EFFICIENCIES OF HOMOGENISATION METHODS: our present knowledge and its limitation

Peter Domonkos\textsuperscript{1}, Victor Venema\textsuperscript{2}

\textsuperscript{1}Center for Climate Change, Univ. Rovira i Virgili, Campus Terres de l’Ebre, Tortosa, Spain, e-mail: peter.domonkos@urv.cat
\textsuperscript{2}Meteorological institute of the University of Bonn, Germany

7th Homogenisation Seminar, Budapest, 2011.
• 1. Verification of detection parts only (Domonkos, 2006)
• 2. Benchmark verification results
• 3. Problem of small, platform-like IHs (when they are frequent and dominating)
• 4. Correction of IHs.
Methods

• Calculation of efficiency:

\[ \text{Eff} = \frac{\text{RMSE}_{\text{raw}} - \text{RMSE}_{\text{hom}}}{\text{RMSE}_{\text{raw}}} \times 100\% \]

• Network-means:
  a) Average of \( K \) stations for year \( j \):

\[ \bar{X}_{j,K} = \frac{1}{K} \sum_{k=1}^{K} x_{j,k} \]

denotation when \( K = K_{\text{tot}} \):

\[ \bar{X}_j \]
Methods

• Network-mean for year $j$ ($T_j$), when $K$ available is less than $K_{\text{tot}}$:

$$T_j = \overline{X}_{j,K} + \frac{1}{m} \sum_{i=1}^{m} \left( \overline{X}_i - \overline{X}_{i,K} \right)$$

$i = 1,2,…m$ are years with available data in each station.
Methods

• ANOVA (Calculation of correction-terms with equation-system, Caussinus and Mestre, 2004).

\[ x_{\text{clima}} + x_{\text{site}} + \varepsilon = x_{\text{obs}} \]

\( x_{\text{clima}} \) for a given time is common for each site. the solution searched for \( \varepsilon = 0 \).
Denotations

ES – RMSE of annual values
ET – RMSE of trend-bias
Bay – Bayes method
C-M – PRODIGE
E-P – Easterling – Peterson method (FTP)
MAS – Multiple Analysis of Series for Homogenization (MASH)
MLR – Multiple Linear Regression
SNH – Standard Normal Homogeneity Test for shifts only
SNT – Standard Normal Homogeneity Test for shifts and trends
tts – sequential t-test
WRS – Wicoxon Rank Sum test
Detection part only: 1 break/series
Detection part only: 5 breaks/series
Detection part only: HU standard
Detection skill: HU standard

Bar chart showing detection skill for different categories.

Categories: Bay, C-M, E-P, MAS, MLR, SNH, SNT, tts, WRS.
Benchmark: annual RMSE; Trend
Benchmark: network-mean trends

PROD MASH ACMA USHC Crad SNHT Clim PMT
1900-1999 1950-1999
Detect. part only: small platform-IHs

![Bar Chart]

- **Bay**
- **C-M**
- **E-P**
- **MAS**
- **MLR**
- **SNH**
- **SNT**
- **tts**
- **WRS**

Legend:
- ES
- ET

Values:
- Bay: -249
- C-M: -104
- E-P: -230
- MAS: -104
- MLR: -230
- SNH: -104
- SNT: -230
- tts: -230
- WRS: -230

X-axis: Variables
Y-axis: Values
ANOVA!: annual RMSE; Trend
ANOVA!: network-mean trends

The chart shows the network-mean trends for various locations (PROD, MASH, ACMA, USHC, Crad, SNHT, Clim, PMT) from 1900-1999 and 1950-1999. The x-axis represents the locations, and the y-axis represents the percentage change from the mean. The chart highlights the differences in trends between the two time periods.
Conclusions

• PRODIGE, MASH, ACMANT, Craddock have been proven to be the most effective homogenisation methods. USHCN also has some noteworthy characteristics.

• The true efficiencies are substantially lower than that had been calculated earlier from testing detection parts only, since the efficiencies are affected by problems out of the detection-part.
Conclusions

• In real data IH-magnitudes are often lower, and platform-shapes are likely more frequent than in the Benchmark. Therefore the true efficiencies could be even lower.

• The ANOVA for calculating correction-term is a tool for improving the accuracy of homogenisation, so it is recommendable for each homogenisation method.
Thank you for your attention!