

Homogenisation of daily station data in England and Wales

Kay Shelton, Sarah Warren, Richard Davis,
Duncan Faulkner



Background

- Environment Agency want to develop a gridded historical daily potential evapotranspiration (PET) dataset for England and Wales based on observed data.
- Observed PET dataset will provide alternative to existing PET datasets known to include inhomogeneities.
- PET is an important quantity for hydrological prediction. Needed to close water balance and allow accurate estimates of surface and sub-surface runoff.
- PET calculation using Penman-Monteith equation requires knowledge of temperature, humidity, radiation and windspeed.



Approach

- First step to the gridded PET dataset is to homogenise the daily station observations.
- No prior experience homogenising climate data or using homogenisation software.
- Process needs to be:
 - repeatable (to allow consistent future updates),
 - objective (to minimise errors introduced by inexperienced users),
 - automatic (to minimise time spent on the process and keep costs down),
 - appropriate to all climate variables (to allow application of a single consistent process),
 - applicable to daily data.
- MASH (Multiple Analysis of Series for Homogenization) meets all criteria, and provides additional QC at the end of the process.



Data

- Met Office station data for England and Wales for 1961-2012:
 - 0900 2m air temperature (AIRTEMP)
 - 0900 2m dewpoint temperature (DEWPOINT)
 - 24h maximum 2m air temperature (AIRMAX)
 - 24h minimum 2m air temperature (AIRMIN)
 - 24h sunshine duration (SUNSHINE)
 - 24h mean windspeed (WIND)
- Data automatically archived in the Met Office Integrated Data Archive System (MIDAS) since 1997; includes automatic and manual quality control (QC).
- Data prior to 1997 added to MIDAS archive without MIDAS QC. Historical QC processes may have been applied to data.

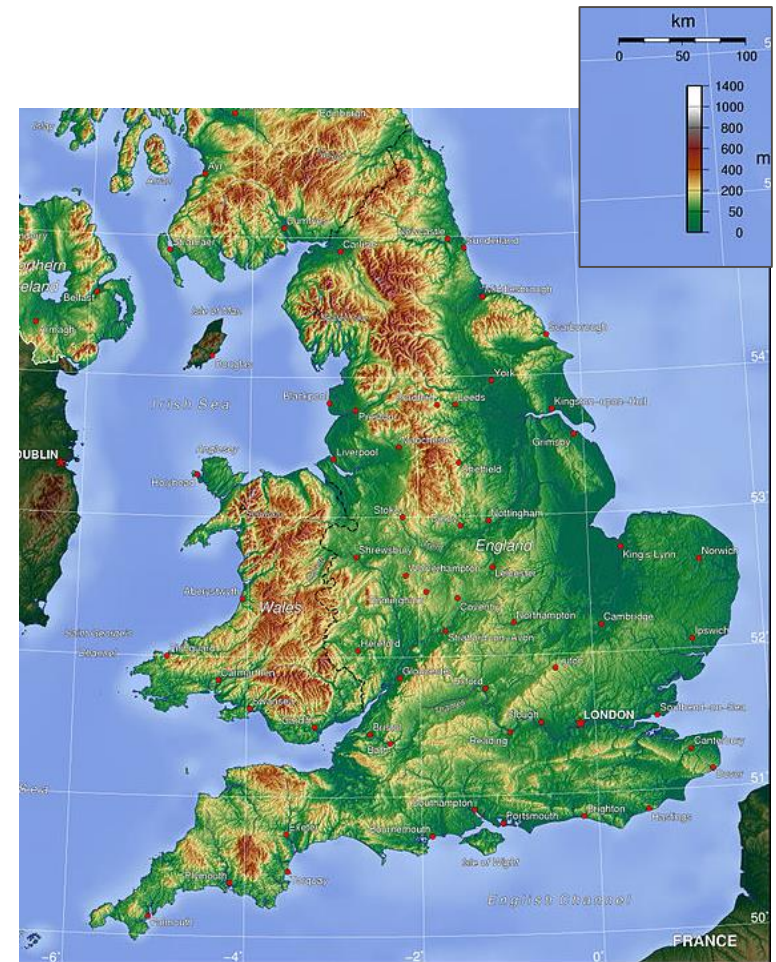


Image from Wikipedia.



Data Caveats (i)

- AIRTEMP and DEWPOINT:
 - Limited data available in MIDAS prior to 1972 (~50 stations).
 - Homogenisation for these limited to 1972-2012.
- DEWPOINT:
 - ~65% DEWPOINT sites missing during 1994-1999.
- SUNSHINE:
 - Observations are a mix of Campbell-Stokes (CS; manual) and Kipp and Zonen (KZ; automatic), dependent on site.
 - CS observations are converted to KZ-equivalent by inverting the regression equation and using monthly coefficients derived for daily data by Legg (2014).
 - This approach is not ideal, but refining the CS to KZ-equivalent conversion was beyond the scope of the project.



Data Caveats (ii)

- WIND:
 - Variety of mast heights across network; correction to effective mast height of 10m required.
 - Incomplete metadata records for a few stations; some mast height records had to be retrieved from paper archives National Meteorological Archive at the Met Office in Exeter.
 - Where no effective or mast height information is available, observations are assumed to be valid at 10m.
- Data QC:
 - Majority of data have not passed through MIDAS QC, so to maximise the amount of data to be homogenised, the data are used “as is”.



HOMOGENISATION RESULTS (1ST ROUND)

AIRMAX



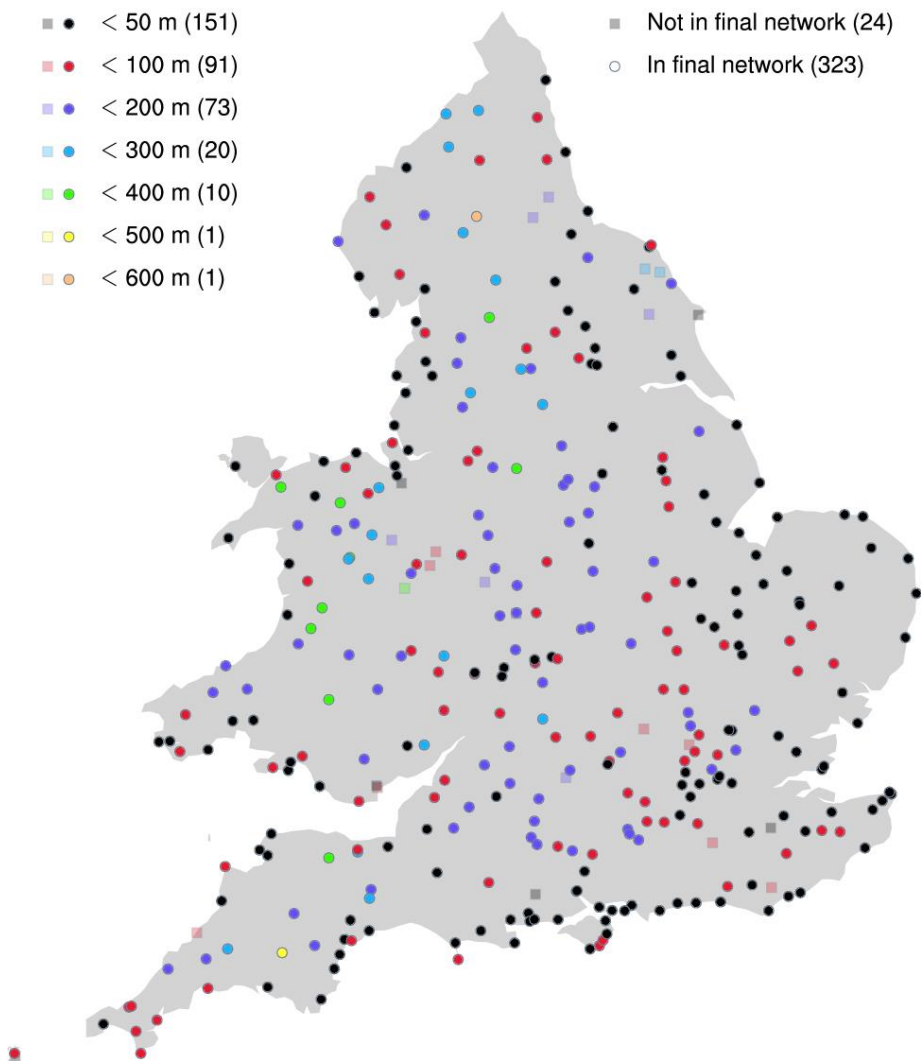
AIRMAX: Network

Station elevation

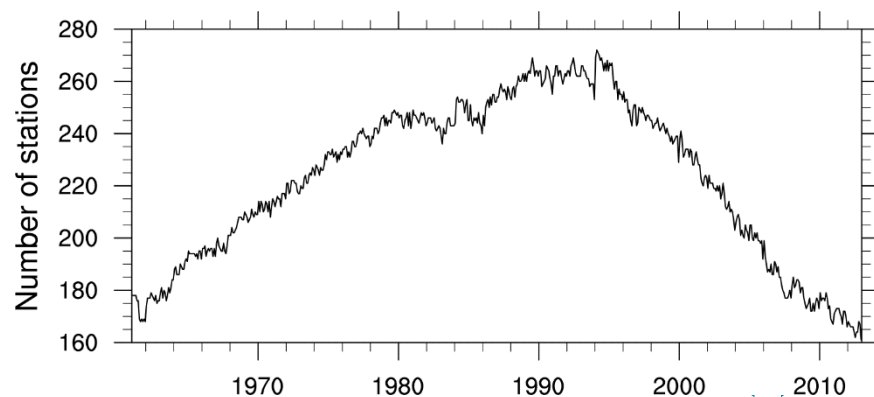
- < 50 m (151)
- < 100 m (91)
- < 200 m (73)
- < 300 m (20)
- < 400 m (10)
- < 500 m (1)
- < 600 m (1)

Network membership

- Not in final network (24)
- In final network (323)



- Imposed minimum record length:
 - Years with daily data 90% complete
 - 20 years
- 24 stations removed after 9 iterations of MASH process
- Final network:
 - 323 stations
 - Variety of altitudes
 - Good spread across domain
 - No coverage in northern tip of England

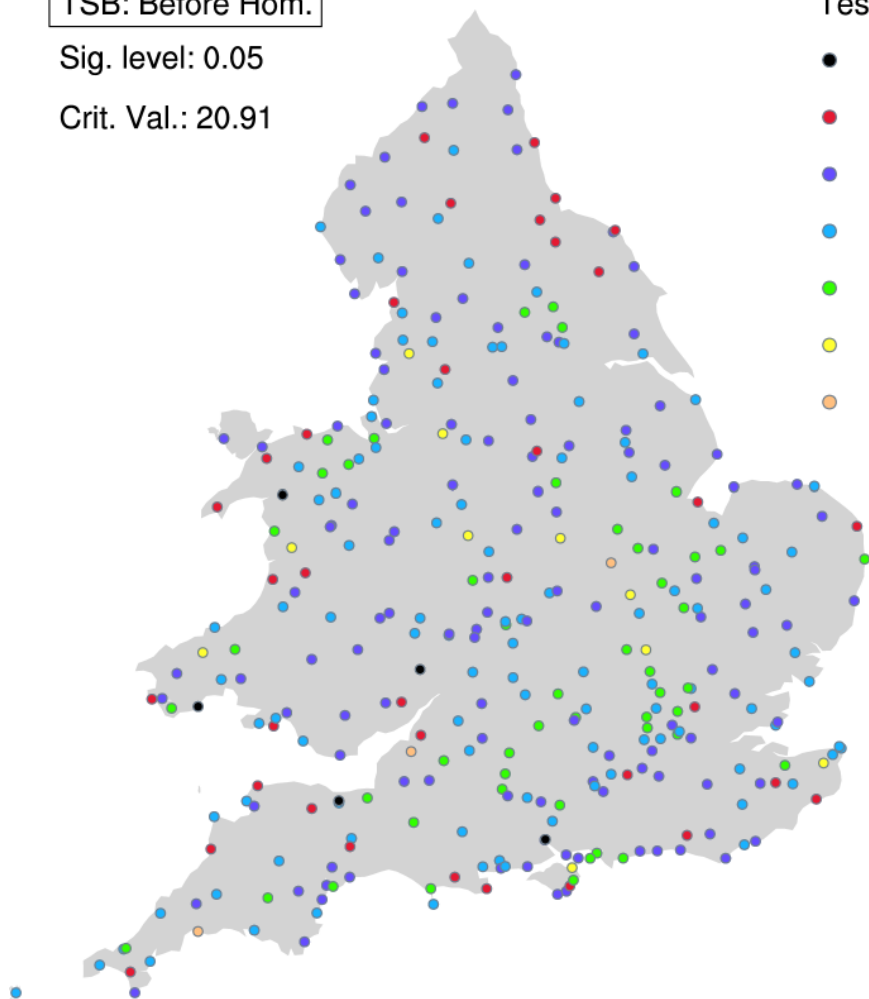


AIRMAX: Homogenisation performance

TSB: Before Hom.

Sig. level: 0.05

Crit. Val.: 20.91



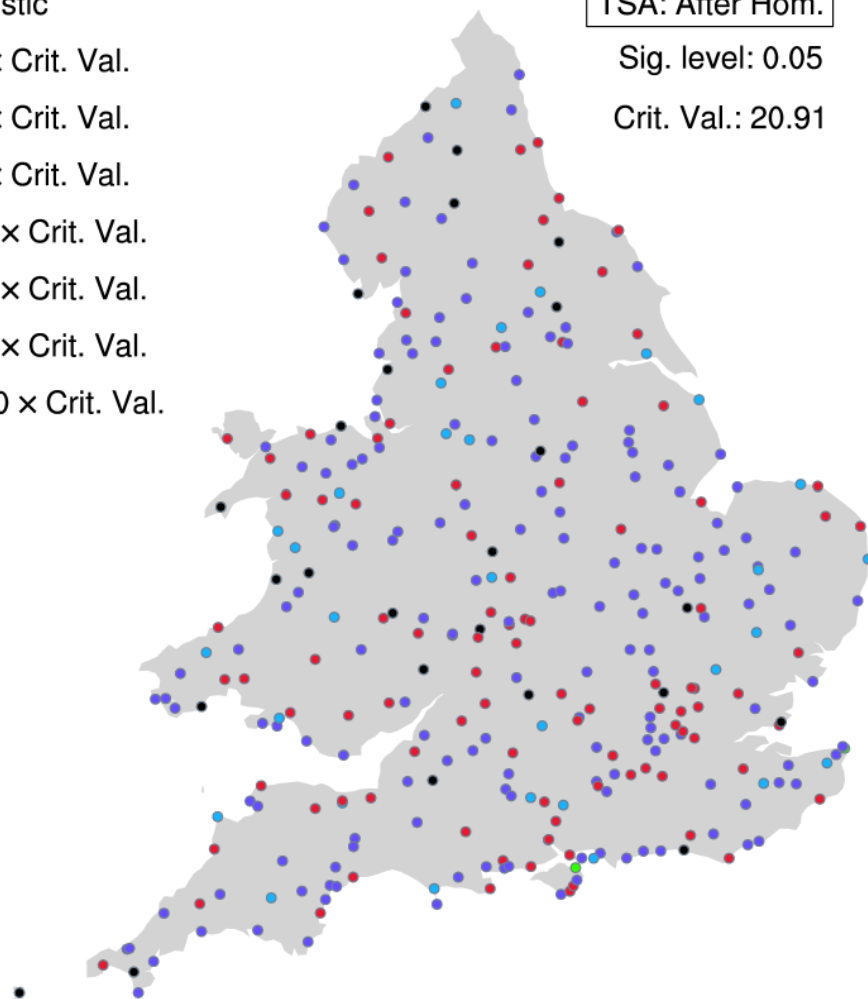
Test statistic

- $\leq 1 \times$ Crit. Val.
- $\leq 2 \times$ Crit. Val.
- $\leq 5 \times$ Crit. Val.
- $\leq 10 \times$ Crit. Val.
- $\leq 20 \times$ Crit. Val.
- $\leq 50 \times$ Crit. Val.
- $\leq 100 \times$ Crit. Val.

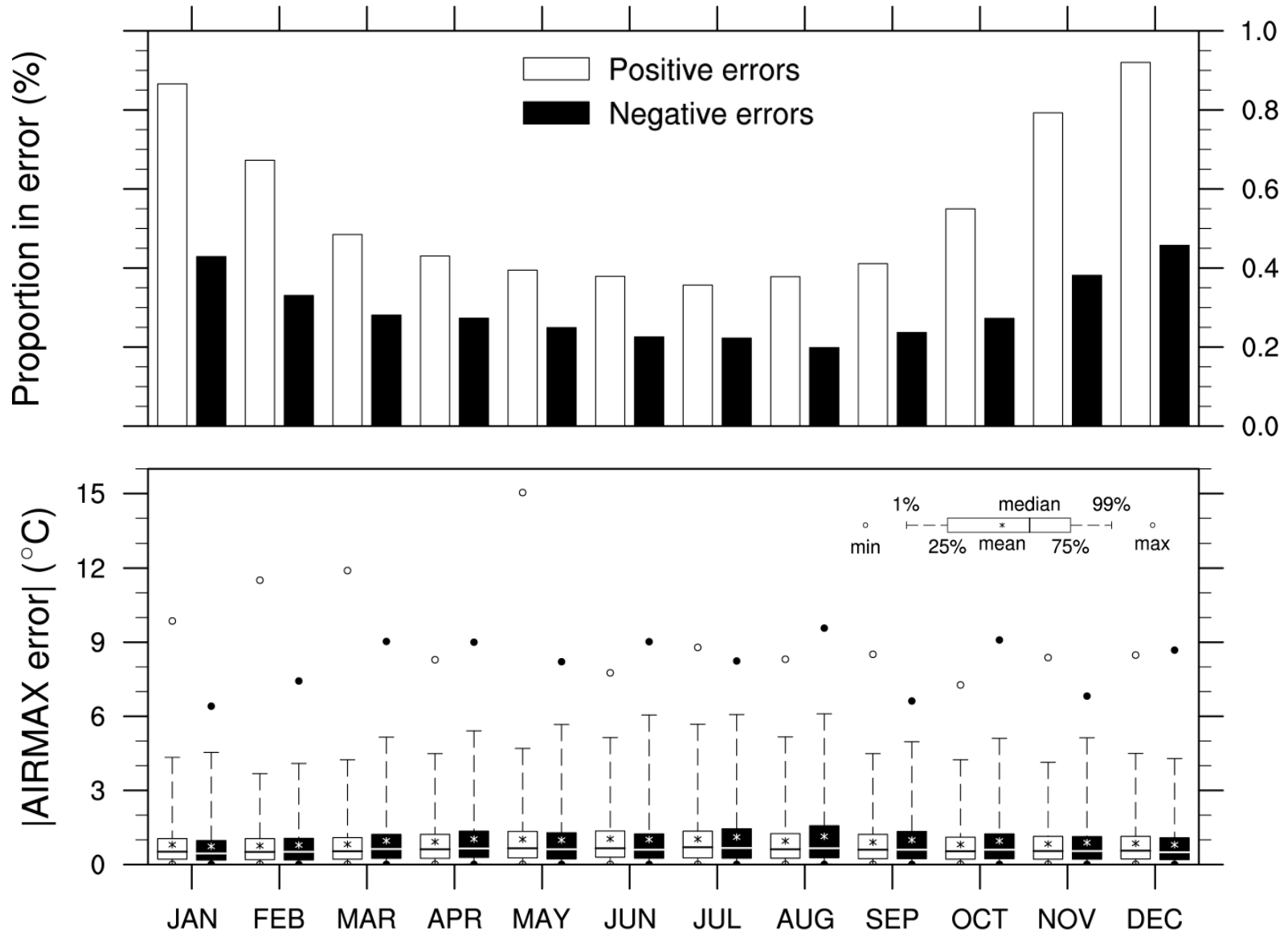
TSA: After Hom.

Sig. level: 0.05

Crit. Val.: 20.91

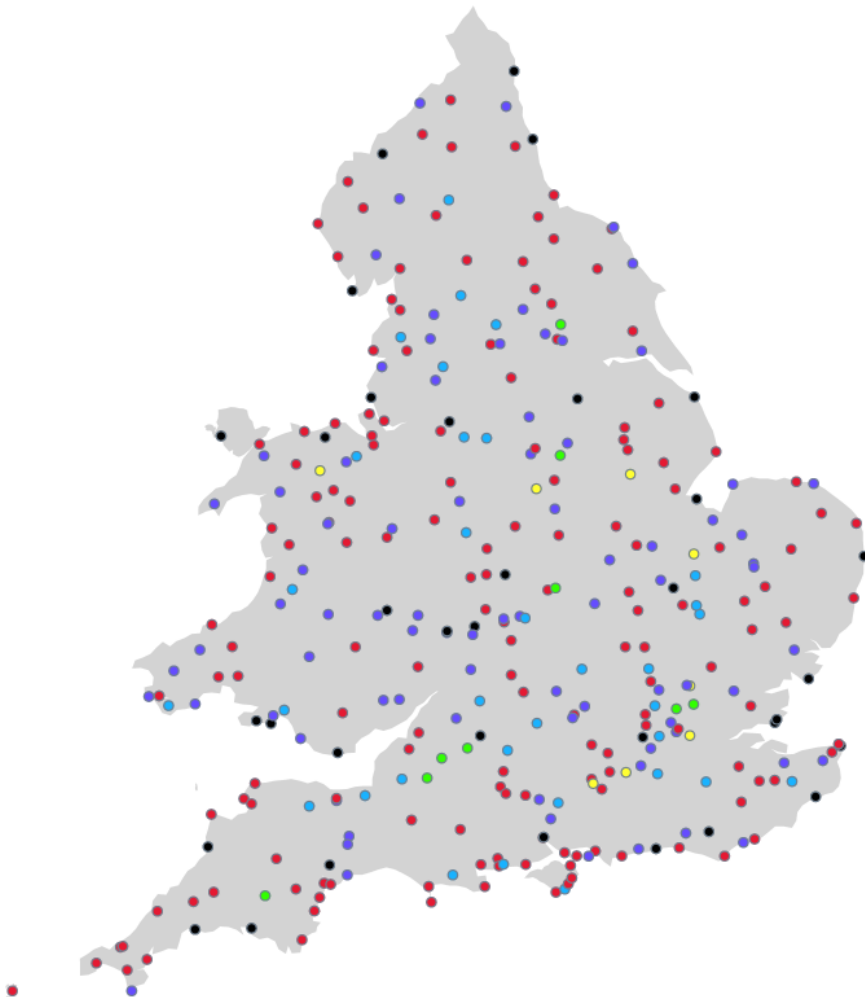


AIRMAX: Homogenised daily QC (seasonal)



AIRMAX: Homogenised daily QC (geographic)

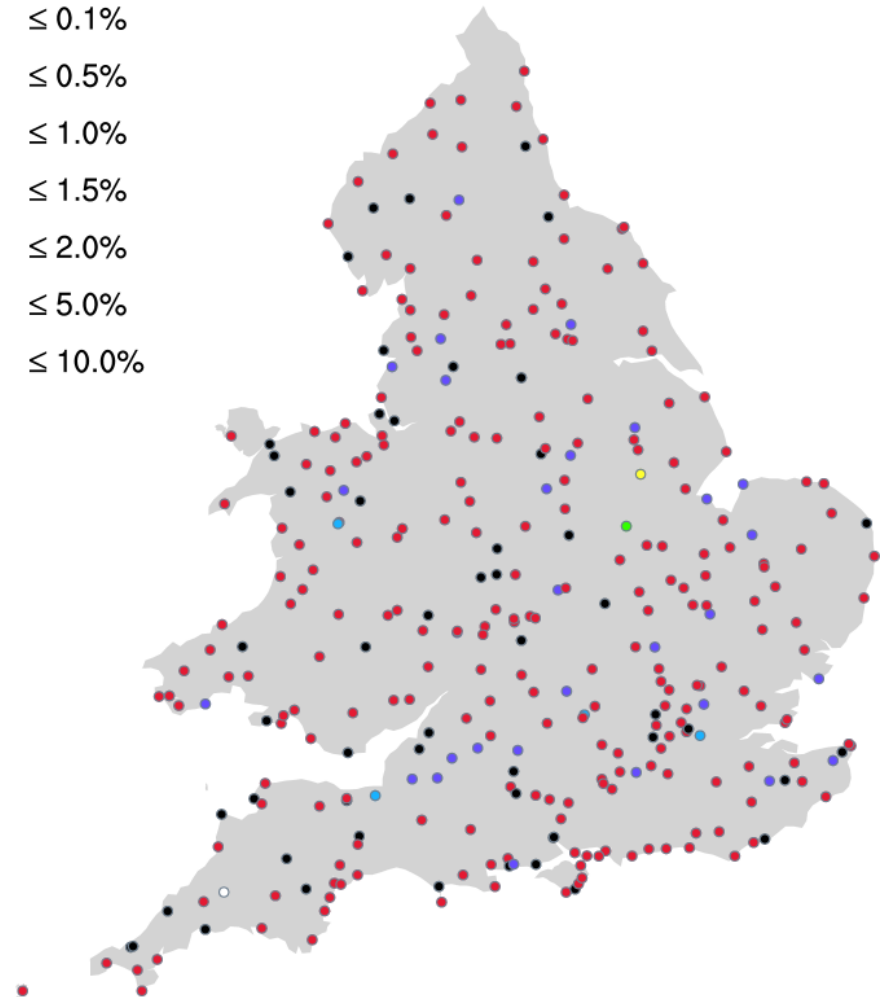
Positive errors



Proportion in error

- No errors
- ≤ 0.1%
- ≤ 0.5%
- ≤ 1.0%
- ≤ 1.5%
- ≤ 2.0%
- ≤ 5.0%
- ≤ 10.0%

Negative errors



HOMOGENISATION RESULTS (1ST ROUND)

Issues



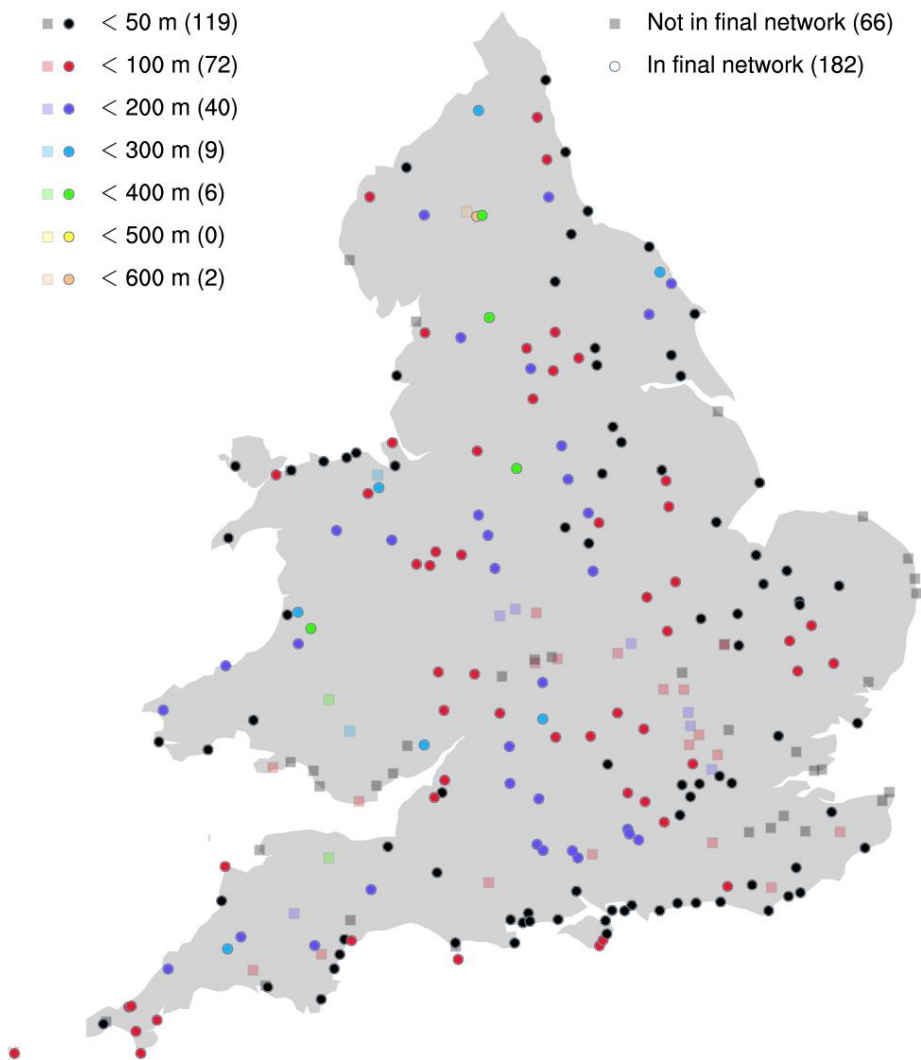
SUNSHINE: Network

Station elevation

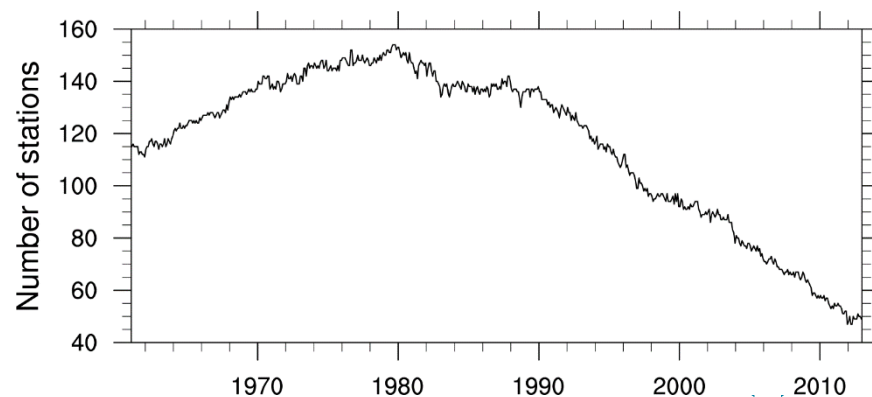
- < 50 m (119)
- < 100 m (72)
- < 200 m (40)
- < 300 m (9)
- < 400 m (6)
- < 500 m (0)
- < 600 m (2)

Network membership

- Not in final network (66)
- In final network (182)



- Imposed minimum record length:
 - Years with daily data 90% complete
 - 15 years
- 66 stations removed after 11 iterations of MASH process
- Final network:
 - 182 stations
 - Variety of altitudes
 - Large gaps in coverage: east coast, Wales, central belt



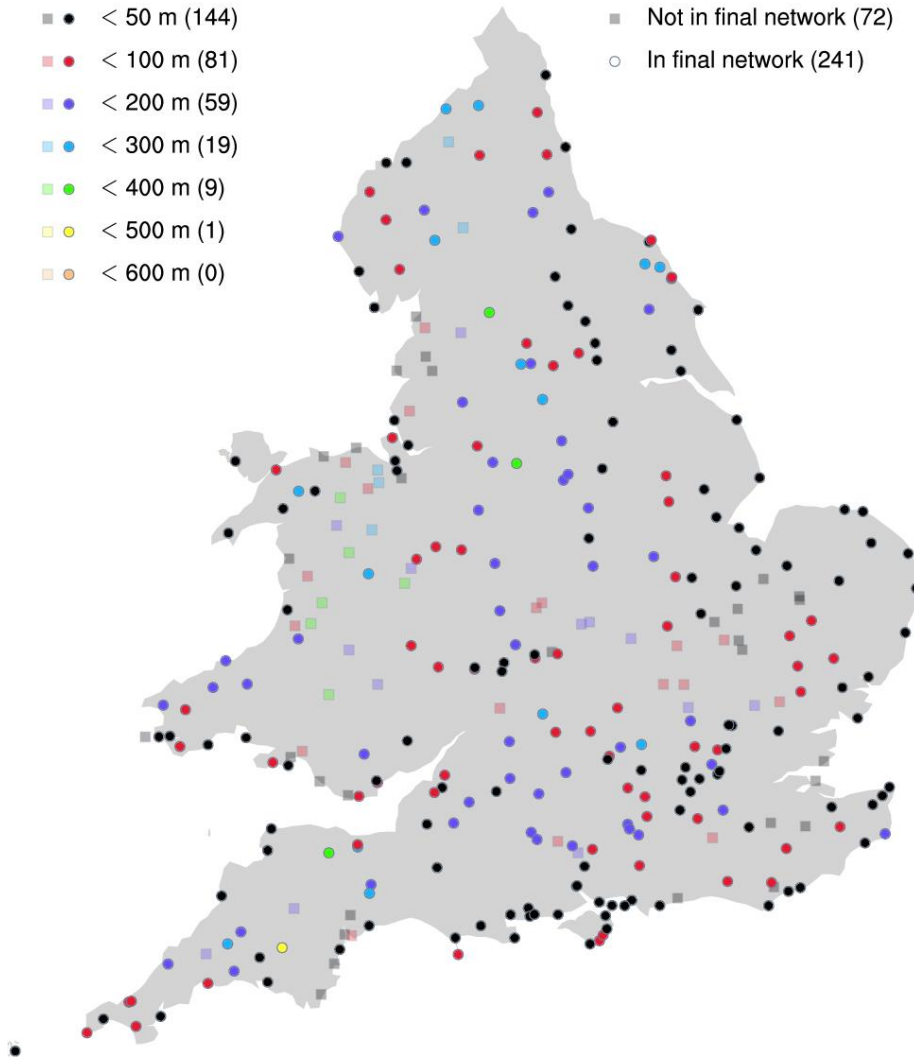
DEWPOINT: Network

Station elevation

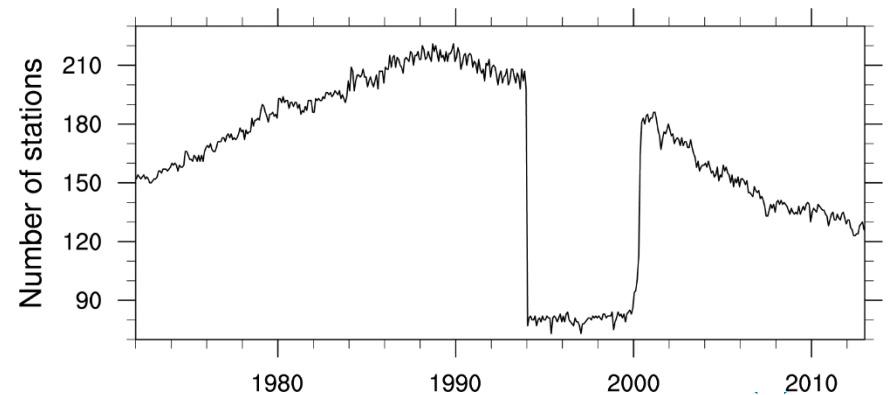
- < 50 m (144)
- < 100 m (81)
- < 200 m (59)
- < 300 m (19)
- < 400 m (9)
- < 500 m (1)
- < 600 m (0)

Network membership

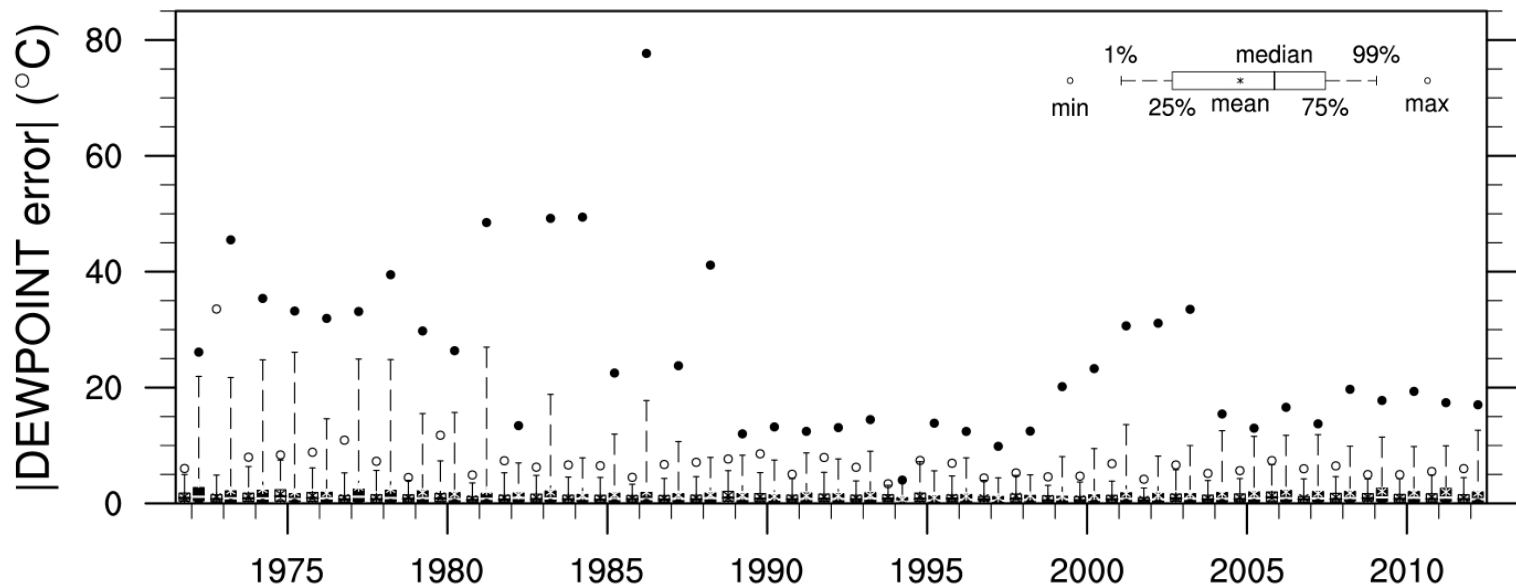
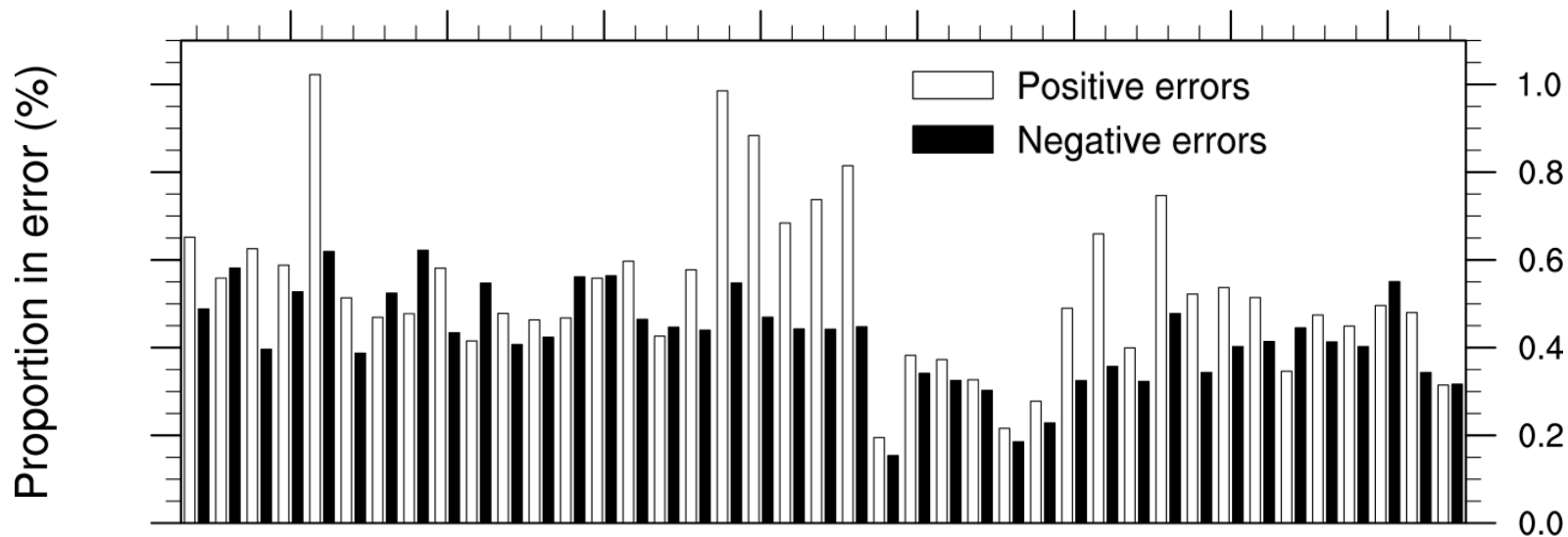
- Not in final network (72)
- In final network (241)



- Imposed minimum record length:
 - Years with daily data 90% complete
 - 15 years
- 72 stations removed after 14 iterations of MASH process
- Final network:
 - 241 stations
 - Removed several higher altitude stations
 - Large gaps in coverage: Wales, central belt



DEWPOINT: Homogenised daily QC (annual)



Summary

- 5 out of 6 parameters have undergone a first round of homogenisation, resulting in overall more homogeneous networks for each parameter.
- Some specific issues have been identified regarding the networks for some parameters:
 - SUNSHINE – large gaps in network, especially in Wales.
 - DEWPOINT – large drop in number of stations for 6 year period in 90s; large gaps in network.
- Erroneous observations flagged by MIDAS should be removed.
- Should consider removing stations with frequent errors.
- MASH QC seems to do a good job identifying erroneous observations.
- Seasonality in errors identified by MASH QC for some parameters is very interesting.



Next steps

- Extension to project to:
 - Include Scottish border sites for all parameters
 - Extend the record to end of 2015
 - Add radiation data, either homogenise separately, or (more likely) convert sunshine to radiation and homogenise together
 - Investigate interpolation of 0600 and 1200 T and Td to 0900 for pre-1972 period
 - Include additional wind meta data information
- Re-do homogenisation for all parameters with updates to records and networks.
- Remove / update stations based on identified errors.
- Employ further rounds of homogenisation to further improve homogeneity.



Questions for you

- How best to perform future updates?
 - Repeat entire process with updated raw series or append raw series updates to existing homogenised series and re-run?
 - Resulting PET dataset used to re-calibrate rainfall-runoff and groundwater models. Costly to keep repeating process if entire historical PET dataset changes when new data is added.
- How to fill gaps in data-sparse regions?
 - Short historical record sites, new sites, sites rejected by MASH
 - Could these be added in to subsequent homogenisation rounds?
- Recommendations for spatial interpolation methods



THANK YOU FOR YOUR ATTENTION

Any questions?

Kay.Shelton@jbaconsulting.com

