



Australian Government

Bureau of Meteorology

White paper 8: creation of quality controlled homogenised datasets from the databank

Blair Trewin, Xiaolan Wang, Phil Jones,
Matthew Menne, Robert Lund, Graham
Machin

September 2010

Why is homogeneity an issue?

- Inhomogeneities can arise in climate records for numerous reasons (e.g. site changes, changes in instruments or observing procedures, changes in local site environment)
- Inhomogeneities can mask, or exaggerate, a real climate trend



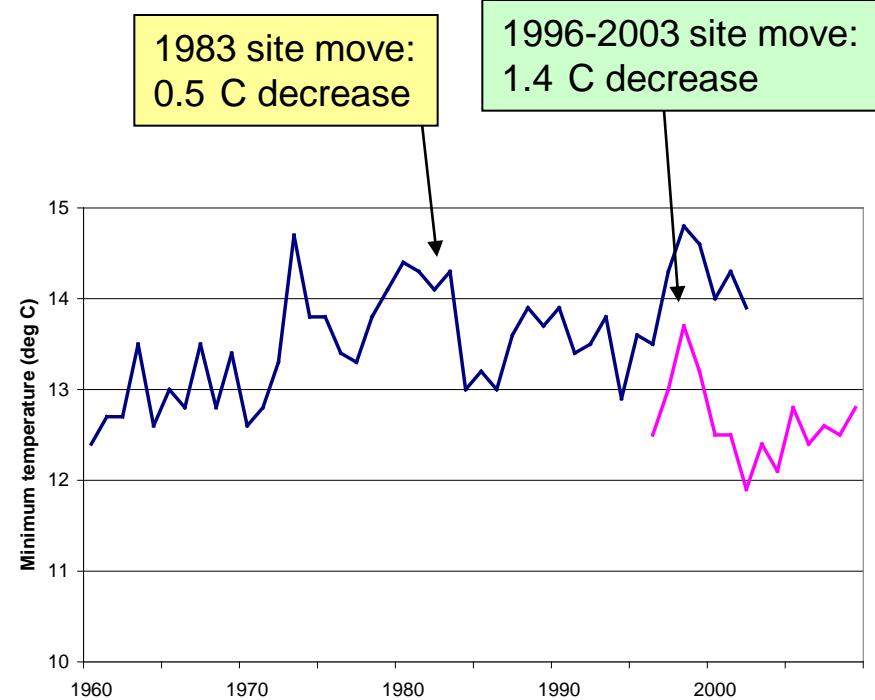
Example of an inhomogeneous data set



Site until 1983

Site 1983-2003

Site since 1996



What is involved in creating a homogeneous dataset?

- Detection of potential inhomogeneities
- Removal of inhomogeneities by adjustment
- Data quality control is also important, especially for extremes.

Detection of inhomogenities

- Metadata
- Statistical methods

Statistical methods for detection of inhomogenities

- Several methods exist (e.g. two-phase regression (TPR), standard normal (SNH))
- Different methods likely to be best suited to different situations
- Methods will vary for known and unknown changepoints
- Methods developed for single changepoints may not handle multiple changepoints or anomalous trends

Reference series

- Usual to use reference series (e.g. neighbours) to reduce noise in time series being used for changepoint detection
- Options for exploring alternative reference series in some cases (e.g. SSTs, upper-air, reanalyses)

Network-wide inhomogenities

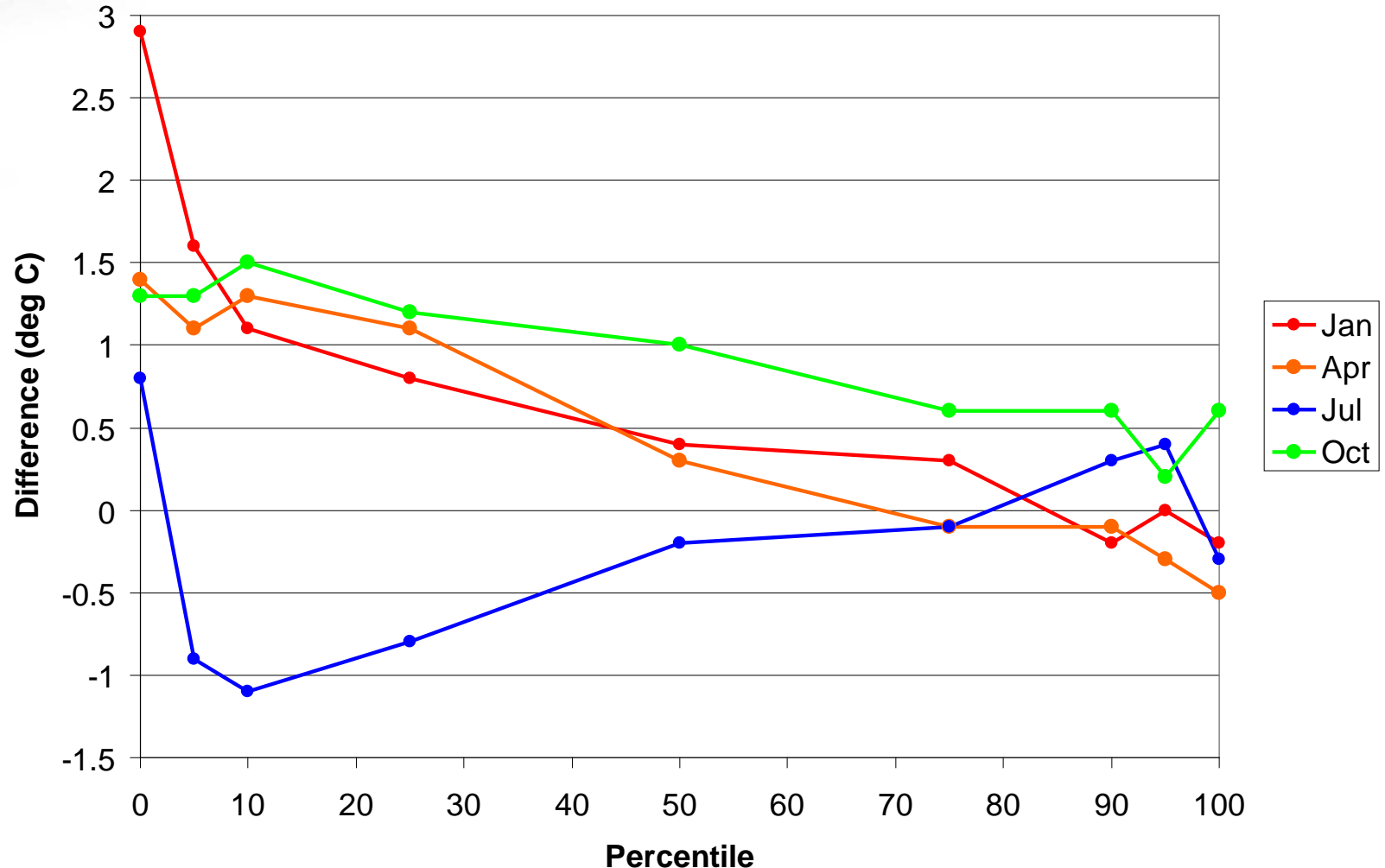
- Sometimes a change will affect a whole network at the same/similar times
 - * Observation time change
 - * Instrument shelter change
- These are difficult to detect/assess as neighbours cannot be used for reference.

Adjustment to remove inhomogeneities

- Adjustments also typically calculated using reference series
- Most methods so far in use base adjustments on monthly/annual data

Sometimes it's a bit more complicated

Comparison – minimum temperatures at Perth Airport and Perth City



Homogenising means will not necessarily homogenise the extremes

Homogenisation of daily and sub-daily data

- Some techniques developed for daily adjustment – limited implementation so far
- Adjustment of sub-daily data almost completely unexplored (though HOME project may address this)

Some homogenisation questions

- What are the most effective techniques for detection and adjustment?
- What timescale should detection be done at?
- What timescale should adjustment be done at?
- What is/are the 'base' series used for analysis?

How might homogenisation be carried out?

- May be easiest at national scale – more access to data, metadata
- Easily portable tools will be important – RHTest is a start here
- Manual intervention is very time consuming – do automated methods produce acceptable results?

Other issues

- Will be important to assess uncertainty, but how?
- Documentation of all decisions very important