WMO guidance on homogenisation - Task Team on Homogenisation

Victor Venema

WMO guidance

- Paper report
  - Part 1: getting started
  - Part 2: Advanced users
- FAQ in internet
- List of homogenisation packages in internet

- Discussion
  - Comments from readers
  - Questions to readers
  - How to proceed

Length period to compute adjustments

- When computing corrections should be take full homogeneous subperiod or limit the length of the period?
  - Longer period: higher chance of remaining breaks
  - Shorter period: additional benefit of more data saturates
  - Is this different for mean adjustments than for adjustments of the variability around the mean?
  - Do we have literature on this?

How do you combine documented break with statistical break?

- Typical to use the date of documented break
  - Do we have any idea of the error rate of documented breaks?
- For a large break the statistical break can be more accurate
- To combine them optimally one would need an estimate of the uncertainty of documented breaks

Resolution for detection

- Do we have studies that show whether daily, serial monthly or annual data is best for homogenisation?
  - Idealised for white noise
  - Realistic with seasonal cycle and autocorrelations
- Parallel monthly or parallel seasonal is different because you can see change in seasonal cycle
  - Also explicit size of the seasonal cycle (ACMANT)

Period to adjust to

- Common is to adjust to the latest homogeneous subperiod.
  - In the time of manual observations this was likely also the most reliable period
  - Otherwise mainly for convenience of updating
- Shouldn’t the principle be to adjust to the most reliable part of the data to minimize errors?
  - Especially in case of only mean adjustments, applying the adjustments to newer data should be easy.

Network-wide inhomogeneities

- Are network-wide inhomogeneities typically well-known by the network operator?
- Are they better known than inhomogeneities in specific stations?

Selection of candidate stations

- Some select the best
  - Smaller dataset is less work
  - Do bad stations make the homogenisation of others or of the network-mean worse?
  - What are selection criteria?
- Some use all available
  - Better signal to noise ratio
  - No need for selection

Precision of break date

- Annual data detection
- Monthly data more precise
  - A: Greater precision at least in part illusory because uncertainty is a few months
  - B: precision depends on SNR, if the break is large (SNR > 1) break would be precise