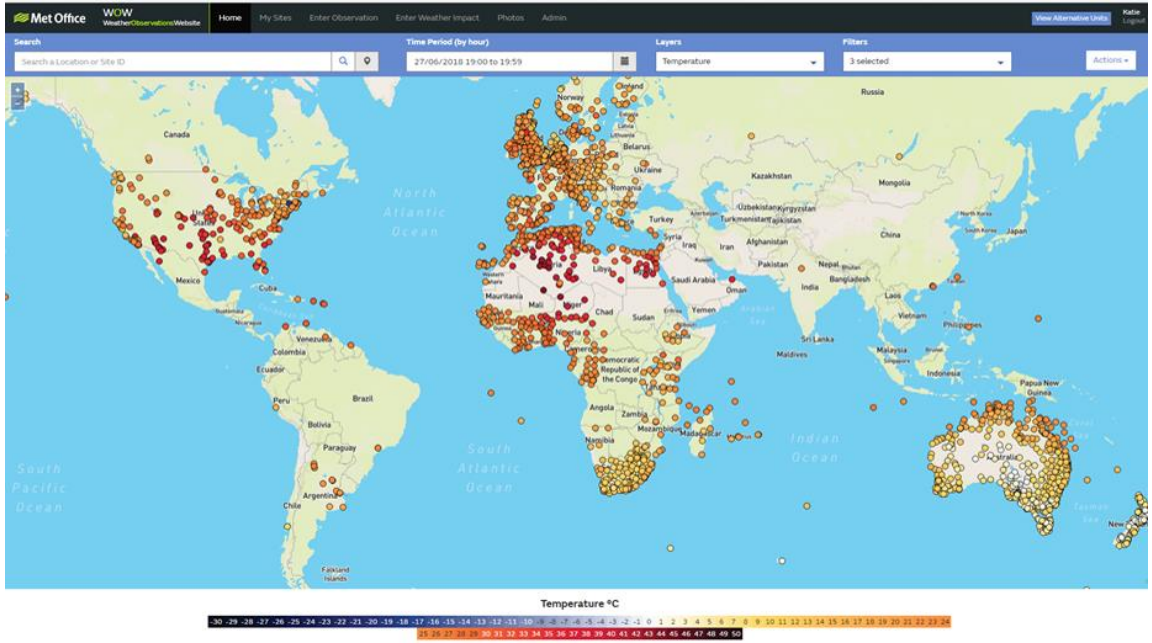
Using crowd-sourced observations to verify post-processed forecasts

Christopher Steele, Philip Gill, Matthew Spurrier

Met Office, UK

9th International Verification Methods Workshop, 20th May 2024

Why use crowd-sourced data?



- Potentially 1000's observations
- Data have been collected over a long period of time
  - WOW: June 2011
  - WUnderground : 1993
  - Weathercloud: 2012
- Data coverage where people live!
- Public engagement

Met Office

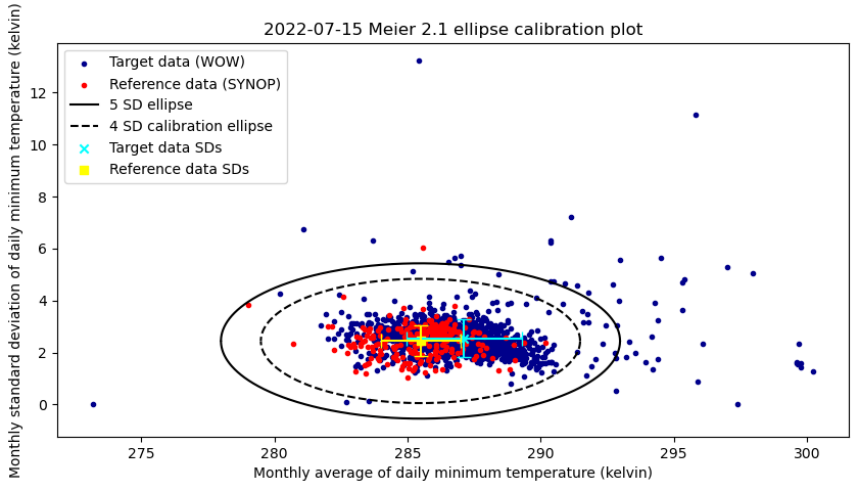
But is it good quality?

But is it good quality?

- It is true that the instrumentation may not be the same quality as those used in the official meteorological network
  - It varies throughout the network!
- The stations may also not be sited in the best locations
  - If you go to the trouble of purchasing a weather station and uploading the data to a crowd-source data centre, then it’s likely not been sited without some thought.
  - Again, this is the weather where people live and are actively experiencing

WOW industrial placement

- Assessed WOW temperature observations for possible use in operational verification.
- Applied simple Quality Control (QC) measures and assessed the impact at different stages
  - Majority of sites passed QC
  - Some outliers remained
  - Short time period studied



An opportunity…

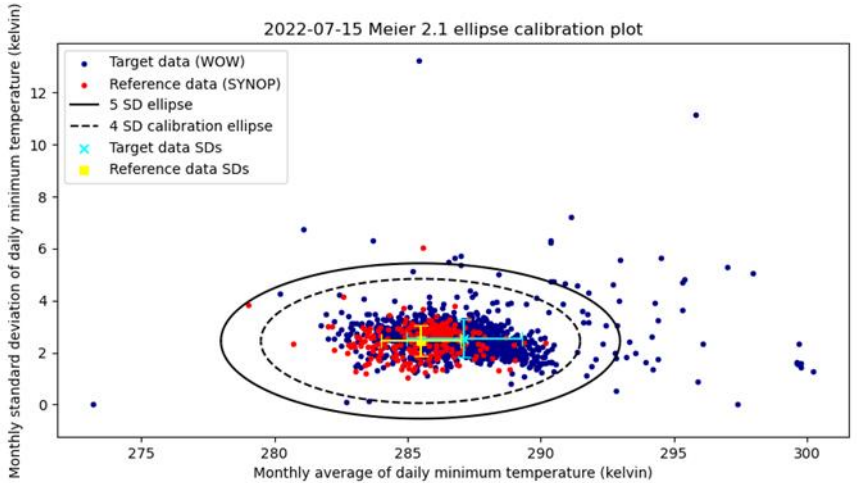
- The Met Office is in the process of replacing our old post-processing system with IMPROVER
- Usually, we would evaluate using a set of PWS sites, but the old system is tuned to these locations.
- We need to find an alternative source of observations…so why not WOW data?

Methodology

1. Extend the QC and run over a 12-month period to produce a static site list
2. Match to the nearest spot forecast location
   - Mimics what you would see on web/app
3. Evaluate forecasts
   - Between January – March 2024
   - Proportion of errors <=2K (POE2), RMSE, MAE, ME
   - T+1 – T+120
   - 1.5m temperature

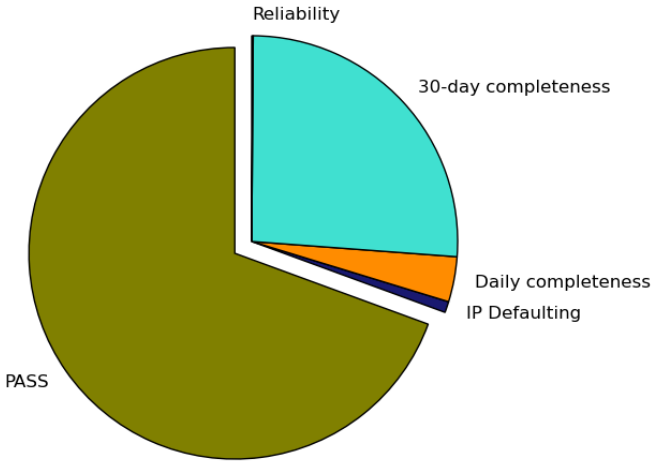
Quality Control

1. **IP defaulting** – Identical locations
2. **Data completeness**
   a) 80% reporting in 24hrs
   b) 80% reporting over rolling 30 days
3. **Data reliability**
   - Fails if the station 30-day mean minimum temperature is larger than 5σ of the average WOW temperature or has σ > 5 times a reference data set.



Meier, F., Fenner, D., Grassmann, T., Otto, M. & Scherer, D. (2017) Crowdsourcing air temperature from citizen weather stations for urban climate research. Urban Climate, 19, 170–191.

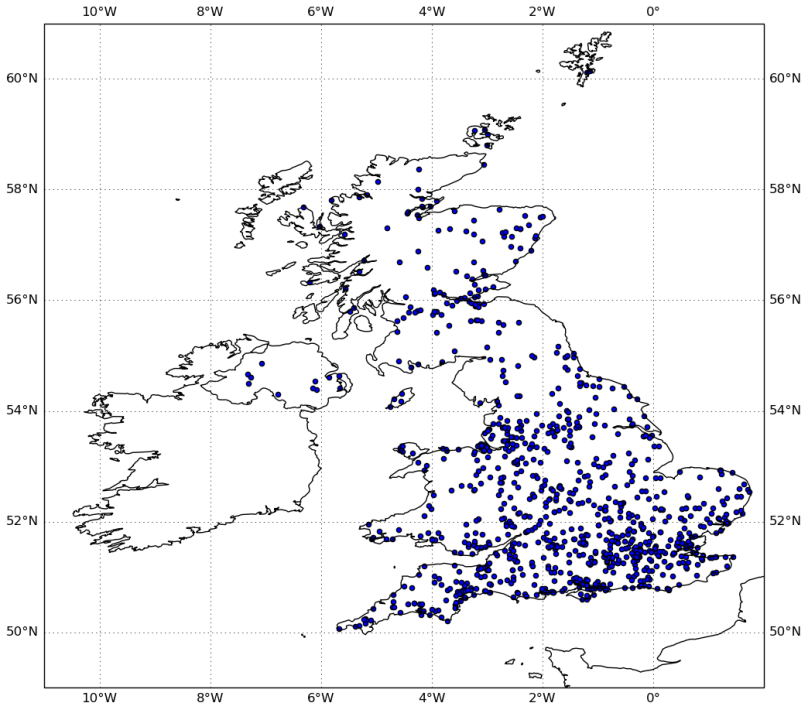
Quality control



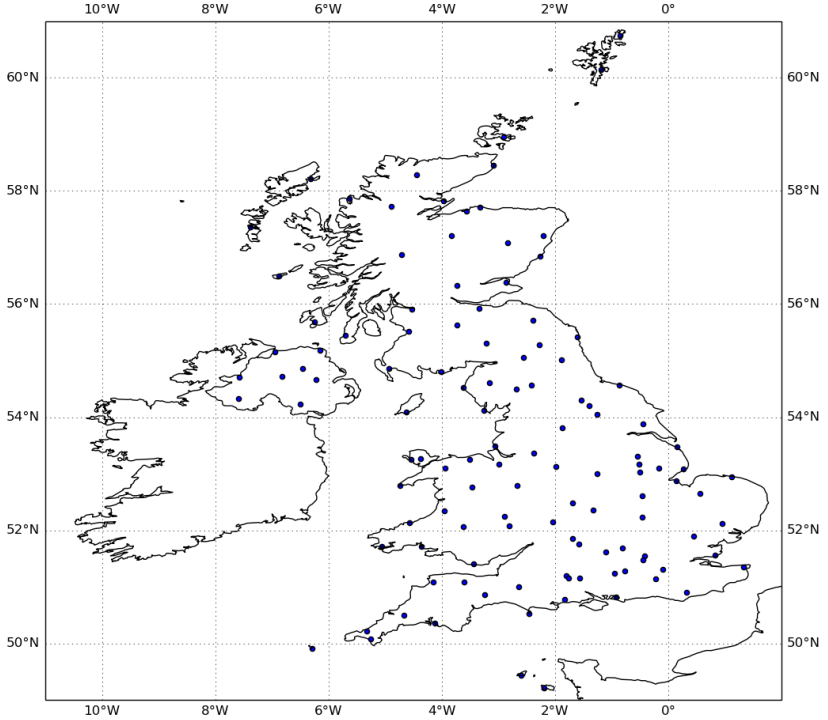
- 69% of observations passed QC
- Most common cause for rejection is data completeness.
- A site list of 975 stations that consistently passed QC was created.

Met Office

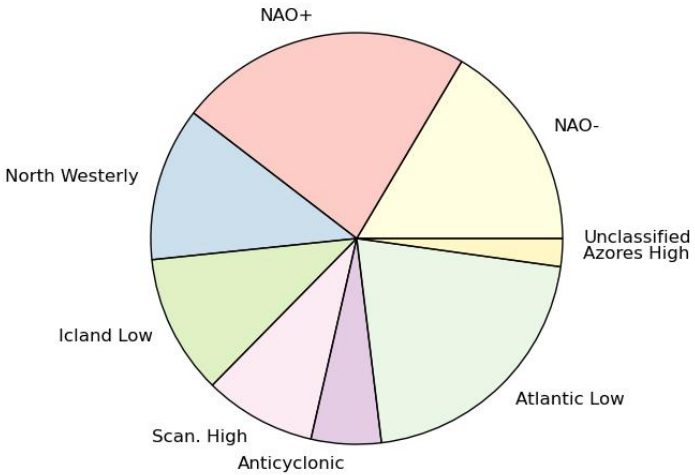
WOW



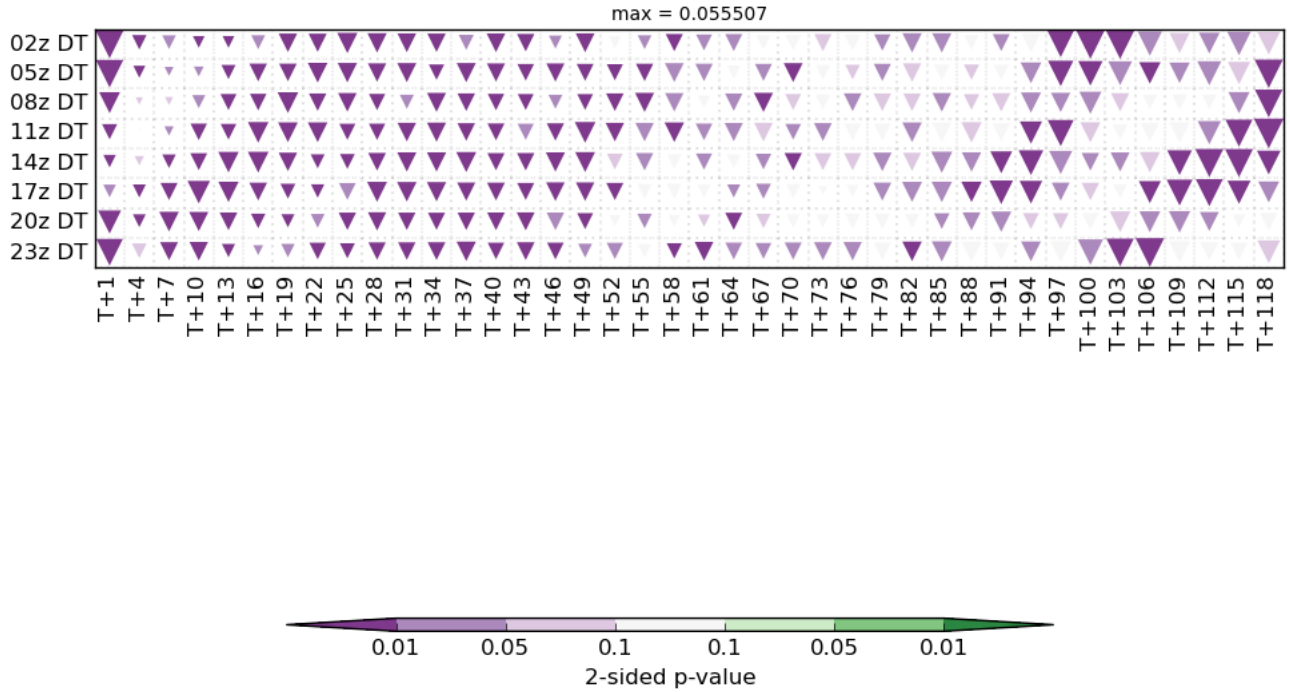
PWS



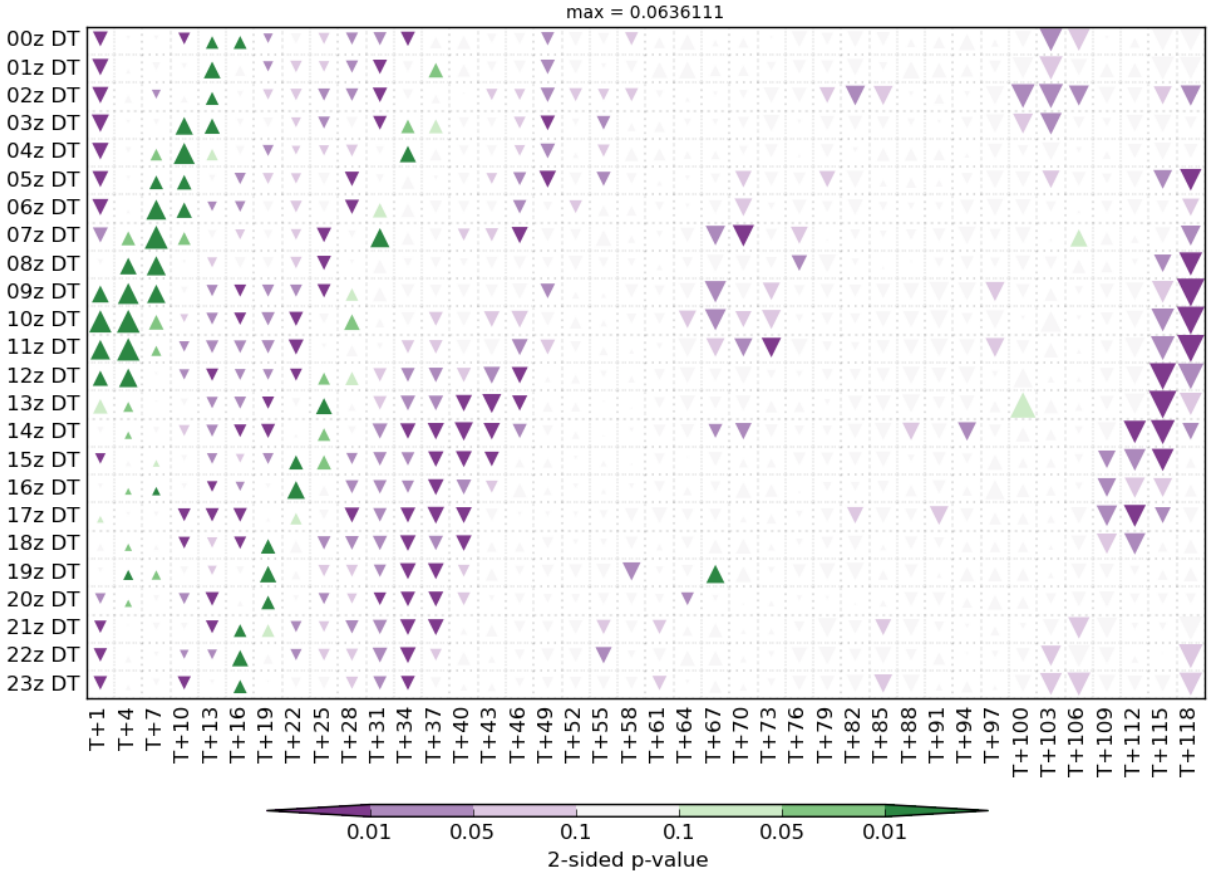
Weather regimes



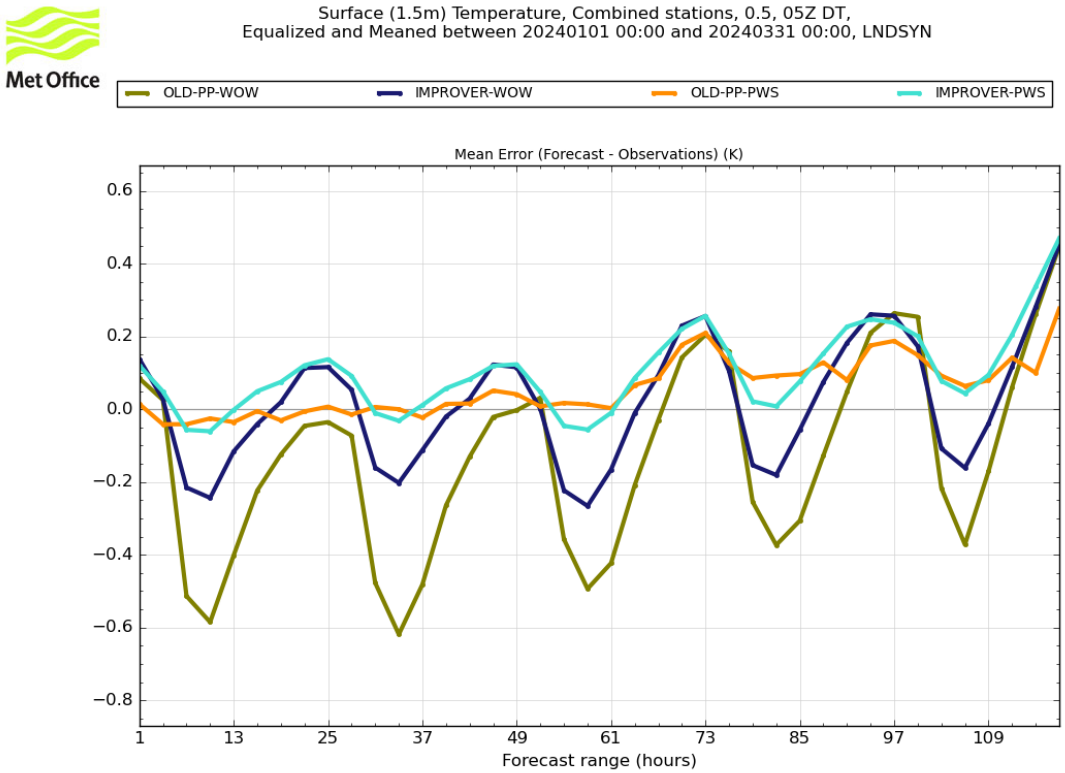
- Early January dominated by NAO^{-} , transitioning to NAO+, NW's and Icelandic Low
- February dominated by NAO+ and Icelandic Low (72%)
- March more changeable. Early dominance of Scandinavian High moving to wetter regimes later in the month


max = 0.055507
02z DT
05z DT
08z DT
11z DT
14z DT
17z DT
20z DT
23z DT
T+1
T+4
T+7
T+10
T+13
T+16
T+19
T+22
T+25
T+28
T+31
T+34
T+37
T+40
T+43
T+46
T+49
T+52
T+55
T+58
T+61
T+64
T+67
T+70
T+73
T+76
T+79
T+82
T+85
T+88
T+91
T+94
T+97
T+100
T+103
T+106
T+109
T+112
T+115
T+118
0.01
0.05
0.1
0.1
0.05
0.01
2-sided p-value

Met Office


max = 0.0636111
00z DT
01z DT
02z DT
03z DT
04z DT
05z DT
06z DT
07z DT
08z DT
09z DT
10z DT
11z DT
12z DT
13z DT
14z DT
15z DT
16z DT
17z DT
18z DT
19z DT
20z DT
21z DT
22z DT
23z DT
T+1
T+4
T+7
T+10
T+13
T+16
T+19
T+22
T+25
T+28
T+31
T+34
T+37
T+40
T+43
T+46
T+49
T+52
T+55
T+58
T+61
T+64
T+67
T+70
T+73
T+76
T+79
T+82
T+85
T+88
T+91
T+94
T+97
T+100
T+103
T+106
T+109
T+112
T+115
T+118
0.01
0.05
0.1
0.1
0.05
0.01
2-sided p-value

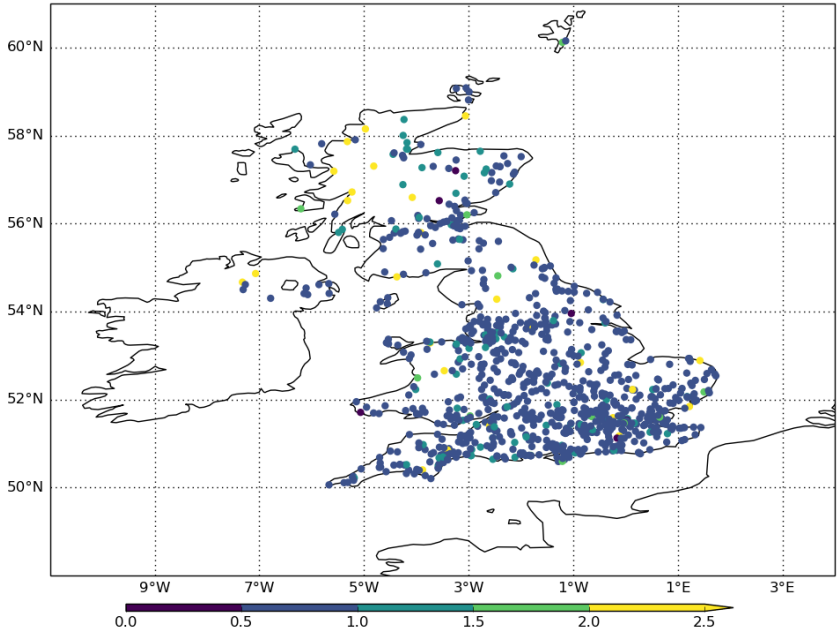
Met Office



Met Office

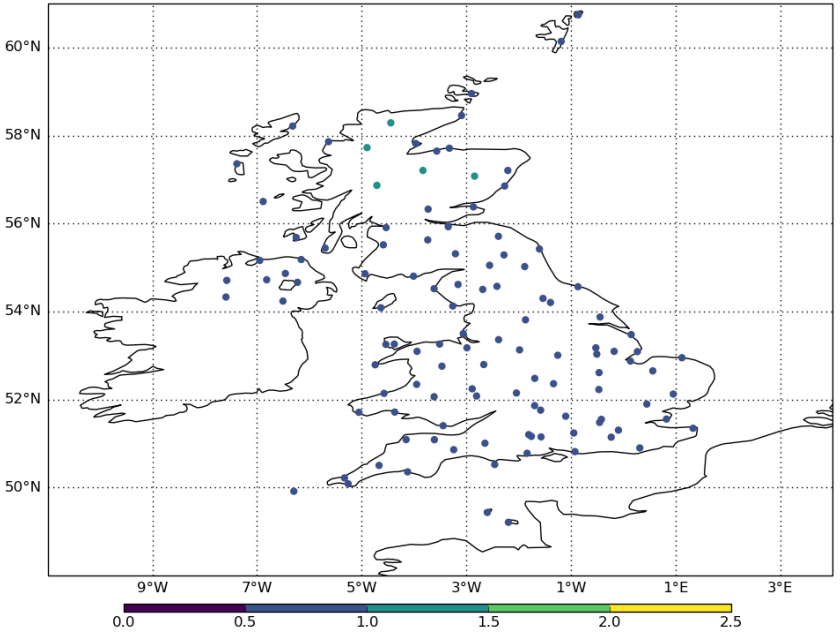
WOW

Surface (1.5m) Temperature (K), Mean Absolute Error (Forecast - Observations), 05Z DT,
20240101 00:00 to 20240331 23:00, LNDSYN, IMPROVER



PWS

Surface (1.5m) Temperature (K), Mean Absolute Error (Forecast - Observations), 0.5, 05Z DT,
20240101 00:00 to 20240331 23:00, LNDSYN, IMPROVER-PWS



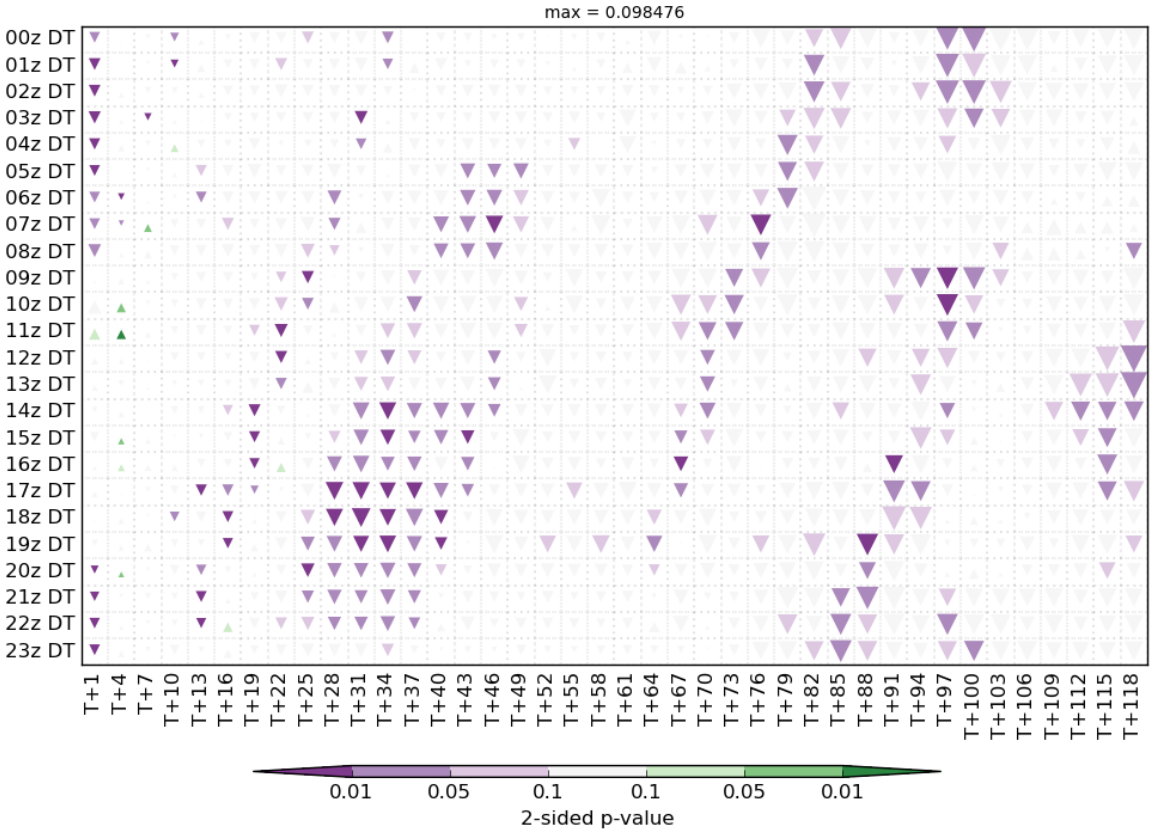
Met Office

Initial comments…

- Using only PWS sites, the old system is consistently better than IMPROVER at all lead times
- Using WOW, the difference depends on time of day with daytime temperature better predicted by IMPROVER
- IMPROVER ME more neutral than the old system
- Most WOW sites have a comparable forecast error when compared to PWS
  - Though there are a few outliers, mostly in more remote locations

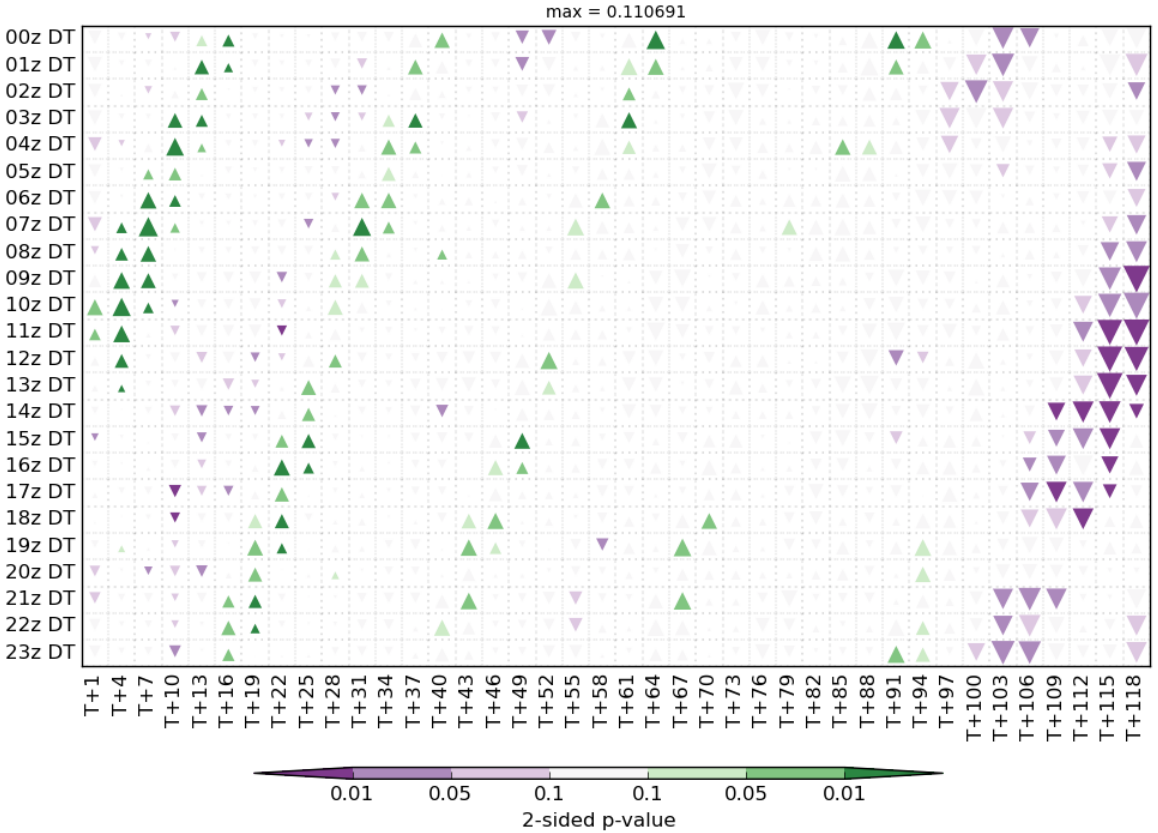
Met Office

Surface (1.5m) Temperature (K), Proportion of Errors <=2K (Forecast - Observations),
Combined stations, Equalized, 20240101 00:00 to 20240131 00:00, LNDSYN,
Difference (IMPROVER - Old System)



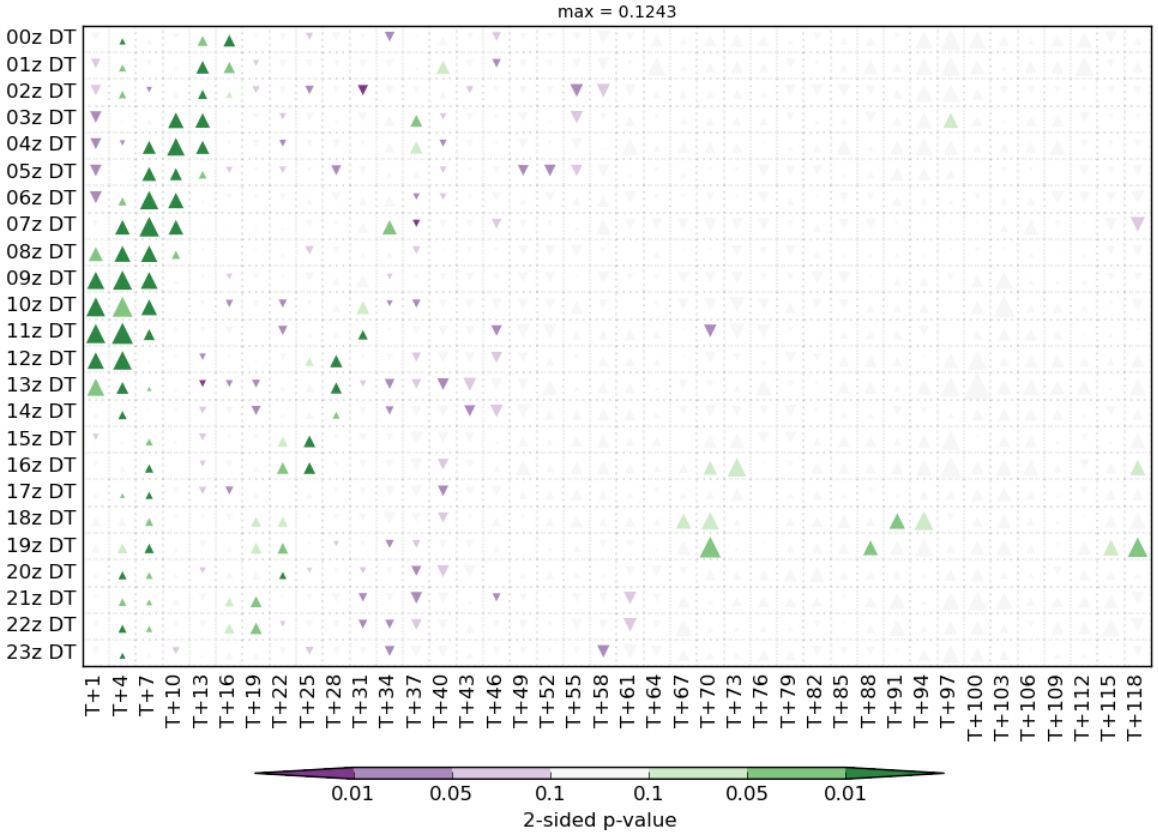
Met Office

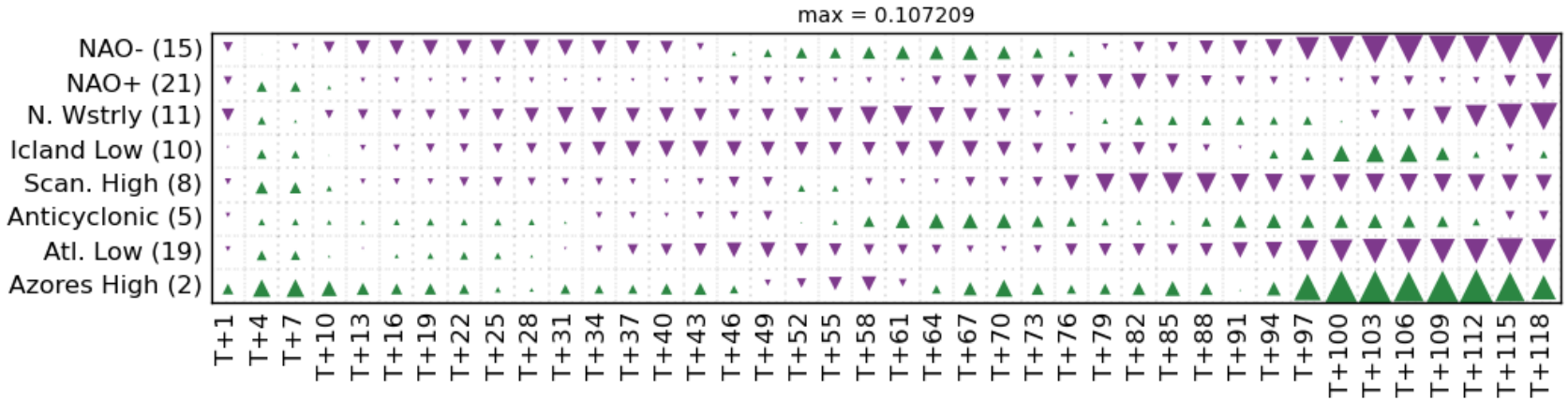
Surface (1.5m) Temperature (K), Proportion of Errors <=2K (Forecast - Observations),
Combined stations, Equalized, 20240201 00:00 to 20240229 00:00, LNDSYN,
Difference (IMPROVER - Old System)



Met Office

Surface (1.5m) Temperature (K), Proportion of Errors <=2K (Forecast - Observations),
Combined stations, Equalized, 20240301 00:00 to 20240331 00:00, LNDSYN,
Difference (IMPROVER - Old System)



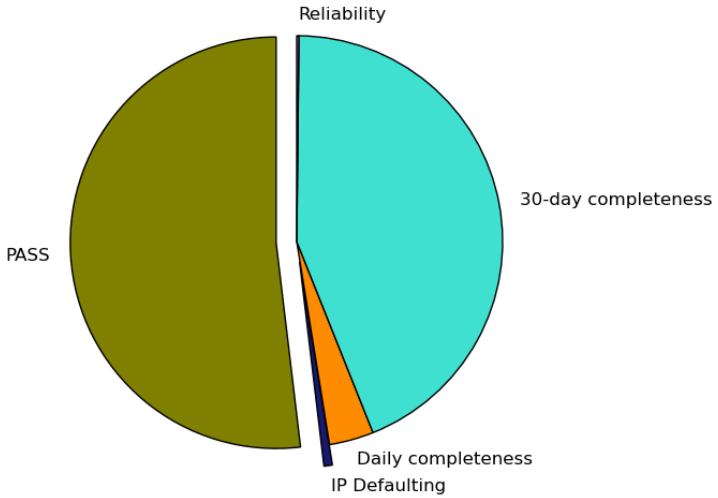

max = 0.107209
NAO- (15)
NAO+ (21)
N. Wstrly (11)
Icland Low (10)
Scan. High (8)
Anticyclonic (5)
Atl. Low (19)
Azores High (2)
T+1
T+4
T+7
T+10
T+13
T+16
T+19
T+22
T+25
T+28
T+31
T+34
T+37
T+40
T+43
T+46
T+49
T+52
T+55
T+58
T+61
T+64
T+67
T+70
T+73
T+76
T+79
T+82
T+85
T+88
T+91
T+94
T+97
T+100
T+103
T+106
T+109
T+112
T+115
T+118

Monthly performance…

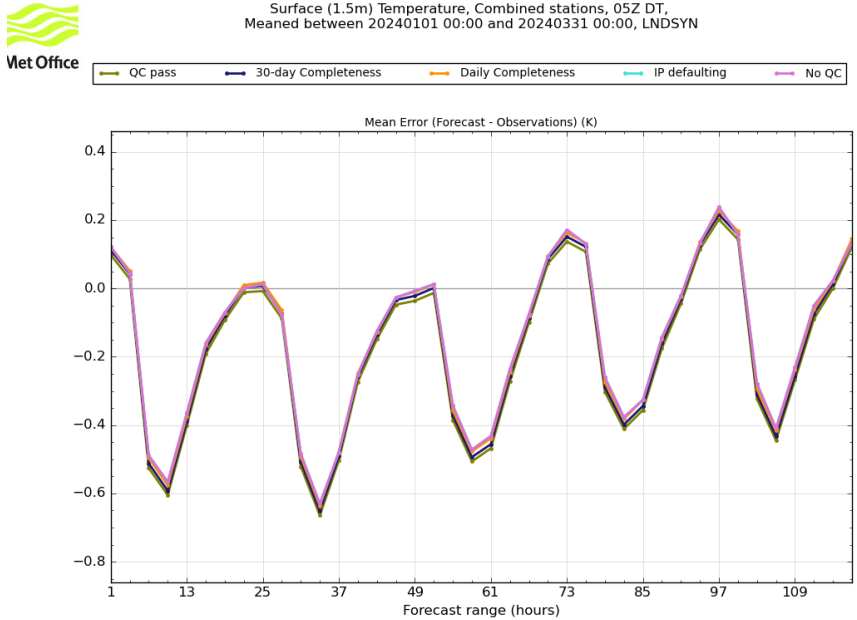
- January worse for IMPROVER than the other months
  - Weak/no signal of better daytime IMPROVER temperatures
- February/March were known for being particularly wet
  - Some regime differences but too few data to be certain about any regime-based differences in performance

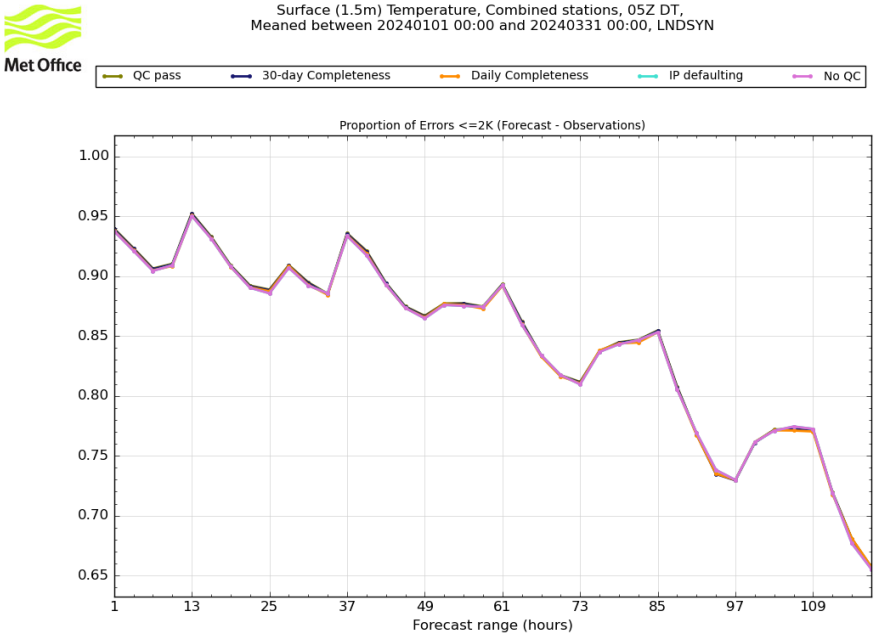
Quality control of study period

- Significant rejection based on 30-day completeness
- Just over half of data passed (51.9%)

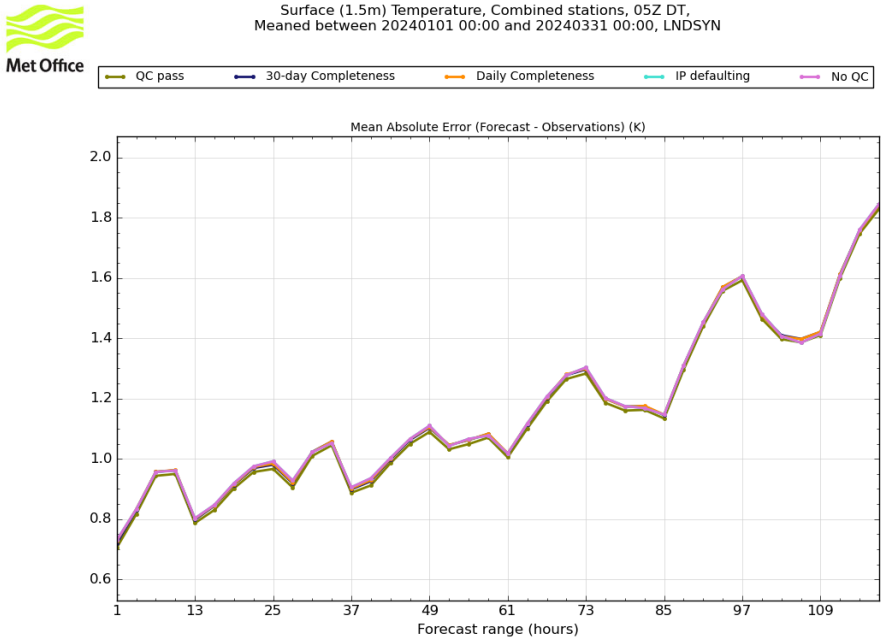


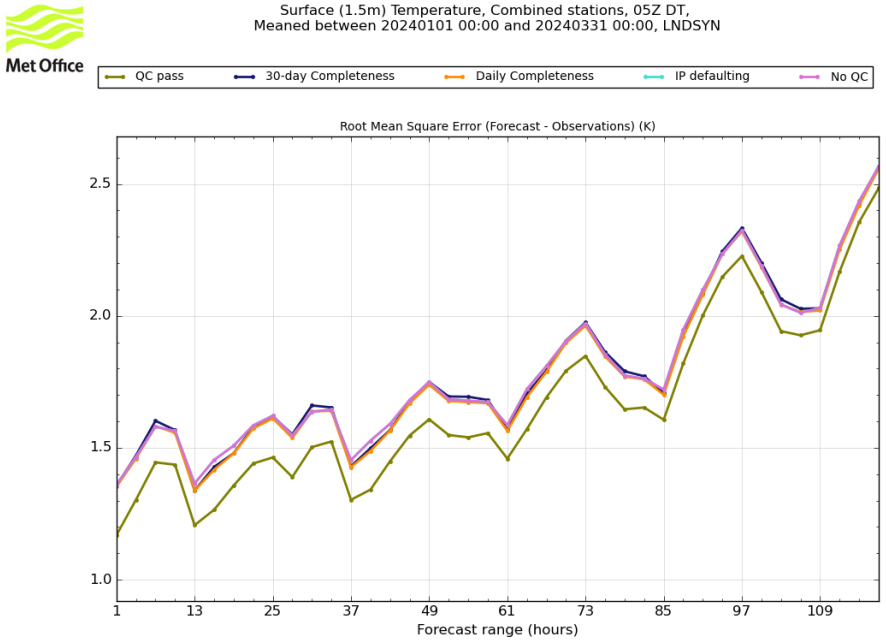
Met Office





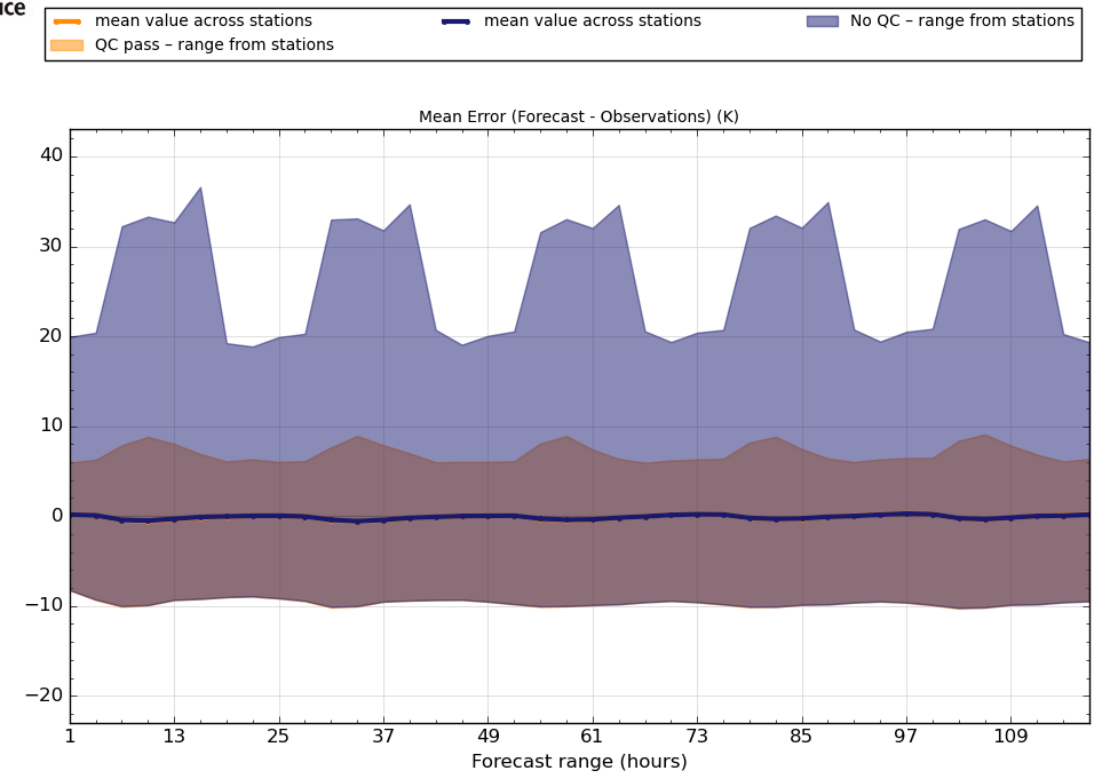
Met Office





Met Office

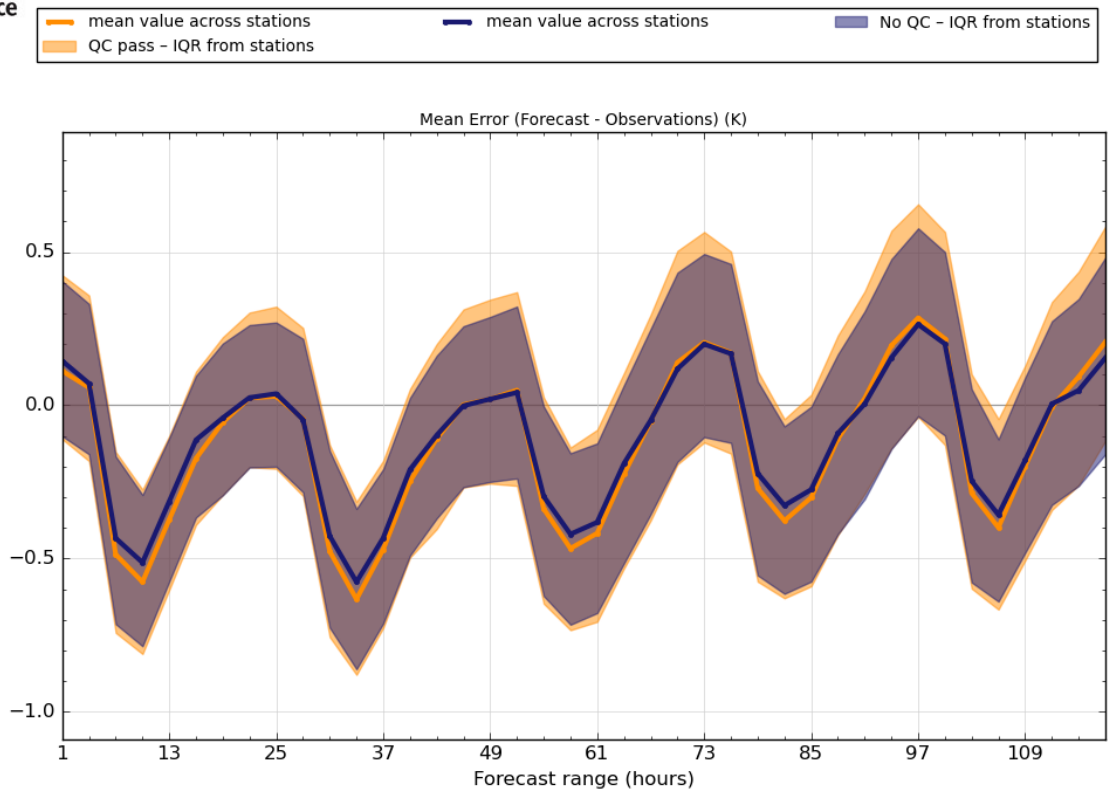
Surface (1.5m) Temperature, 05Z DT, Meaned between 20240101 00:00 and 20240331 00:00, LNDSYN



Met Office

Met Office

Surface (1.5m) Temperature, 05Z DT, Meaned between 20240101 00:00 and 20240331 00:00, LNDSYN



Summary of QC

- Quality control has a little influence on the overall score
  - There are a large number of good sites!
  - Unless using a metric that depends on the squared error (RMSE)
- Most influential stage is the data reliability check
- Still the occasional large error, most often associated with more remote sites.

Overall…

- Crowd-sourced data are a useful resource for operational verification!
- Most WOW sites after QC had a comparable forecast error when compared to PWS sites
- Very little change in score was observed with and without QC (except RMSE)
  - Large number of good sites compared to a few outliers

Met Office

What next?

- Generate larger volumes of data
- Add additional parameters?
- Explore different Quality Control Schemes?


Met Office
Thank you and any questions?
www.metoffice.gov.uk
© Crown Copyright 2024, Met Office