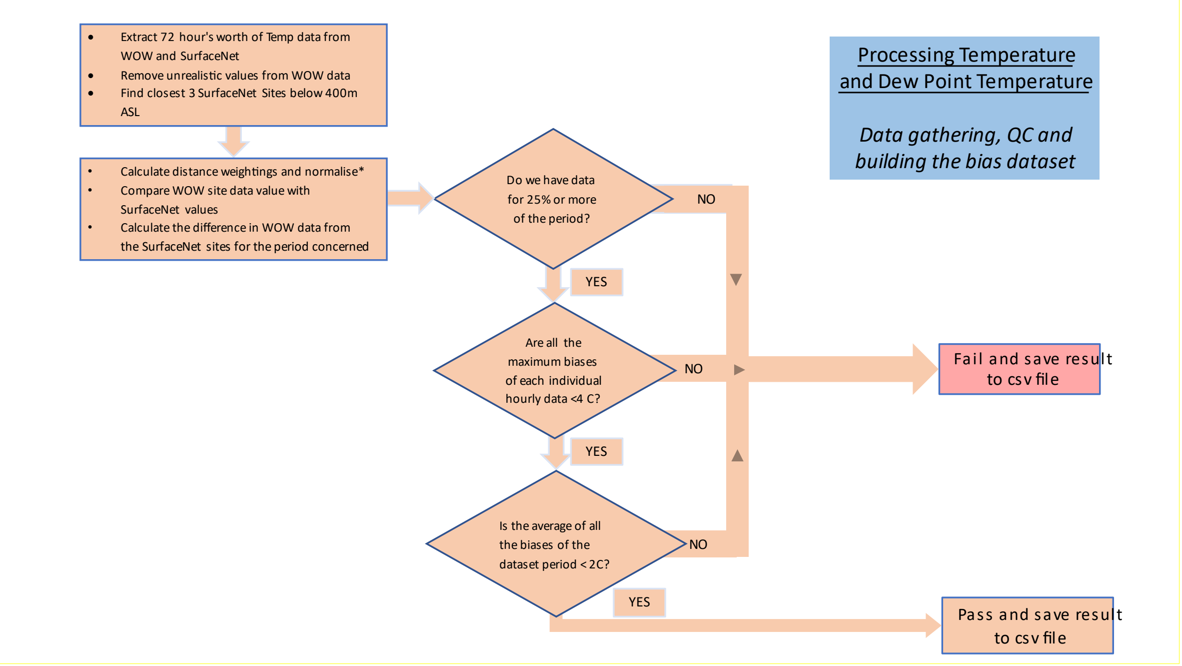


Fig 10b Temperature – Application of processing results to current data

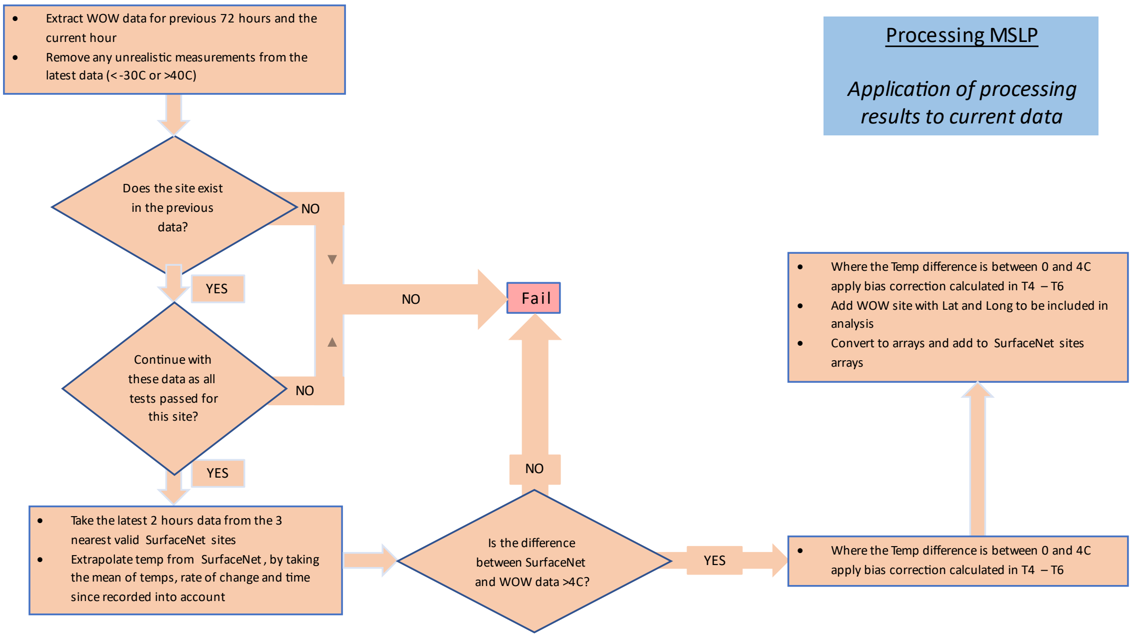


Fig 10c Weightings calculation visualisation

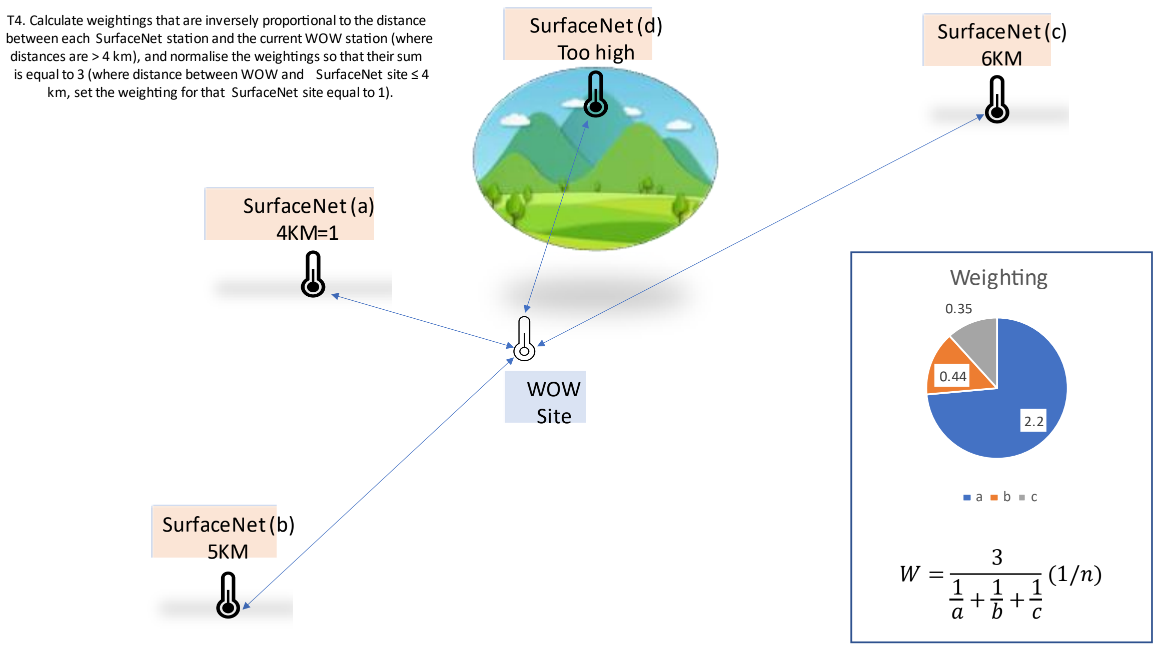


Fig 11a Pressure – Data Gathering

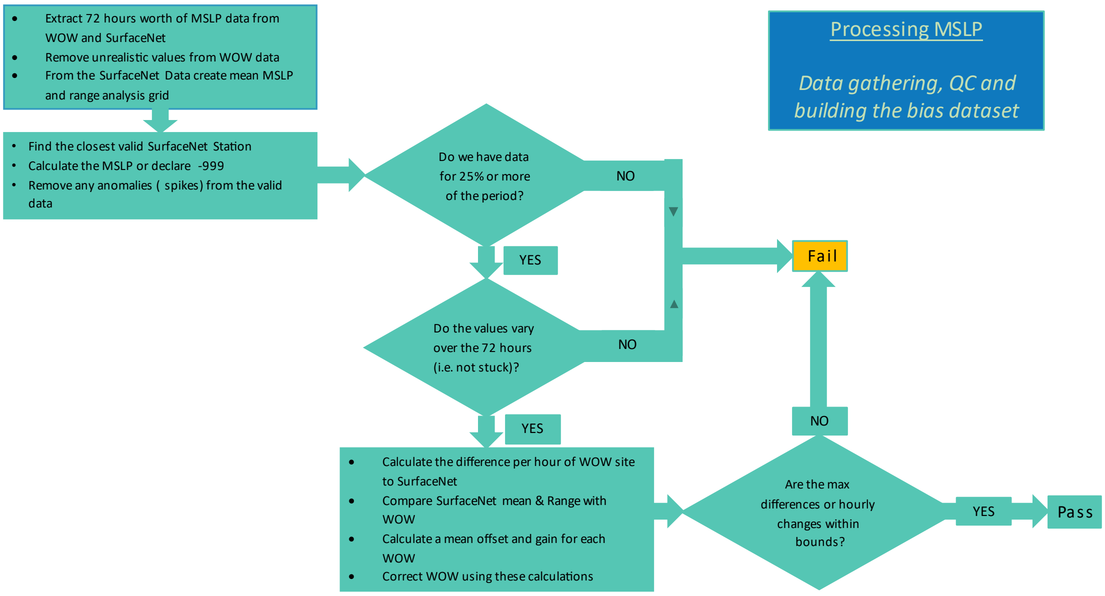


Fig 11b Pressure – Application of processing results to current data

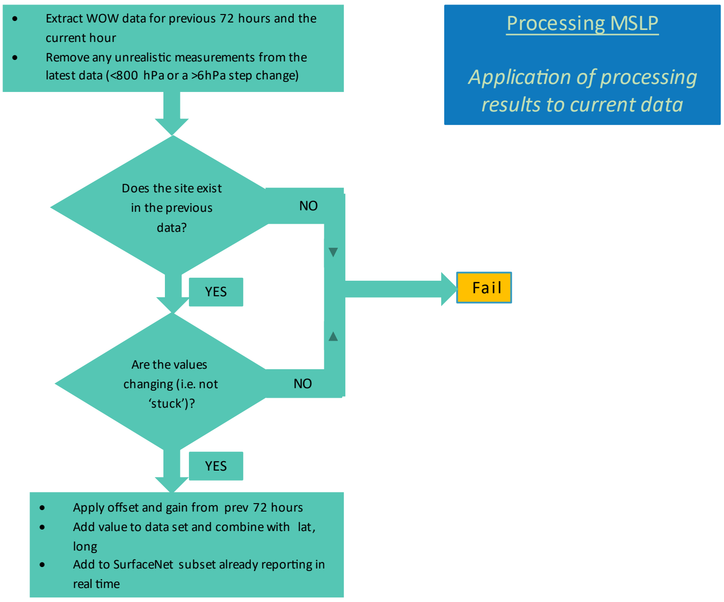
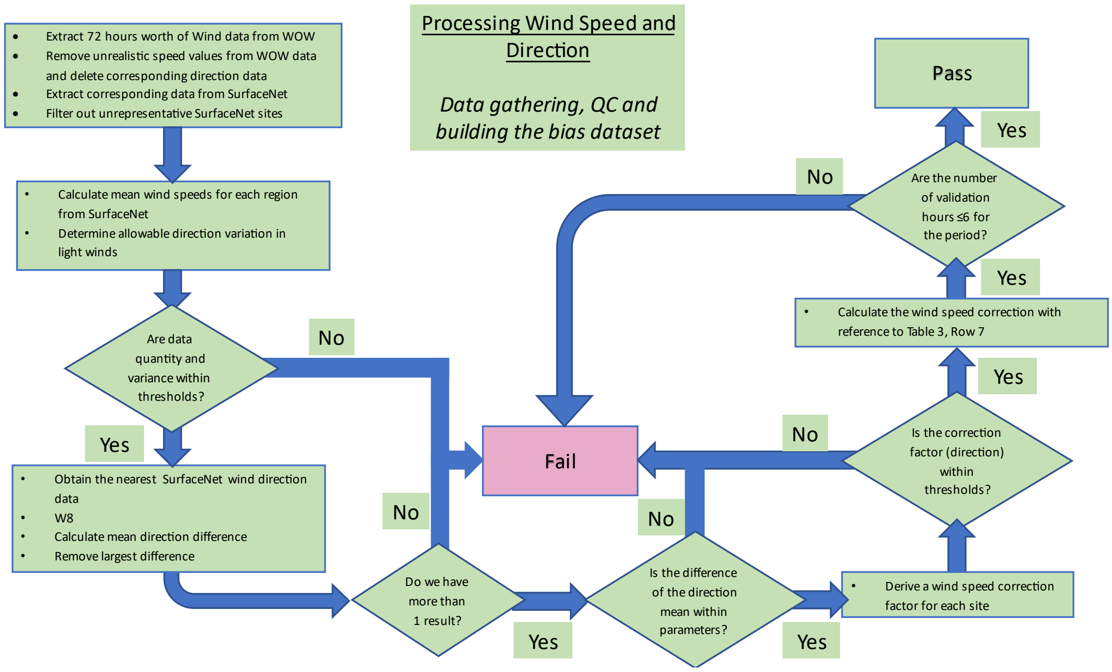
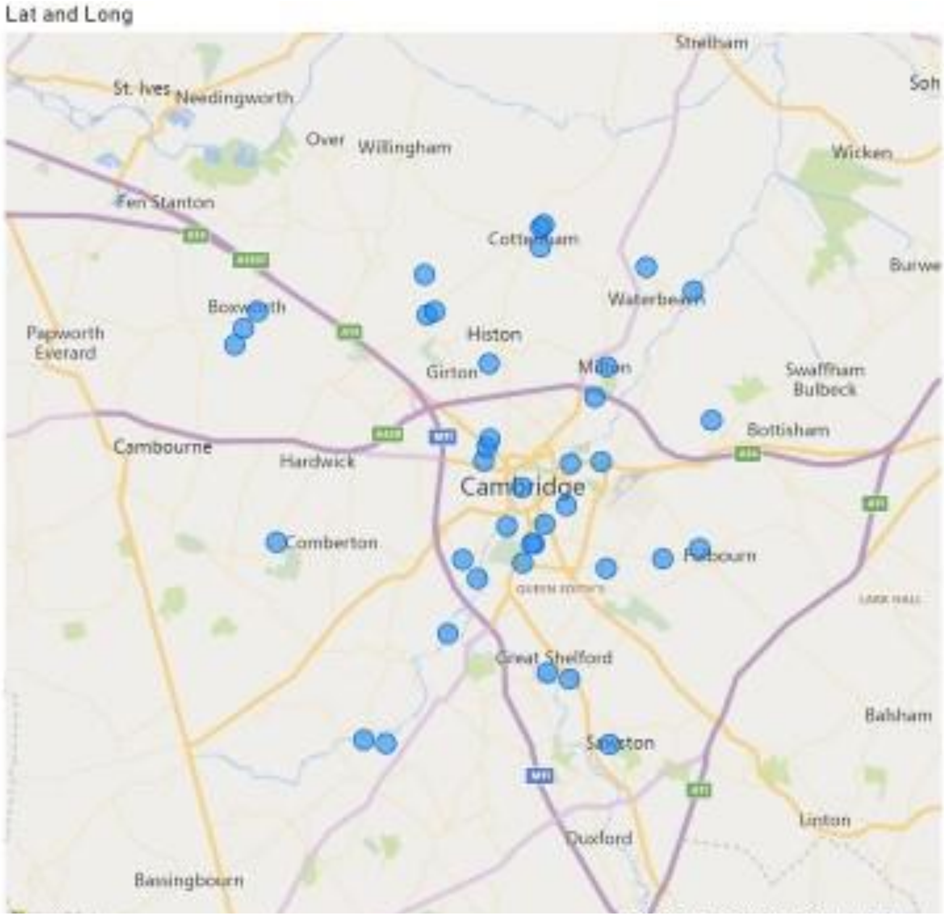


Fig 12 Flowchart Wind Speed and Direction – Data Gathering



Annex 7b Location of Cambridge Area stations.



ID	Name
4475	BARRINGTON
4496	BOTTISHAM LOCK
450	BOXWORTH, HOME PADDOCK
4427	BOXWORTH, MANOR HOUSE
451	BOXWORTH, SAMSON FIELD
4487	CAMBRIDGE
18993	CAMBRIDGE GUILDHALL
4490	CAMBRIDGE P STA
4494	CAMBRIDGE S WKS
4485	CAMBRIDGE, 33 BARROW ROAD
4486	CAMBRIDGE, BARROW ROAD
454	CAMBRIDGE, BOTANIC GARDEN
4489	CAMBRIDGE, CORPORATION STORE YARD
4492	CAMBRIDGE, NEWMARKET ROAD
455	CAMBRIDGE, NIAB
4491	CAMBRIDGE, QUEEN EDITH WAY
4488	CAMBRIDGE, SHERLOCK CLOSE
4484	CAMBRIDGE, TRUMPINGTON ROAD
4493	CAMBRIDGE, WOODLARK ROAD
4434	COTTENHAM
4435	COTTENHAM COLLEGE
4433	COTTENHAM NO 2
4478	FOXTON S WKS
4501	FULBOURN HOSP
4502	FULBOURN P STA
4482	GRANTCHESTER
4483	GRANTCHESTER MEADOWS
4481	GRANTCHESTER, LYNDEWODE
4462	GREAT SHELFORD VICARAGE
4479	HASLINGFIELD S WKS
4495	MILTON
4432	OAKINGTON
16734	OAKINGTON MET OFFICE
4431	OAKINGTON NO 2
4503	QUY HALL
4453	SAWSTON
4461	SHELFORD, STAPLEFORD HOUSE
4480	TOFT
16737	WATERBEACH MET OFFICE