Comparison of Homogenisation methods

Everything you wanted to know about homogenisation but where afraid to ask

Olivier Mestre  Ecole Nationale de la Météorologie
Institut de Mathématiques de Toulouse

Victor Venema  Bönn University

On behalf of COST Action ES0601 group
COST Action ES0601

- EU funded project  COST funds NETWORKING
- Start: June 2007  end: October 2011
- Intercomparison of homogenisation methods
- Synthesis of the best aspects of each method

SOFTWARE

- Methods for Daily data
- Methods for Urban trends
Salzburg annual mean temperatures

MEAN TEMPERATURE ID000014 SALZBURG
Homogénéisation : pourquoi?

- Salzburg metadata (ZAMG)

Section 5: 1884-01 to 1903-07
High School (Oberrealschule), z=419m

Section 6: 1903-08 to 1941-02-28
Studiengebäude-Lehrerbildungsanstalt, z=423m

Section 7: 1939-03-01 to 1996-06-15
Airport station 1, z=434m
Observation hours:

Section 8: Since 1996-06
Airport station 2, z=437m
Positions of station
Instruments, shelters

- 1935
- Since 1996
Comparing Salzburg with neighbours

- Package “segclust” (Picard) + penalized likelihood criteria
DETECTION

A statistical problem

But information provided by METADATA must not be forgotten

- Change-point detection: a constantly evolving field
  - Signal processing
  - Econometrics
  - Biology (DNA segmentation)
  - etc...
Change-point models

Simple linear regression

Change-point model (mean)
How many change-points?
How many change-points?
How many change-points?
How many change-points?

AIC

CAU

BIC
How many change-points?
Different procedures have different properties

Some behave well in some situations
Maybe some behave well in every situation?

We are interested in those who behave well in our case

Simulated benchmarks, synthetic noise

Courtesy of Peter Domonkos (URVCAT), Emilie Lebarbier (AgroParisTech), Franck Picard (U. Lyon)
ROC scores

ROC SCORES: $a=2.0$ sigma=1.0 noise: White

DETECTION POWER

FALSE ALARM RATE (CHANGE-POINT/SAMPLE)
Conclusion

- On those experiments, Jong, Caussinus & Lyazrhi, Zhang & Siegmund are best criteria when combined with DP, MASH, MDL and Lebarbier behave well.

- Robustness to skewness: rather good for parametric methods: the use of non-parametric procedures is not justified!

- Robustness to autocorrelation: C&L and Jong tend to over-detect, MASH is robust, Lebarbier tend to under-detect

- On same configurations, but with 10 times more data (n=1500), many procedures behave much better
TESTING THE WHOLE PROCESS

- COST Action ES0601 « HOME » : 25 contributions
- Realistic surrogate benchmark created by Victor Venema
- 20 temperature networks, 20 precipitations networks
- Realistic temporal and spatial correlations
- 6-15 series/network, varying density

Multiple change-points, seasonal behaviour, outliers, local trends

Scores computed for every method, on Truth vs Corrected, to be compared to Truth vs Raw
A BLIND EXPERIMENT
Overall results: RMSE

**TEMPERATURE MONTHLY RMSE**

**TEMPERATURE ANNUAL RMSE**
Correction of trends (SNHT)
Correction of trends USHCNv2 52x
Correction of trends (PRODIGE)
Correction of trends (MASH)
Correction of trends (SNHT)
Correction of trends (USHCNv2 52x)
Correction of trends (PRODIGE)
Correction of trends (MASH)
First conclusions

• Relative methods are to be privileged versus absolute
• Methods not relying on « homogeneous references » behave better
• Detection is improved by multiple change-points dedicated methods
• For expertise, training is important
• Many manual methods are worse than the best automatic ones
• Automatic algorithms can achieve good results and are necessary for large networks
• Many more conclusions… way of comparing series, corrections, references vs pairwise, etc...
Software

- Detection
  - DP + penalized likelihood criteria
  - Implement joint detection on summer/winter series (ACMANT)

- Attribution of changes
  - Improvement of USHCNv2 system using bayesian posteriors (Hannart)
  - Joint segmentation using MLM (Lebarbier – Picard)

- Correction
  - ANOVA two factors model (PRODIGE)
  - GLS Estimation (introduction of spatial covariances, MASH)
CONCLUSION

- Most homogenisation methods improve the data
- Some methods behave better than others
- Still room for improved methods
- Comparing methods:

Do not underestimate the difficulty of the task!