European Geosciences Union General Assembly 2012

Vienna, Austria, 22 - 27 April 2012

Session: HS7.4/AS4.17/CL2.10 Climate, Hydrology and Water Infrastructure

Investigation of methods for hydroclimatic data homogenization

E. Steirou and D. Koutsoyiannis

Department of Water Resources and Environmental Engineering,
National Technical University of Athens, Greece

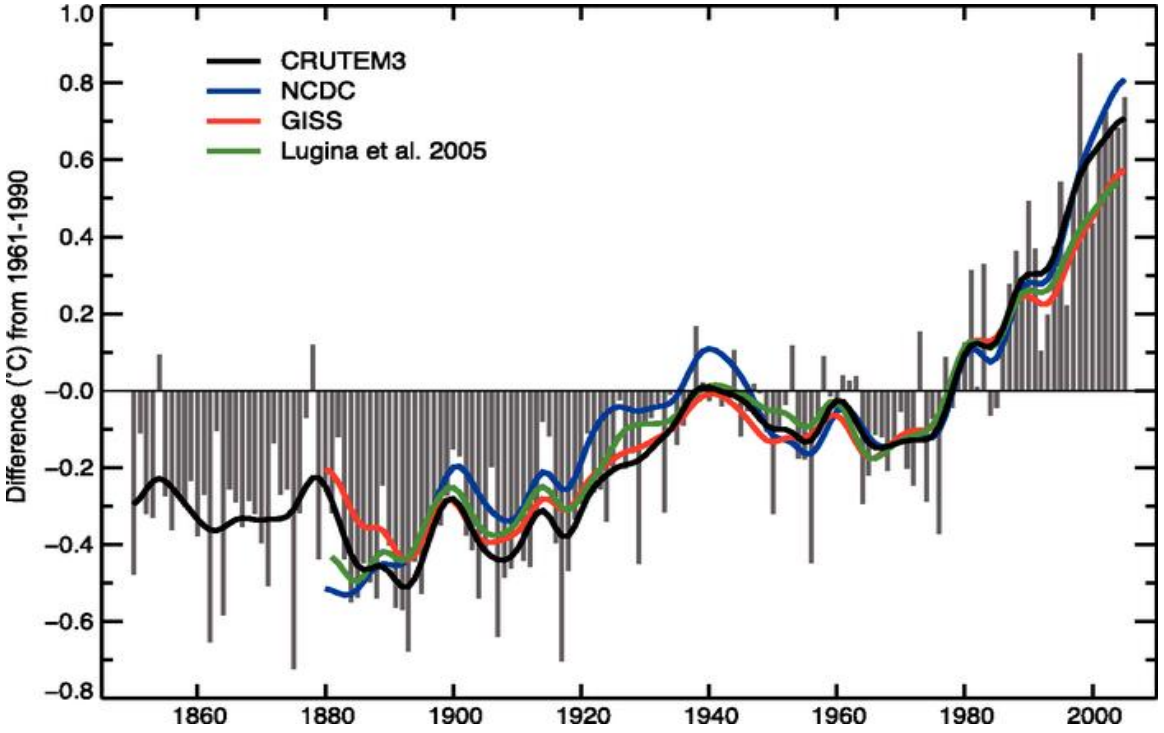
Presentation available online: itia.ntua.gr/1212

Temperature increase during the last century

- The dominant view concerning the climate change is summarised by the IPCC (Intergovernmental Panel on Climate Change) Assessment Reports.

- Fourth Assessment Report (2007): a non-uniform but clear temperature increase of 0.6 - 0.7°C is estimated during the last hundred years

► Estimations are not based on raw data but on data **adjusted** in order to remove errors.

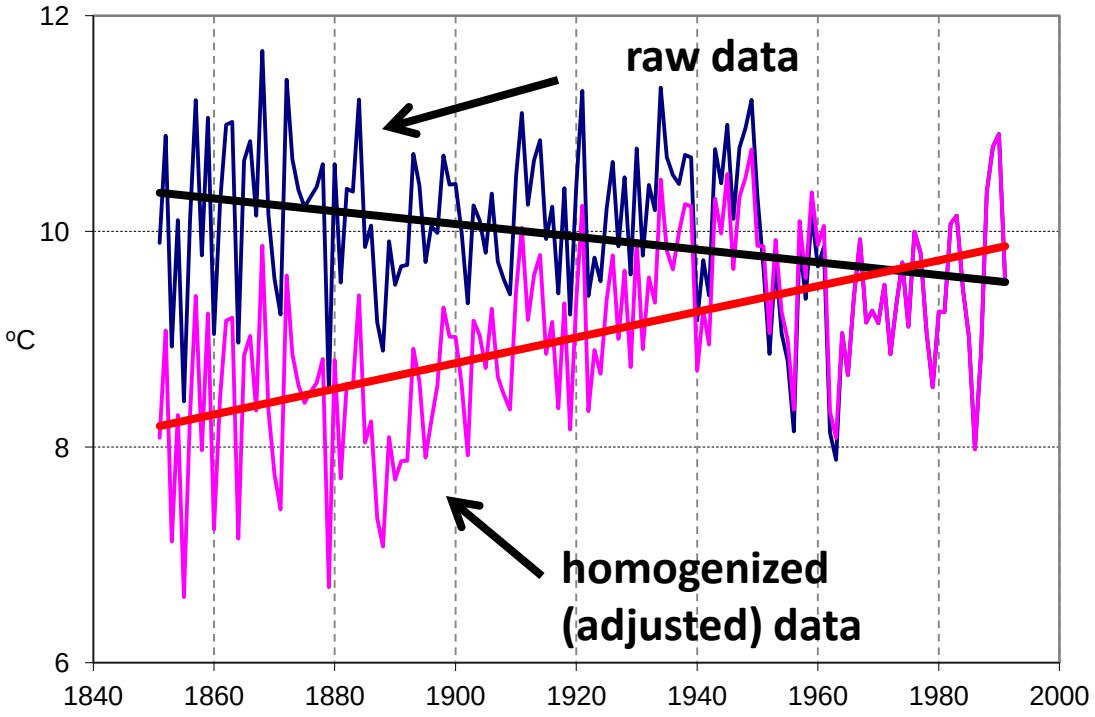


Different estimates of global temperature changes (IPCC, 2007)

The problem

- Historical and contemporary climatic time series contain **inhomogeneities** – errors introduced by changes of instruments, location etc.

- The **homogenization** of climatic time series is made with mainly statistical methods of identification and correction of recorded and non-recorded inhomogeneities and is a subject of debate.



The difference between the trends of the raw and the homogenized data is often very **large**.

De Bilt station – The Netherlands
Source: Database
GHCN-Monthly Version 2
(aggregated to annual)

Aim of our work

1. To classify and evaluate the observed inhomogeneities in historical and modern time series, as well as their adjustment methods.

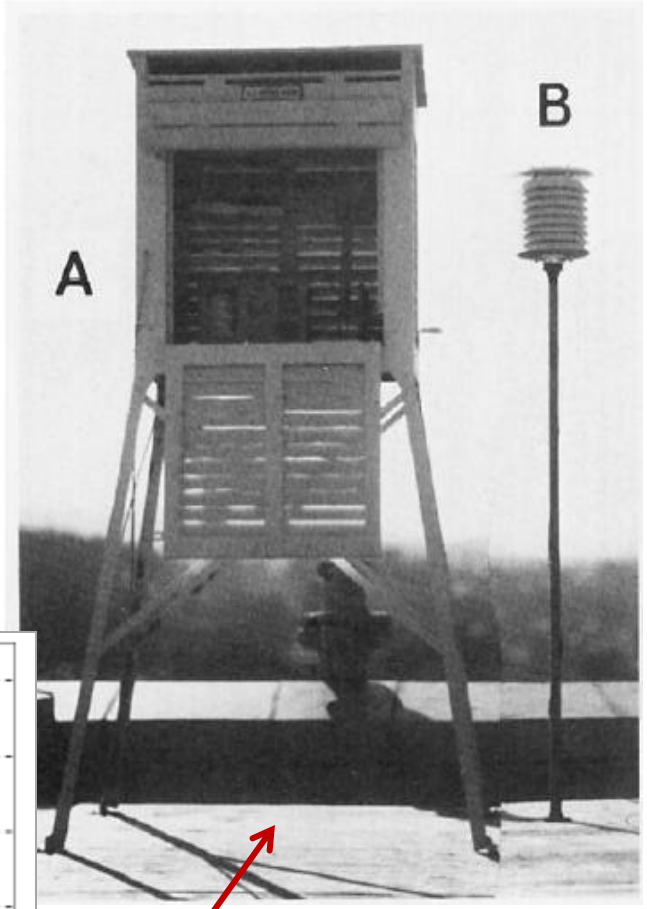
2. To investigate if and how the homogenization procedure affected temperature trends worldwide.

3. To investigate the behaviour of common homogenization methods, when applied to synthetic time series with specified statistical characteristics.

In this presentation we focus on points 2 and 3.

Inhomogeneities

- Different types (shifts, trends, outliers)
- Different causes (thermometer/recording errors, changes in measurement conditions, differences in observational hours and in the methods used to calculate the mean temperature)



Changes of instruments – shelters in the USA in the 1980s

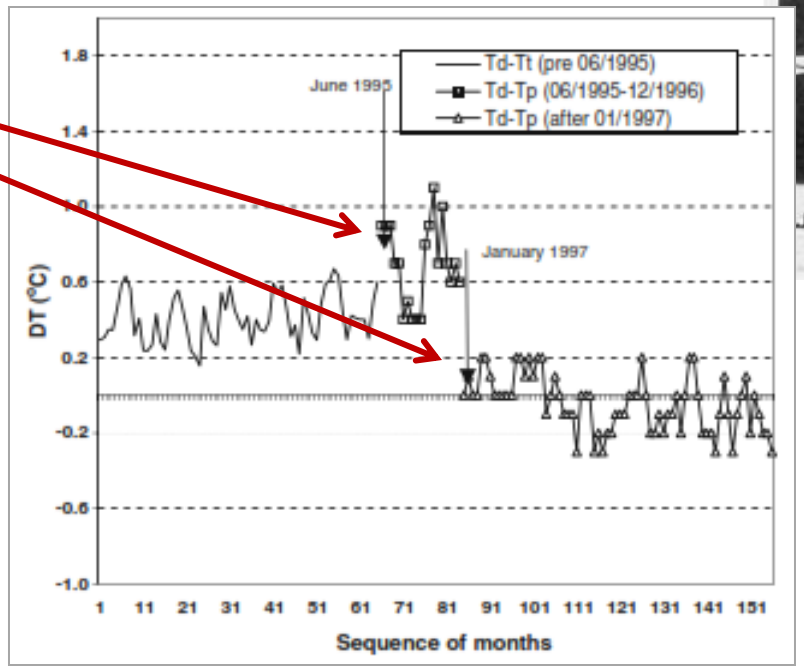
A: initial wooden Cotton Region Shelters

B: modern plastic shelters

Discontinuities in the air-temperature time series at the National Observatory of Athens:

– instrument change in June 1995

– calibration of the new thermometer in January 1997



Homogenization methods

The homogenization procedure usually consists of three basic steps:

1. Removal of outliers – usually values out of a range of ±(3σ to 5σ) are rejected
2. Corrections to account for different data/methods to estimate mean daily temperatures and corrections for recorded changes of measurement conditions
3. Application of statistical methods to remove shifts or false trends identified in a single time series (absolute methods) or in comparison of a "candidate" time series to one or more "reference" time series (relative methods, more common)

► Relative methods: they require high statistical correlation between candidate and reference series.

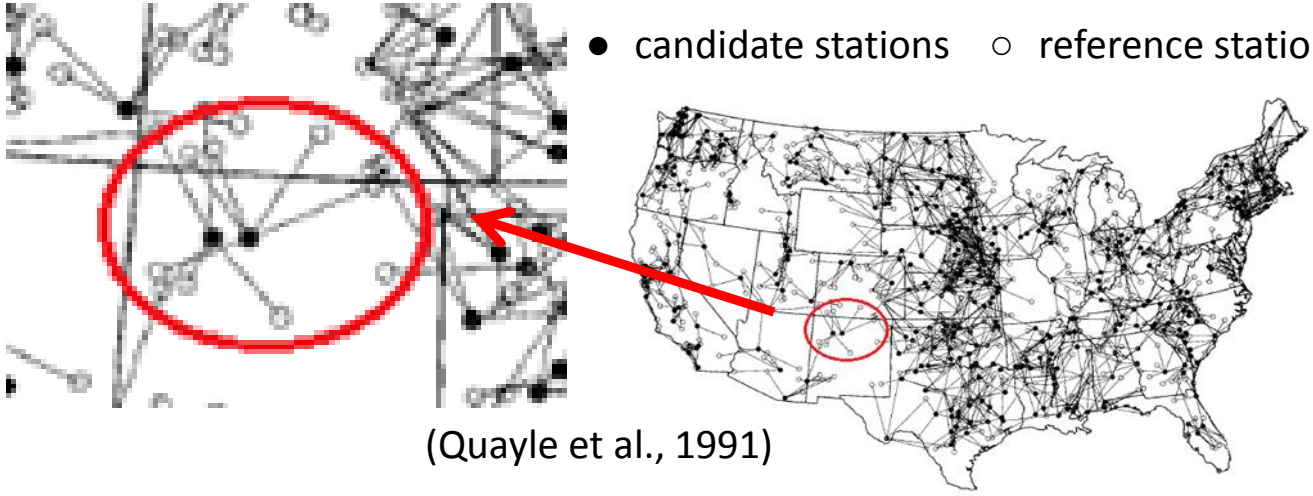
► Common assumption of homogenization methods is that temperature data (and generally hydroclimatic data) are independent and normally distributed.

Discussion on the homogenization-1

► Homogenization results are usually not supported by metadata or experiments (a known exception in literature is the experiment at the Kremsmünster Monastery, Austria).

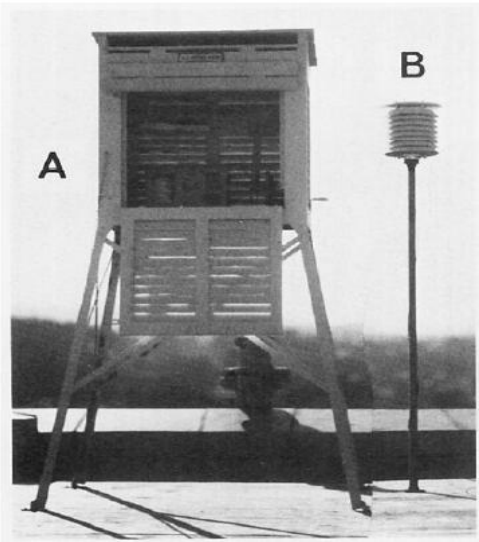
Example: change of thermometers–shelters in the USA in the 1980s (Quayle et al., 1991)

- No single case of an old and a new observation station running for some time together for testing of results is available!
- On the contrary, comparison and correction were made using statistics of remote (statistically correlated) stations.



(Quayle et al., 1991)

Two neighbouring stations are corrected based on two groups of reference stations located at distances of hundreds of km.



Discussion on the homogenization-2

► Homogenization methods do not take into consideration some characteristics of hydroclimatic data (long-term persistence, microclimatic changes, time lags).

► Some inhomogeneities detected are statistically non-significant and they can lead to false corrections.

Example: Adjustments of daily summer maximum temperatures in the Greater Mediterranean Region (Kuglitch *et al.,* 2009)

REGION	CORRECTION
Western Mediterranean	+0.03 ±0.38 °C
Central Mediterranean	+0.16±0.52 °C
Eastern Mediterranean	+0.19±0.30 °C

► Inhomogeneities not reflecting systematic instrumentation changes in a specific period are expected to have a random character, not introducing a consistent bias in long time series that needs to be corrected.

► Corrections may introduce bigger errors than the errors they try to remove.

Evaluation of homogenization results

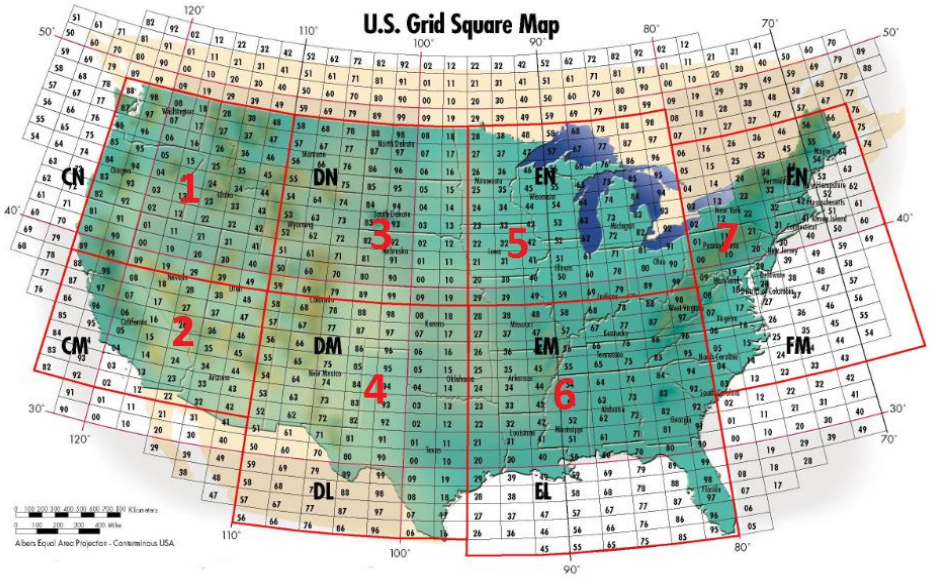
Data selection:

From the total number of stations of the database GHCN-Monthly Version 2 we examined **163 stations** worldwide satisfying certain criteria:

- They have both raw and adjusted data.
- Each time series contains ≥ 100 years of data.
- Each time series contains ≤ 4 successive missing values.
- In each time series the percentage of missing years does not exceed 10%.
- Time series end at or later than 1990.

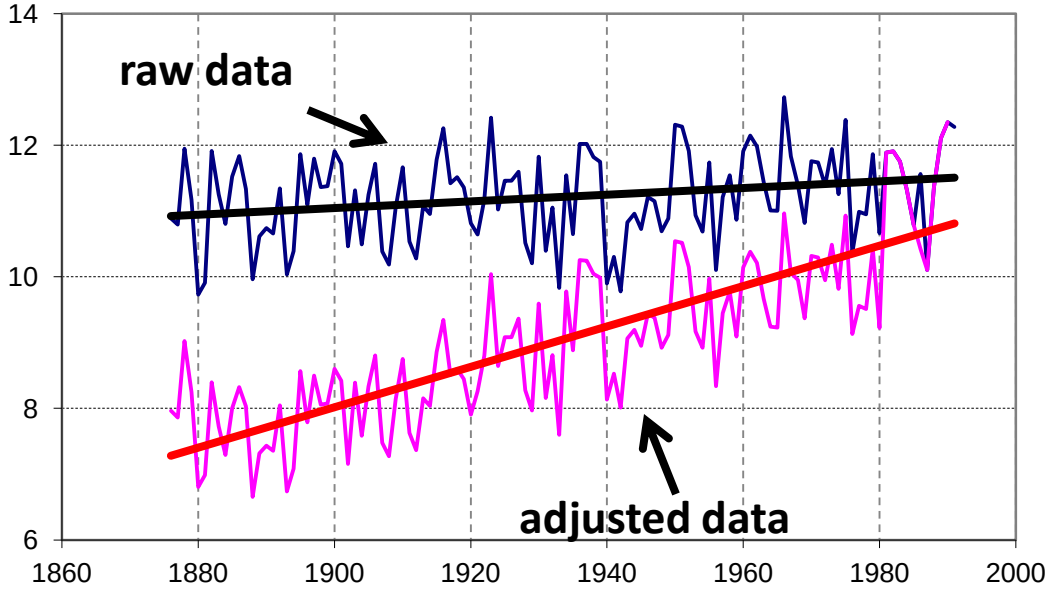
REGION	STATIONS
Africa	3
Europe	44
Asia	40
South America	5
North America	54
Oceania	17

In the USA, due to the large number of stations satisfying the criteria, we divided the region into 7 sections and selected a number of stations in proportion to their area.



Data analysis

- We calculated annual values from monthly values (a year with more than 4 missing months in total or 3 consecutive missing months was considered 'missing').
- We calculated trends for both raw and adjusted data.
- We calculated the Hurst coefficient in two cases of stations with a big difference between the trends of raw and adjusted data.



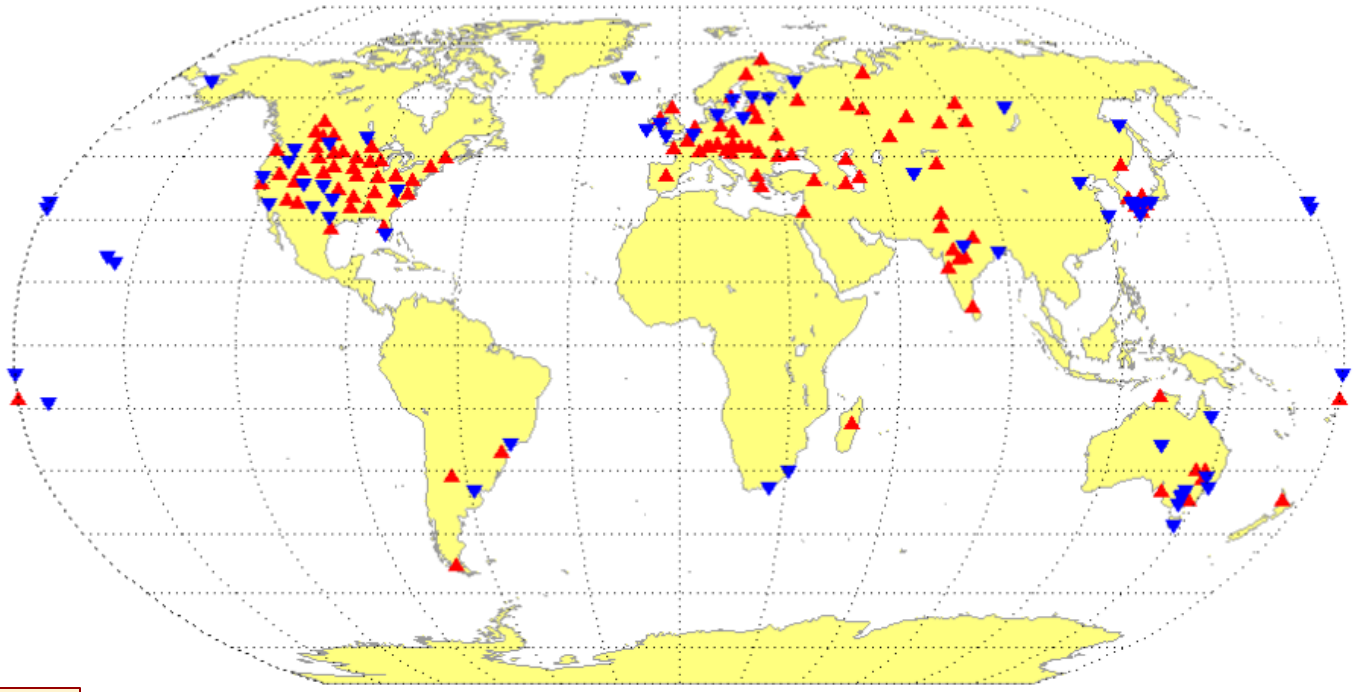
► The Hurst coefficient increased due to the trend increase of the time series.

Sulina station – Romania
Source: Database
GHCN-Monthly Version 2

Results

In 2/3 of the stations examined the homogenization procedure increased positive temperature trends, decreased negative trends or changed negative trends to positive.

► The expected proportion would be 1/2.



Trend difference due to homogenization

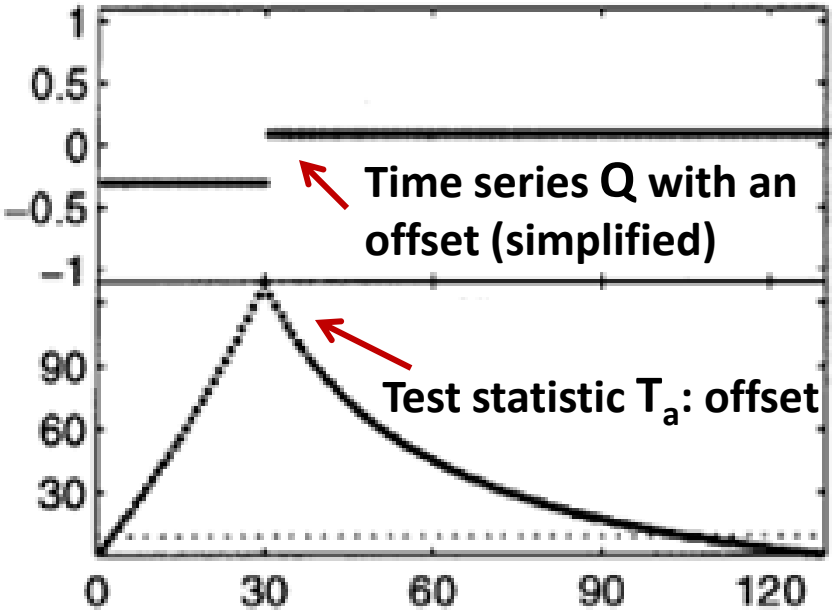
Increase ▲ Decrease ▼

Global Temperature Increase (from the examined series)	
Raw data	0.42°C
Adjusted data	0.76°C

Homogenization has amplified the estimation of global temperature increase.

Evaluation of the SNHT performance

► **Standard Normal Homogeneity Test (SNHT) for single shifts** is one of the most common homogenization methods (GHCN - Version 3) for temperature data. A version of the method is used for precipitation data.



- A time series Q is formed as a function of the candidate (tested) time series Y and a number of reference time series X_j.

$$Q_i = Y_i - \left\{ \sum_{j=1}^{k} \rho_j^2 \left[X_{ji} - \bar{X}_j + \bar{Y} \right] / \sum_{j=1}^{k} \rho_j^2 \right\}$$

- The time series Q is normalised to time series Z.
- The test creates a test statistic T_a which at the point of a shift takes its maximum value.

inhomogeneity point

$$T_{max}^{s} = \max_{1 \le a \le n-1} \left\{ T_a^s \right\} = \max_{1 \le a \le n-1} \left[a\bar{z}_1^2 + (n-a)\bar{z}_2^2 \right]$$

SNHT for single shifts

The method was applied in three different cases of synthetic time series:

1. independent data normally distributed with a shift
2. homogeneous data with long-term persistence
3. data with long-term persistence and a shift

We created two time series X, Y each one containing 100 elements and time series W as a linear function of X, Y.

$$W = \frac{X + \kappa Y}{\lambda}$$

Time series X, Y: μ=0 and σ=1
Data with long-term persistence: H=0.85, SMA model (Koutsoyiannis, 2000)
The coefficients κ, λ were calculated so that ρ_{WY}=0.9 and σ_W=1.

W: candidate series
Y: reference series

$$Q_i = Y_i - \left\{ \sum_{j=1}^{t} \rho_j^2 \left[X_j - \bar{X}_j + \bar{Y} \right] / \sum_{j=1}^{t} \rho_j^2 \right\}$$

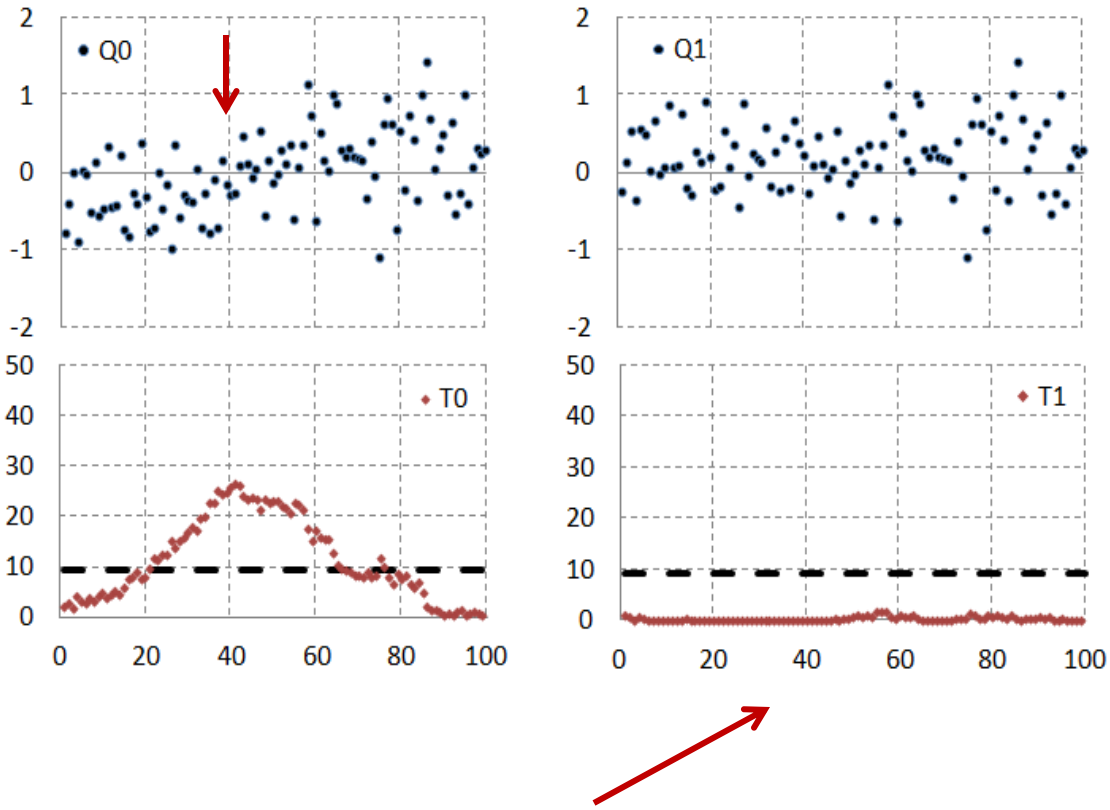
$$Q_i = W_i - Y_i + \bar{Y} - \bar{W}$$

1. Independent data normally distributed with a shift

We induced a shift of 0.5 °C to the candidate time series.

SNHT located and corrected the shift of 0.5°C.
The original trend of the time series was recovered.

Time series	Trend
W (original)	0.0038
W (adjusted)	0.0032



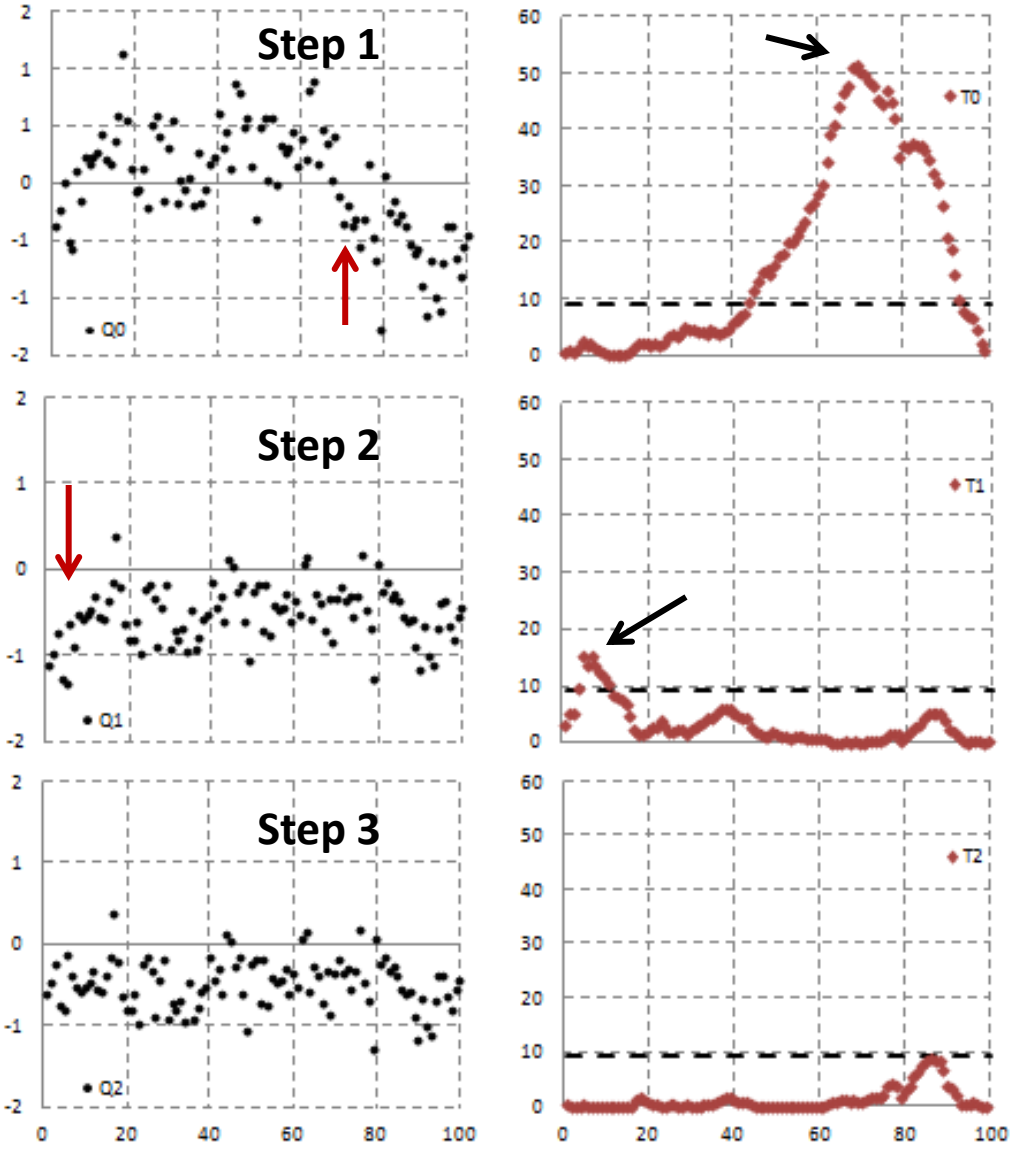
The time series is considered homogeneous

► SNHT seems to be satisfactory when applied to independent data normally distributed.

2. Homogeneous data with long-term persistence

The method detected two false (non existing) inhomogeneities. The time series was corrected in two steps even if it was already homogeneous.

► The homogenization changed the trend of the time series.



Time series	Trend	Hurst coef.
W (initial)	0.0103	0.76
W (1st correction)	0.0198	0.88
W (2nd correction)	0.0179	0.86

The observed increase of the Hurst coefficient is caused by the increase of the trend of the time series.

3. Data with long-term persistence and shift

- We induced a shift of 0.5°C after time 40.
- We applied the homogenization method until a homogenous time series was derived.

1st step – false inhomogeneity

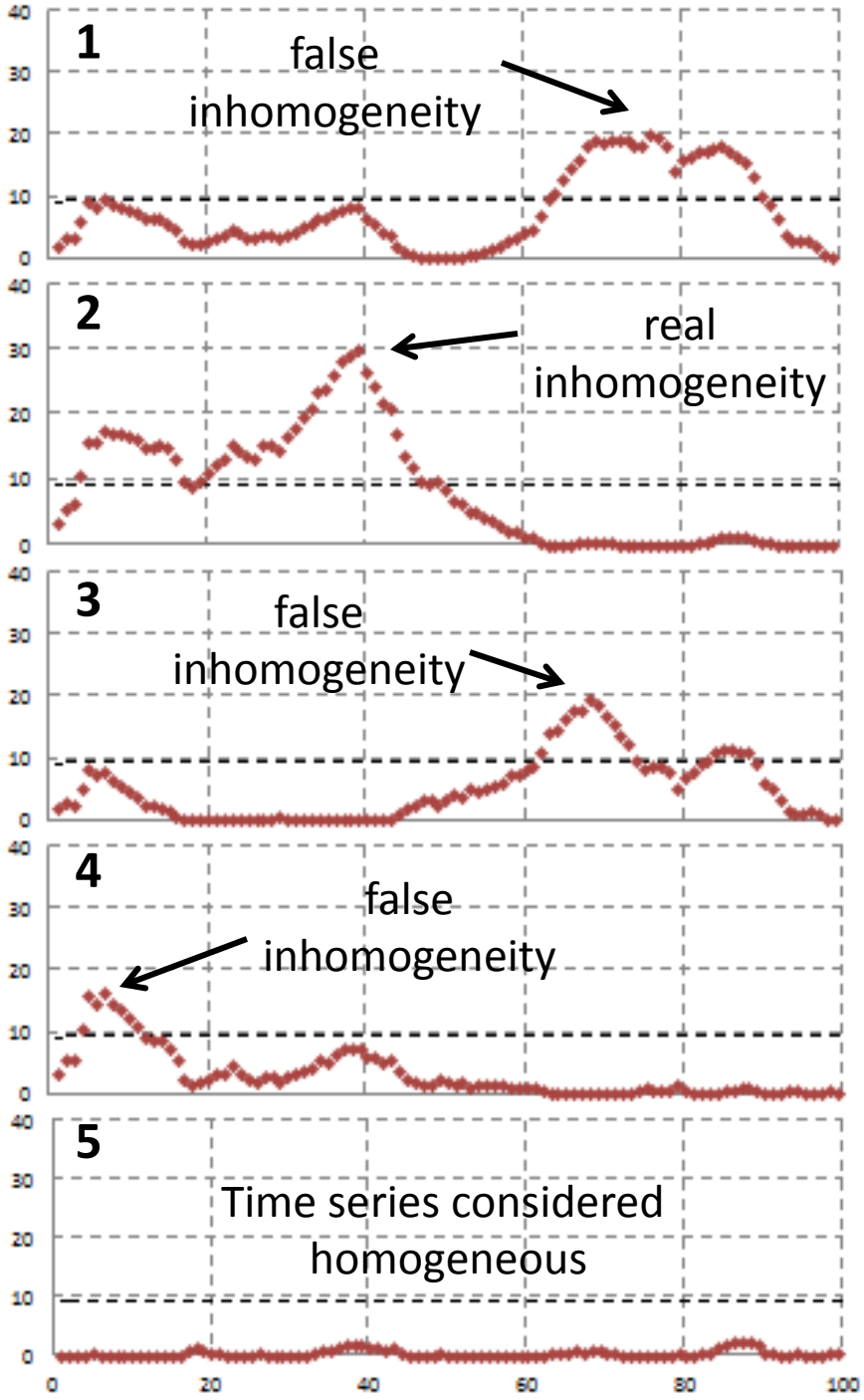
2nd step – **real** inhomogeneity

3rd step – false inhomogeneity

4th step – false inhomogeneity

- The homogenization changed the trend of the time series. Statistical characteristics similar to the homogenized time series of the previous example.

► SNHT does not seem to have a satisfactory behaviour when applied to data with long-term persistence.



Conclusions

1. Homogenization is necessary to remove errors introduced in climatic time series.

2. Homogenization practices used until today are mainly statistical, not well justified by experiments and are rarely supported by metadata. It can be argued that they often lead to false results: natural features of hydroclimatic time series are regarded errors and are adjusted.

3. While homogenization is expected to increase or decrease the existing multiyear trends in equal proportions, the fact is that in 2/3 of the cases the trends increased after homogenization.

4. The above results cast some doubts in the use of homogenization procedures and tend to indicate that the global temperature increase during the last century is smaller than 0.7-0.8°C.

5. A new approach of the homogenization procedure is needed, based on experiments, metadata and better comprehension of the stochastic characteristics of hydroclimatic time series.

References

Alexandersson H, Moberg A. (1997) 'Homogenization of Swedish temperature data Part I: homogeneity test for linear trends', *Int J Climatol* 17:25–34.

Founda, D., Kambezidis, H.D., Petrakis, M., Zanis, P., Zerefos, C. (2009) 'A correction of the recent air-temperature record at the historical meteorological station of the National Observatory of Athens (NOA) due to instrument change', *Theoretical and Applied Climatology*, 97 (3-4), pp. 385-389.

IPCC (2007) Summary for Policymakers, Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge: Cambridge University.

Koutsoyiannis, D. (2000) 'A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series', *Water Resources Research*, 36 (6), 1519–1533.

Kuglitsch, F.G., Toreti, A., Xoplaki, E., Della-Marta, P.M., Luterbacher, J., Wanner, H. (2009) 'Homogenization of daily maximum temperature series in the Mediterranean', *Journal of Geophysical Research* D: Atmospheres, 114 (15), art. no. D15108 2.

Quayle, R. G., D. R. Easterling, T. R. Karl, and P. Y. Hughes (1991) 'Effects of recent thermometer changes in the Cooperative Station Network" *Bull. Amer. Meteor. Soc.,* 72, 1718–1723.