



Australian Government
Bureau of Meteorology

The WMO homogenization guidance

Blair Trewin

Australian Bureau of Meteorology

With thanks to: Victor Venema, Tamas Szentimrey, Monika Lakatos, Enric Aguilar, Ingeborg Auer, Jose Guijarro, Matthew Menne, Clara Orio, Wilfred Louamba, Ghulam Rasul



Australian Government
Bureau of Meteorology

What is the purpose of the guidance?

- Supporting national meteorological services in creating and maintaining operational homogenized climate data sets
- Providing a baseline set of standards to support this

Many advanced users will be beyond the scope of the guidance (e.g. daily data are out of scope)



Australian Government
Bureau of Meteorology

Many national meteorological services have limited capacity

- Sparse networks and limited historical data/metadata
- Few or no staff with specialised training in data analysis
- Limited computing resources – tools which can be run on a standard PC without needing coding skills are important

Baseline standards important everywhere

These cover areas such as:

- Maintaining homogenized data sets is important!
- The need to maintain good metadata
- Managing change effectively (e.g. through parallel observations)

These standards give guidance to managers (and can be used as part of contracts where relevant)

The structure of the guidance

1. Prerequisites to developing homogenized data sets
2. Homogenization practice
3. Selection of software
4. History of homogenization
5. Theoretical background

(The first three chapters will be most important to most users; chapters 4 and 5 are for advanced users)

Some of the key inputs into homogenized data sets

- Metadata
- Data quality control
- Data rescue
- Effective change management (e.g. system testing, parallel observations)

These won't always have been done historically, but are important going forward

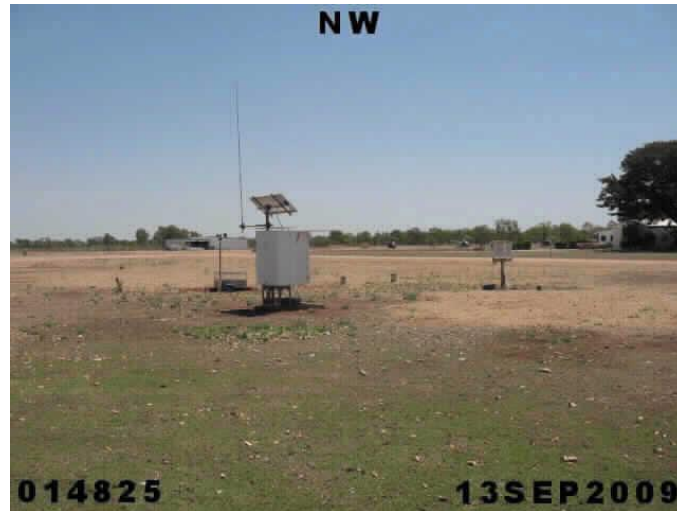


Australian Government
Bureau of Meteorology

Good metadata helps, but usually not the full solution



2005



2009

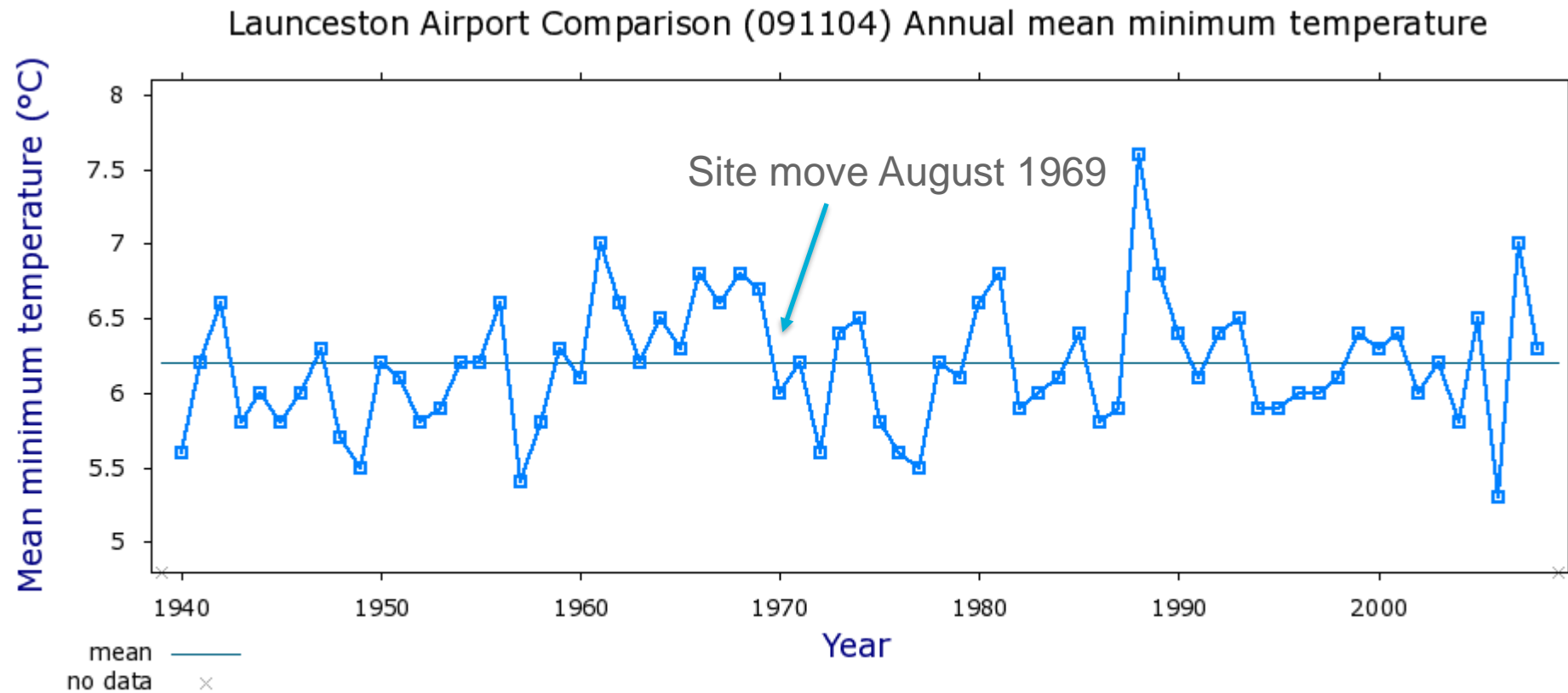
Good metadata often kept for modern period, but often incomplete or imprecise historically, and may require interpretation

Often only limited metadata available digitally

Some key changes are not easily expressed as an “entry in a database”



An example of non-conventional metadata



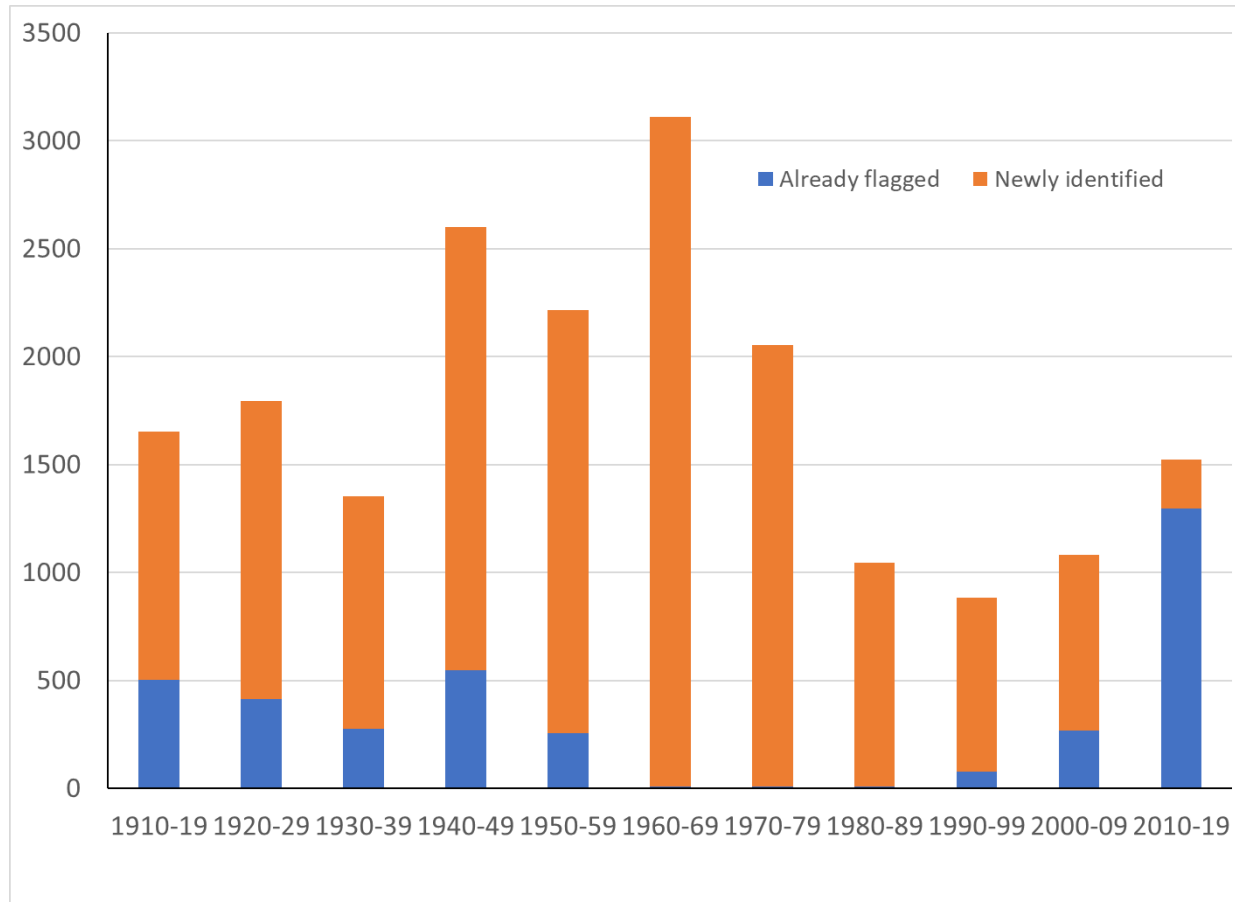
Note: Data may not have completed quality control
Observations made before 1910 may have used non-standard equipment

Climate Data Online, Bureau of Meteorology
Copyright Commonwealth of Australia, 2020



Australian Government
Bureau of Meteorology

Consistent quality control an important part of data sets



Number of suspect observations flagged in ACORN-SAT data set

Most NMHSs will have QC systems for current data

These may not have been in place historically, so QC of the full data set is important

QC of older data sometimes more difficult (lower time resolution, less comparison data)



Data rescue also an important part of many long-term data sets

9

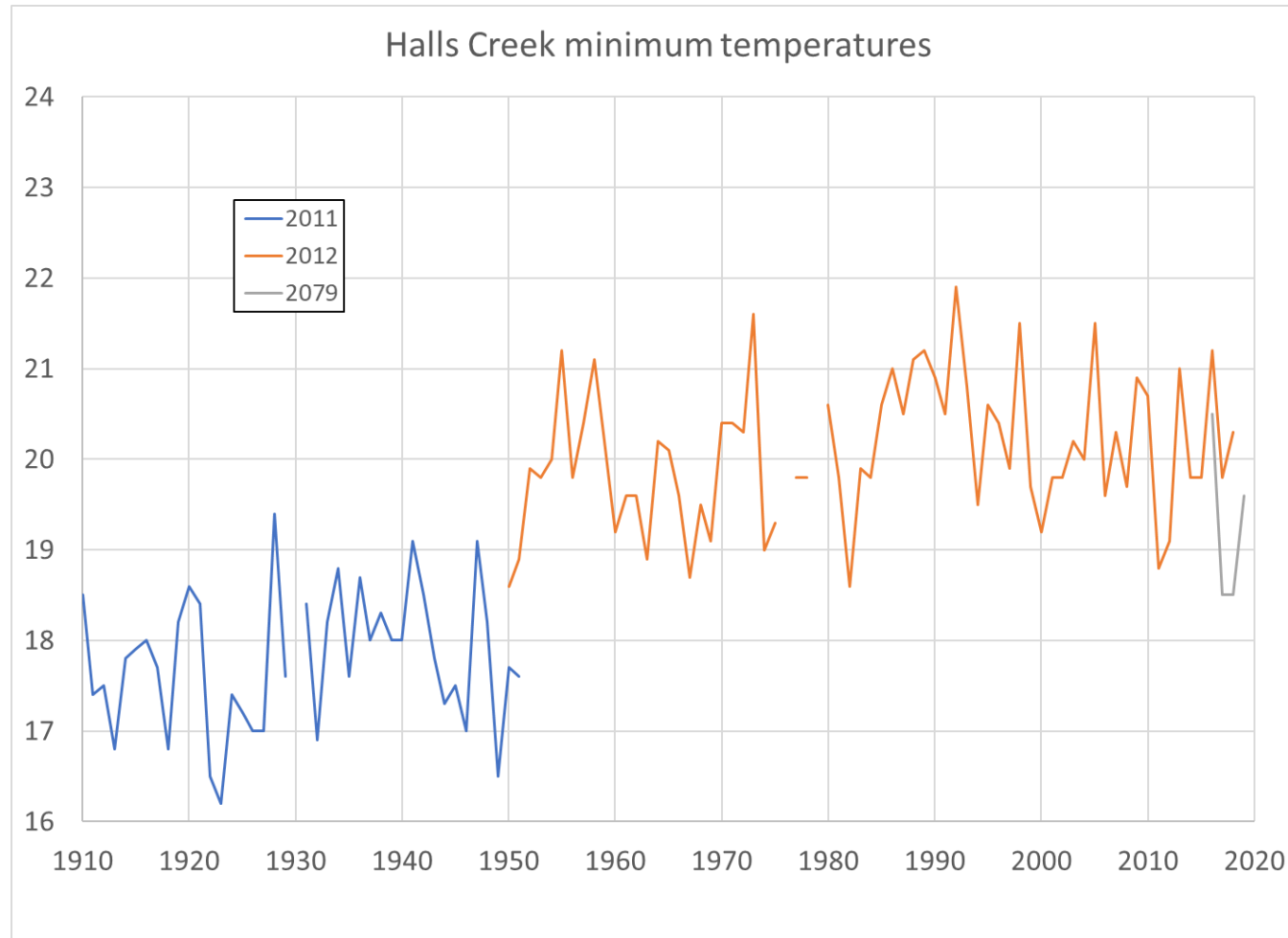
Root Darg (90° or over) underlined in Red.

Daily Max. and Min. Temp

Date.	1857		1858		1859		1860		1861		1862		1863		1864		1865		
	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	
1			90.0	68.0	76.8	64.8	78.0	61.0	99.0	74.0	92.0	63.6	91.0	56.9	84.2	58.5	86.4	53.2	87.0
2	100.2	64.2	82.0	55.3	74.0	59.0	78.0	60.0	100.0	71.0	84.0	70.8	93.9	68.8	81.5	56.5	89.0	57.6	100.0
3			89.0	55.9	71.8	56.8	78.0	58.3	99.0	73.0	94.8	72.0	78.0	64.0	80.8	61.3	85.3	51.8	71.0
4	93.5	68.8	91.7	56.8	72.0	54.0	86.0	51.2	80.2	67.5	97.0	70.2	78.5	59.5	81.8	55.0	98.0	60.0	70.0
5			95.1	57.9	77.7	55.8	92.0	56.2	78.5	56.0	90.0	64.9	87.0	63.1	76.0	56.0	73.5	59.0	71.0
6			104.8	63.7	77.6	50.0	93.5	57.2	75.3	57.0	87.8	63.2	82.0	59.1	83.5	53.2	75.0	49.5	87.0
7	101.5		107.0	68.2	79.0	51.2	88.0	57.2	75.2	57.9	103.6	77.0	81.0	57.9	91.8	63.5	77.2	50.7	91.0
8			98.0	63.4	71.1	53.7	77.0	55.4	80.0	56.8	98.1	74.0	84.0	61.0	91.0	66.3	84.0	55.0	71.0
9			77.8	62.2	73.8	58.0	87.5	58.3	94.0	56.5	93.5	80.8	96.0	64.8	89.2	65.0	83.9	54.0	71.0
10			73.5	54.9	72.0	57.8	102.5	67.0	103.0	64.2	77.7	60.7	86.0	67.2	85.5	65.0	80.0	57.3	81.0
11			75.0	55.0	82.0	50.8	91.0	74.3	103.0	70.0	73.2	59.2	83.5	63.5	96.5	67.5	82.5	53.5	81.0
12			84.0	55.2	93.2	57.2	86.3	69.0	108.0	80.0	74.9	53.2	94.0	63.0	81.5	59.0	99.0	54.0	100.0
13			96.0	57.0	82.7	70.4	88.0	67.0	107.3	78.0	79.5	55.9	80.5	65.3	92.0	56.5	79.5	72.6	71.0
14			95.0	71.0	88.7	64.7	93.7	63.5	82.3	65.0	75.0	57.8	72.0	61.0	91.5	63.5	73.0	56.0	71.0
15			102.0	73.6	78.0	56.8	97.5	73.5	88.3	58.0	75.0	54.6	72.0	58.9	74.5	53.0	74.0	59.5	81.0



Parallel observations important for major changes going forward



Site moves:

1950: town to meteorological office
2015: meteorological office to airport



Australian Government
Bureau of Meteorology

The value of working with network managers



Wittenoom, Australia

- Site established 1952 (good record in remote area) and has exceptionally high data completeness
- Mining ghost town, only a few residents left by 2000
- Installing AWS nearby impossible because of asbestos contamination



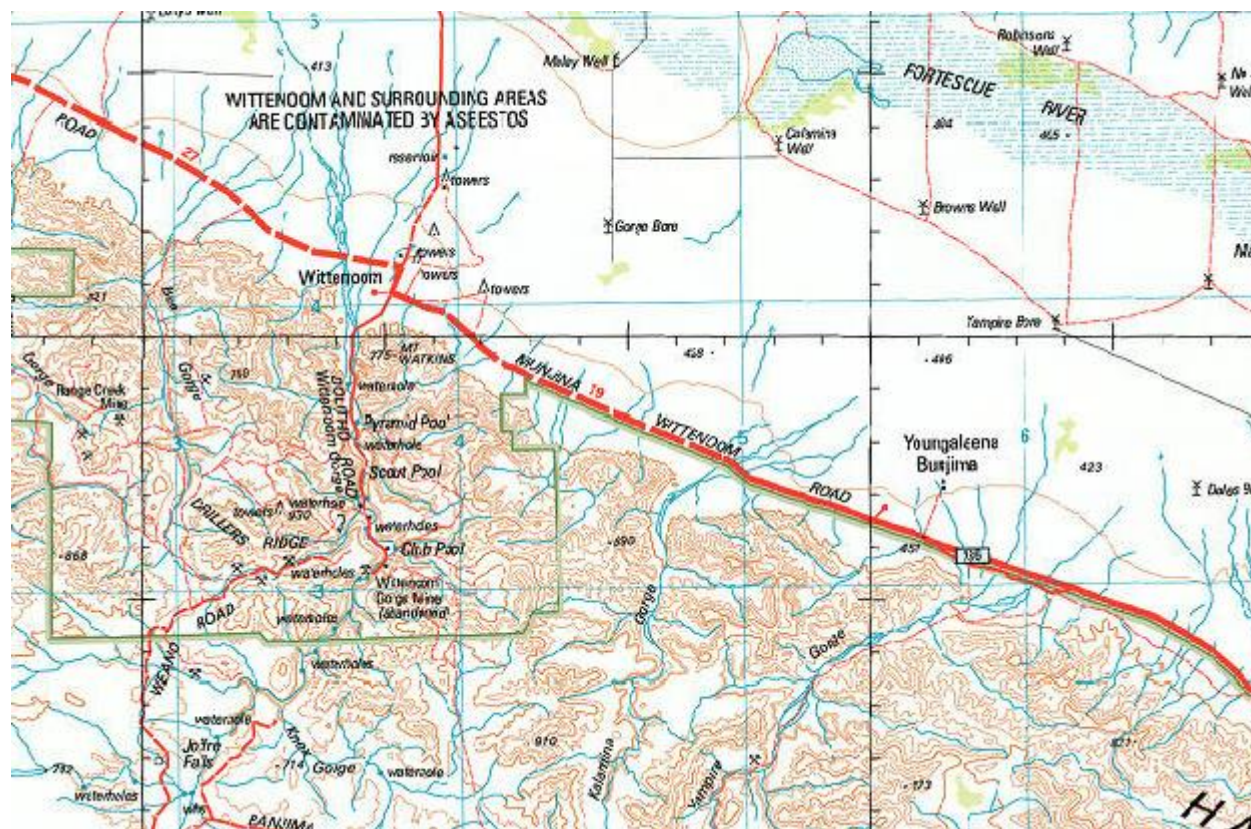
Australian Government
Bureau of Meteorology

Also an area with complex topography





Also evidence of an unusual local microclimate



Wittenoom site has a high frequency of extreme high minimum temperatures

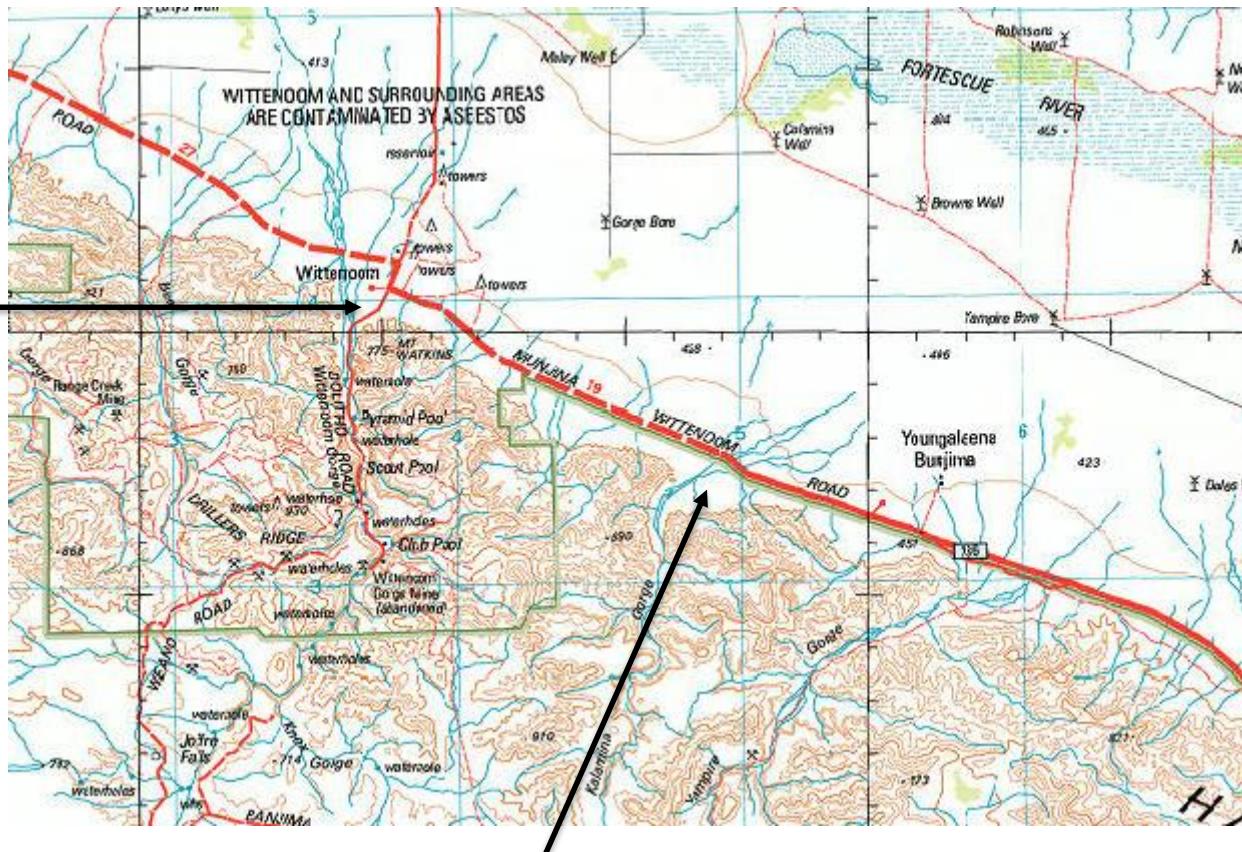
Has Australia's hottest night of the year in ~30% of years

Suspected to be due to mixing of the boundary layer due to katabatic winds flowing out from Wittenoom Gorge



Finding a new site to match the existing one

Old site



New site

Wanted a site near the outlet of a different gorge

Several other gorges exist along the escarpment extending 30km ESE from Wittenoom

Suitable site identified 12km to ESE

12 months of parallel data 2018-19

New site has mean minimum temperatures 0.5 C cooler than old site

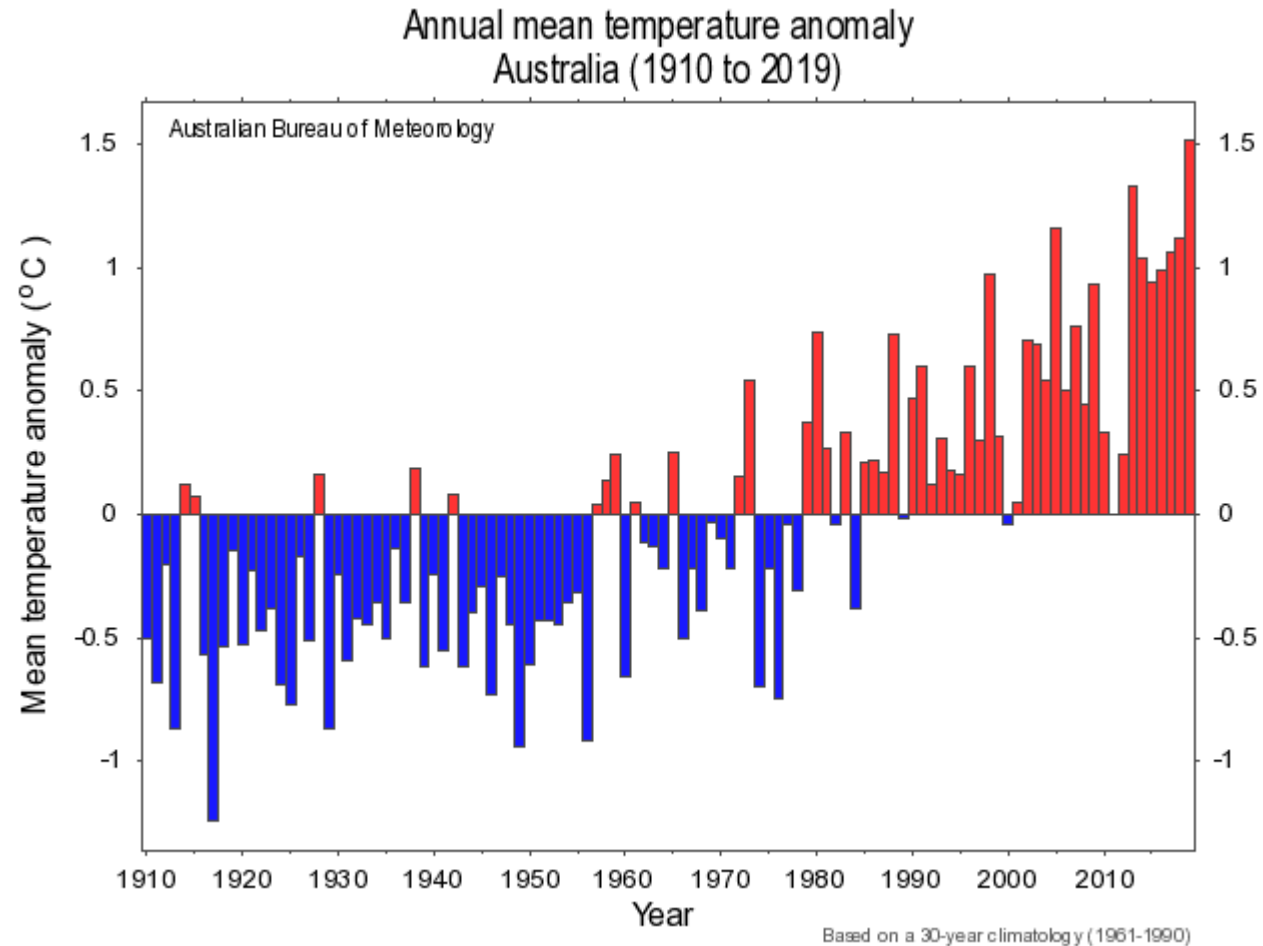
Hottest night of 2018-19 summer similar at both sites (34.2 at old, 33.9 at new)

Homogenisation practice – some of the concepts covered in the guide

- Which data to use
- Detecting inhomogeneities (combining metadata and statistical methods)
- The concept of a reference series and how to construct one
- Methods of adjustment
- Multiple rounds of homogenisation
- Documenting breakpoints and adjustments

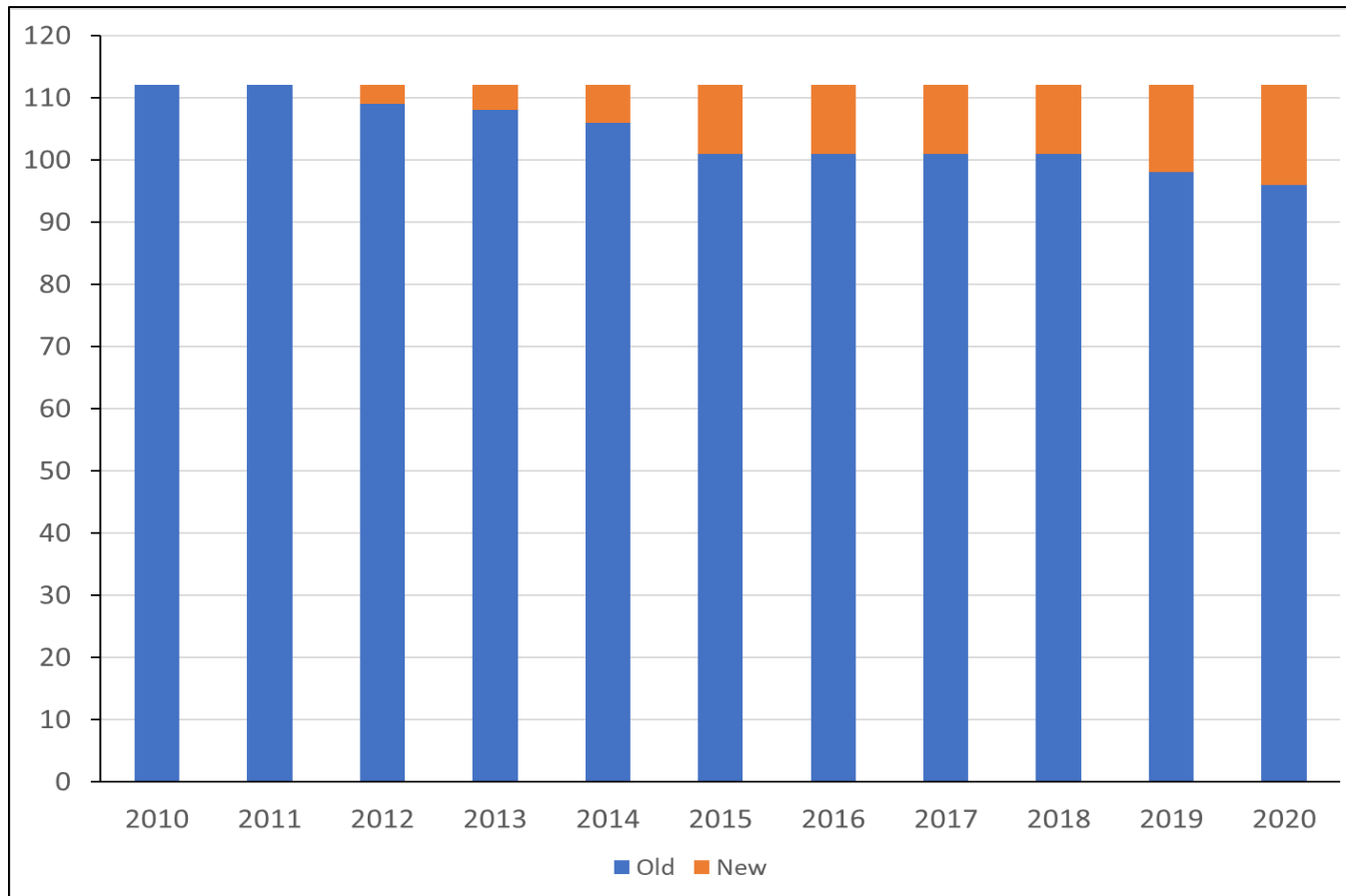


Homogenised data sets need to be maintained if used operationally





Networks evolve over time



Original Australian homogenised data set (ACORN-SAT) had 112 stations when first released in 2012

By 2020, 16 of these stations had closed and been replaced by new ones

(all with parallel observations)

V1.0

V2.0

V2.1



Australian Government
Bureau of Meteorology

Network-wide issues need to be considered specifically



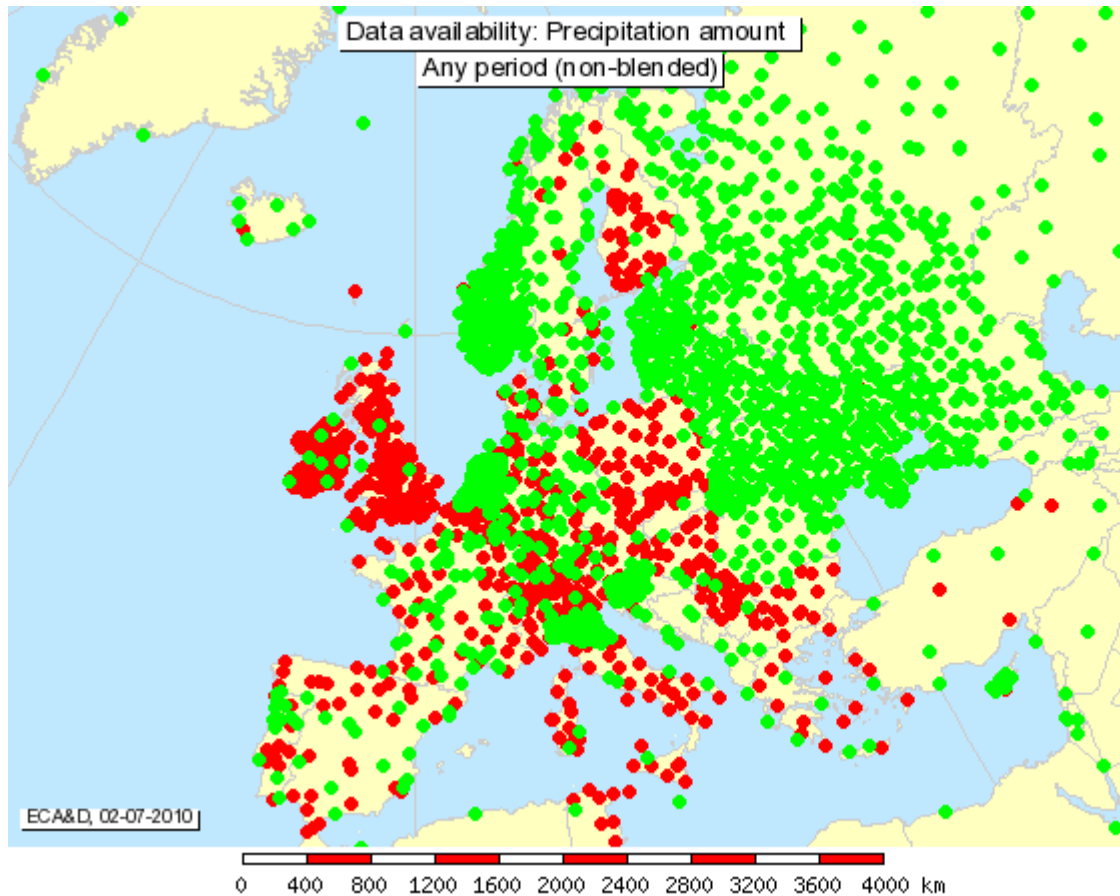
Some issues affect large parts (or all) of a network at similar times:

- Instrument changes
- Screen type changes
- Observation time changes

Reference stations will not always help in resolving these!



Some specific issues for multi-national data sets



- Data sets often large, so only automated methods feasible
- Not all stations are made available internationally, limiting availability of reference data
- Metadata may be difficult to access
- Sometimes multiple data sources must be reconciled



Australian Government
Bureau of Meteorology

Selecting homogenization software

- There are many software packages in existence
- User requirements will vary (data set size, tolerance of missing data, automated/semi-automated/manual, availability of metadata/reference series)
- Users will also have different levels of computing available
- Many users will have little knowledge of software and need guidance

Recommendations are given on some packages useful for particular use cases



Package	Version	License	Open source	Operating System	Program type	Primary operation	Availability
ACMANT	3.1	Freeware	Yes	DOS/Windows	Executable	Automatic	http://www.c3.urv.cat/data/ACMANTv3.1.zip
AnClim ProClimDB	?	Freeware	No	Windows	Executable	Interactive (and automatic)	http://www.climahom.eu/
Climatol	3.0	GPL	Yes	(Most)	R package	Automatic	http://www.climatol.eu/index.html
GAHMDI HOMAD	?	GPL	Yes	(Most)	R source R/Fortran	Automatic Interactive	mail to andrea.toreti at giub.unibe.ch
GSIMCLI	0.0.1	GPL	Yes	(Most)	Python	Automatic (and interactive)	http://iled.github.io/gsimcli/
HOMER	2.6	GPL	Yes	(Most)	R source	Interactive	http://www.homogenisation.org/
MASH	3.03	Freeware	No	DOS/Windows	Executable	Automatic (and interactive)	https://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/
ReDistribution Test	?	Freeware	Yes	(Most)	R source	Interactive	mail to predrag.petrovic at hidmet.gov.rs
RHtests	4	Freeware	Yes	(Most)	R source	Interactive	http://etccdi.pacificclimate.org/software.shtml
USHCN	52i	Freeware	Yes	Some linux versions	Fortran source	Automatic	ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/software/52i/phav52i.tar.gz



Australian Government
Bureau of Meteorology

Package	GUI	Time resolution	Input format	Metadata use	Detection method	Ref. series selection	Detection statistic	Climatic variables
ACMANT	No	Monthly & daily	ASCII	No	Reference	Correlation	Caussinus-Lyazrhi	Temperature and precipitation
AnClim ProClimDB	Yes	Any	ASCII DBF	Yes	Ref. and pairwise	Correlation & distance	Several	Any
Climatol	No	Monthly & daily	ASCII	Yes	Reference	Distance	SNHT	Any
GAHMDI HOMAD	No	Monthly Daily	ASCII	Yes	Pairwise	Correlation	New method	Any Temperature
GSIMCLI	Yes	Monthly & yearly	ASCII	No	Multiple references	Correlation & distance	User defined	Any
HOMER	No	Monthly	ASCII	Yes	Pairwise	Correlation	Penalized Likelihood	Any
MASH	No	Monthly & daily	ASCII	Yes	Multiple references	Correlation	MLR & Hypothesis test	Any
ReDistribution Test	No	Sub-daily	ASCII	No	Distribution	None	SNHT-like	Wind speed and direction
RHtests	Yes	Monthly & daily	ASCII	Yes	Reference	Correlation	Penalized max. t & F tests	Any
USHCN	No	Monthly	ASCII	Yes	Pairwise	Correlation	MLR	Temperature



Package	Correction method	Missing data tolerance	Max. number of series	Outputs				Documentation
				Homogenized series	Corrected outliers	Corrected breaks	Graphics	
ACMANT	ANOVA	Very high	100	Yes	Yes	Yes	No	User's guide
AnClim ProClimDB	Several	User defined	?	Yes	Yes	Yes	Yes	Manuals
Climatol	Missing data filling	Very high	9999*	Yes	Yes	Yes	Yes	User's guide
GAHMDI HOMAD	?	?	?	Yes	No	Yes	Yes	None
GSIMCLI	User-defined & missing data filling	High	9999*	Yes	Yes	Yes	No	Manuals
HOMER	ANOVA	15 year data	?	Yes	Yes	Yes	Yes	User's guide
MASH	Multiple comparisons	30%	500	Yes	Yes	Yes	Yes	User's guide
ReDistribution Test	None	10-20%	?	No	No	Detected breaks	No	None
RHtests	Multi-phase regression	?	1	Yes	No	Yes	Yes	User's guide
USHCN	Multiple comparisons	Very high	9999*	Yes	?	Yes	No	Plain text notes



Australian Government
Bureau of Meteorology

Thank you...