Benchmarking homogenization computer packages: First results of the MULTITEST project

J.A. Guijarro[1], J.A. López[1], P. Domonkos[2], E. Aguilar[2] and M. Brunet[2]

[1] State Meteorological Agency (AEMET), Spain
[2] Universitat Rovira i Virgili, Tarragona, Spain

16th EMS / 11th ECAC (Trieste, September 12-16, 2016)

Outline

Introduction

- The need to homogenize observational series before its use to assess climate variability emerged long time ago and many methodologies, some of them implemented in computer packages, have been developed since then.
- Action COST ES0601 *HOME* was very useful to promote discussion meetings of homogenization specialists and to inter-compare the performances of their methods and software developments.
- As these program are being improved, new benchmarking exercises are needed to show their new achievements. (Some preliminary results are still at http://www.climatol.eu/DARE/testhomog.html)

Introduction

The Spanish project MULTITEST aims at updating and improving those benchmarking experiments in various ways:

- More realistic temperature networks
- Inclusion of precipitation networks with different climatic characteristics (Temperate, Mediterranean and Monsoonal)
- More realistic inhomogeneities
- More tested homogenization methods
  (In automatic mode only!)

Benchmarking methodology

- In general, data benchmarks composed by 100 homogeneous series without missing data. From them, for every test (100 realizations):
  - 10 series samples randomly chosen (*true* series)
  - Inhomogeneities applied to them (problem series)
  - Homogenization methods applied to homogenize the problem series (solutions, adjusted backwards)
  - Homogenized series are compared with the *true* series, computing RMSE, trend differences, and other metrics

Methodology

Tested homogenization programs (those that we could run in completely automatic mode):

- Climatol 3.0 (Guijarro), with constant and variable corrections
- ACMANT 3.0 (Domonkos), versions for temperature (different seasonalities) and precipitation
- MASH 3.03 (Szentimrey)
- RHTestV4 (Wang & Feng), absolute and relative, with or without quantile adjustment. (Average series were given as reference)
- USHCN_v52d (Menne & Williams)
  (We could not compile the current version)
- HOMER 2.6 (Mestre *et al.*), with different iteration strategies

Temp(a) networks & inhom.

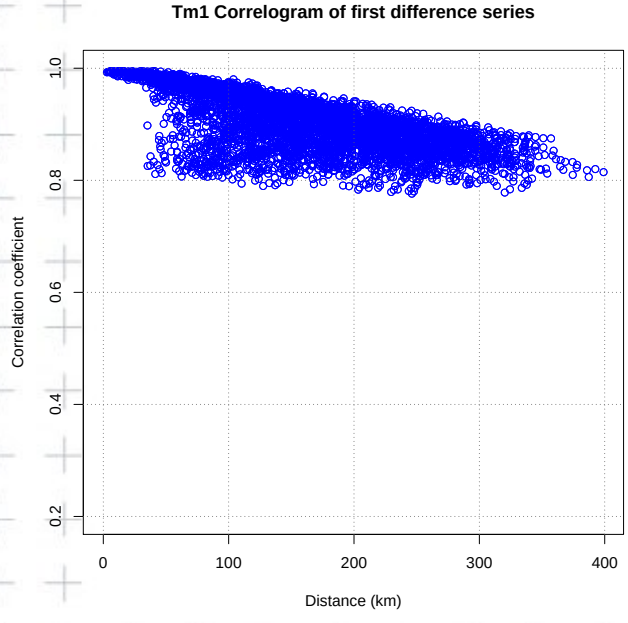
Generation of master networks Tm1, Tm2 and Tm3:

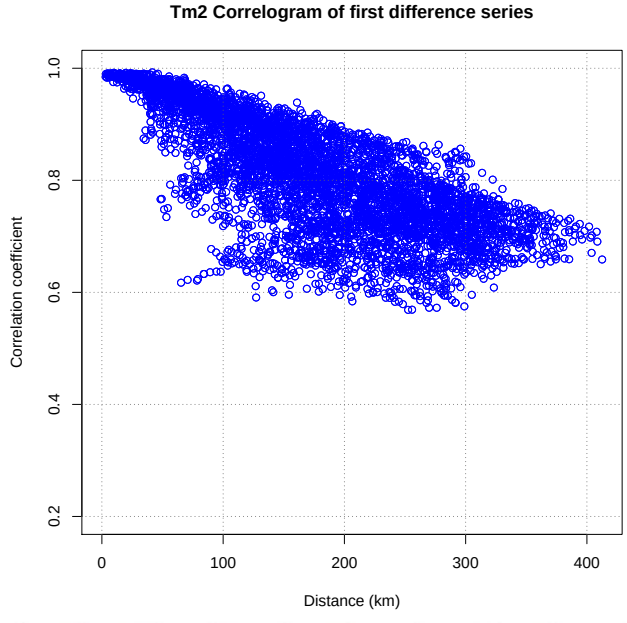
- 100 random points on a 4 x 3° lon-lat area
- Mean monthly homogenized temperatures from Valladolid (Duero basin, Spain) acting as seed series
- Closest point is assigned the same series plus white noise from $C \cdot \mathbf{N}(0, 1.5)$
- Coefficient $C = 0.18, 0.30, 0.65$ yield three master networks with decreasing correlation between stations, called Tm1, Tm2 and Tm3
- Series shifted to account for simulated elevation, 2°C/100yr trend added, and annual oscillation varied $\pm 20\%$

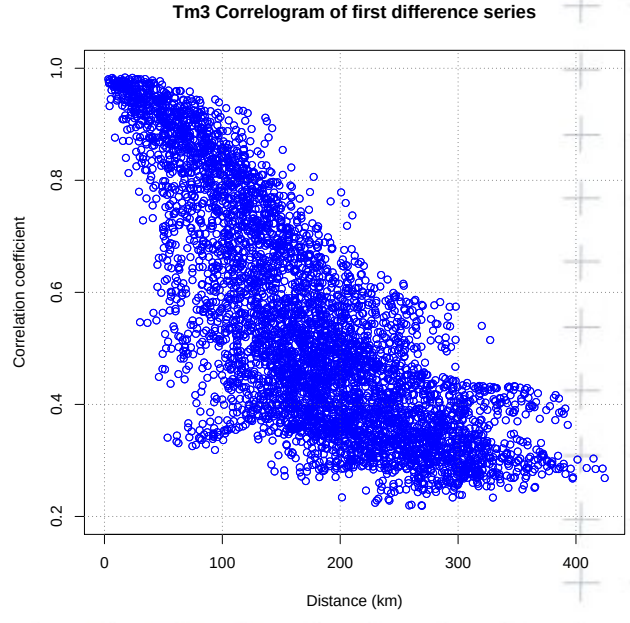
Inhomogeneities (mode 'rs'): Random number of shifts (5/100yr) with random size from $\mathbf{N}(0, 1)$ with sinusoidal seasonality of random amplitude from $\mathbf{N}(0, 0.7)$

Temp(a): Correlograms

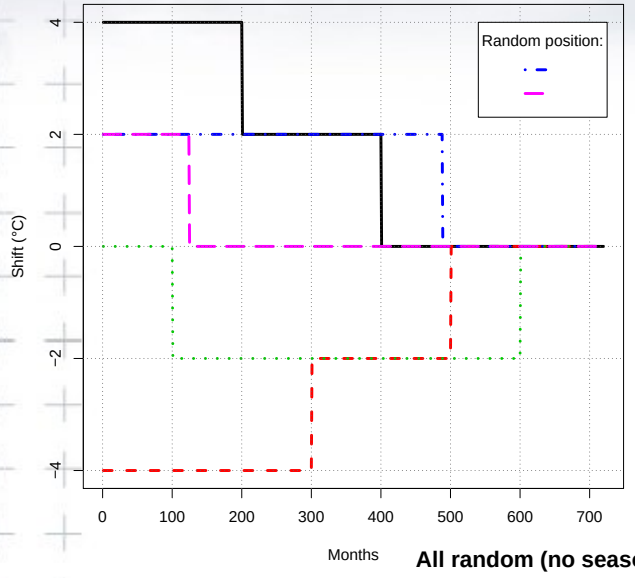
Correlograms of the first differences of the temperature networks Tm1, Tm2 and Tm3:



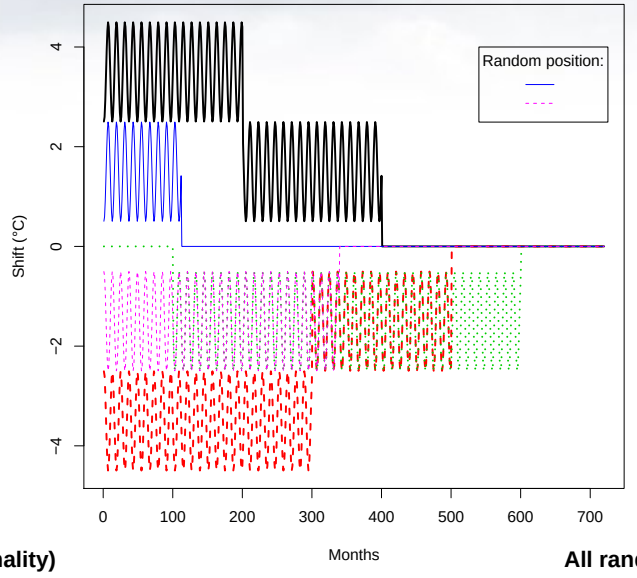


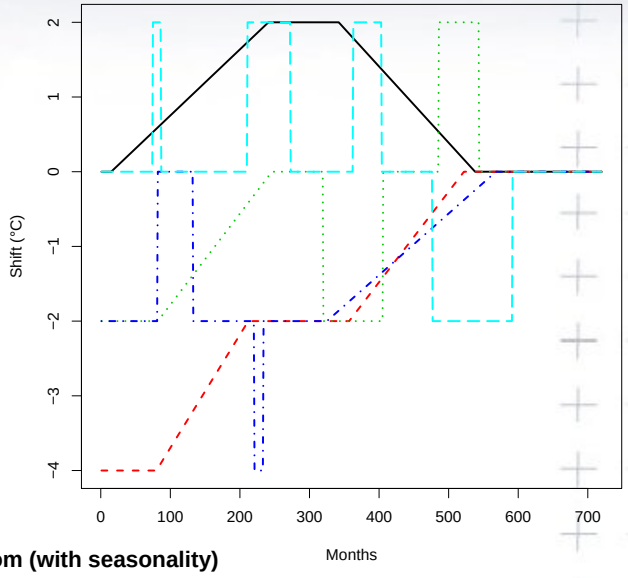


Temp(a): 5 experiments

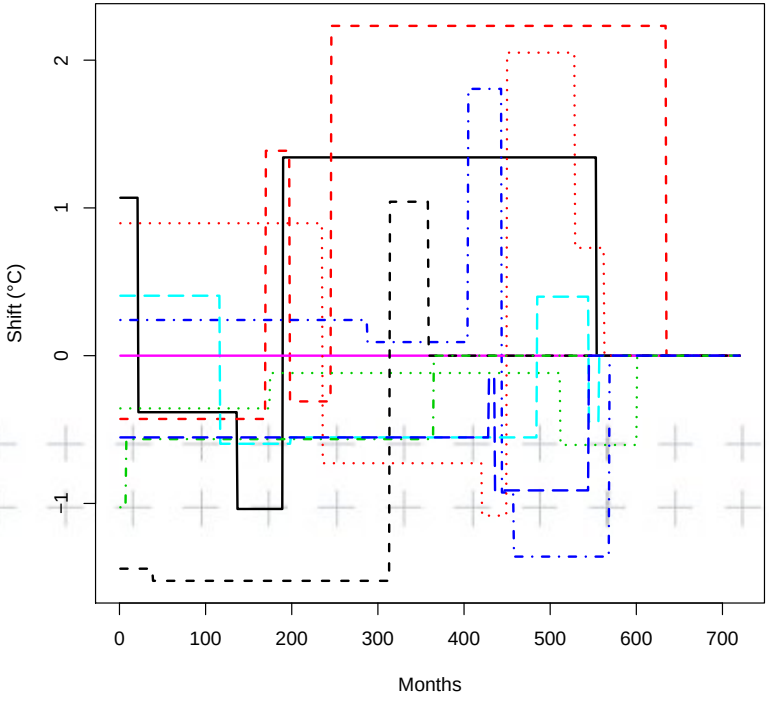


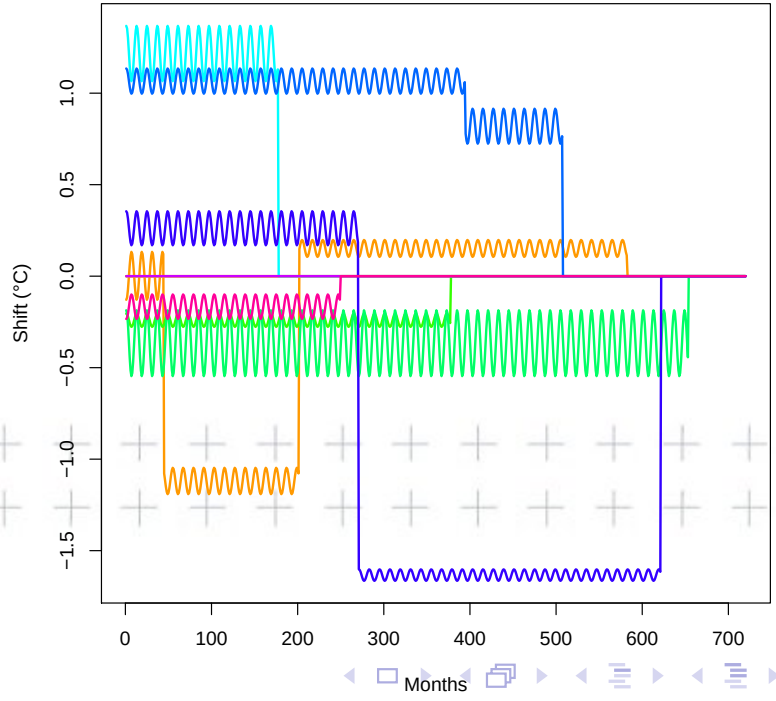
All random (no seasonality)



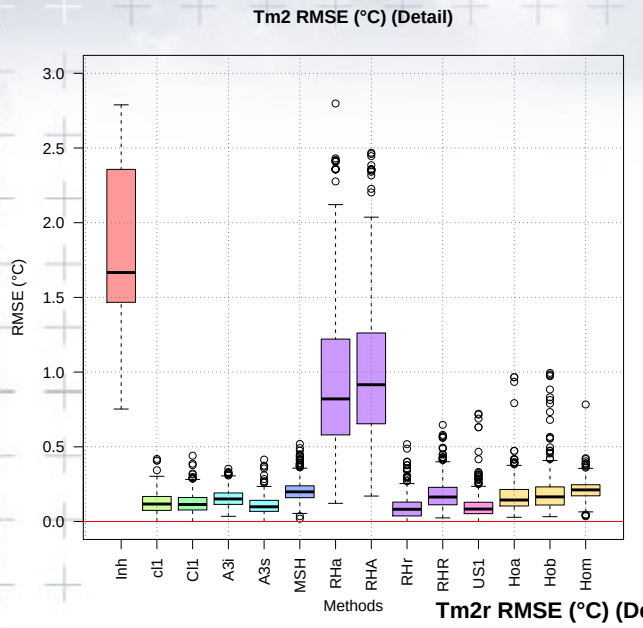


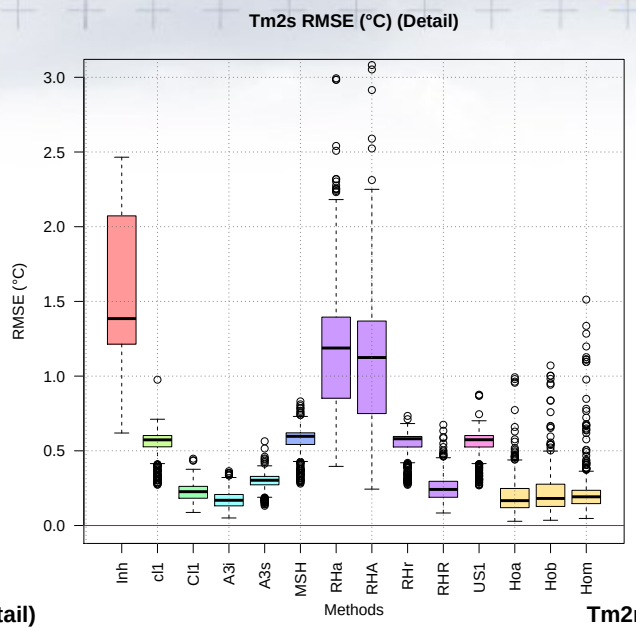
All random (with seasonality)

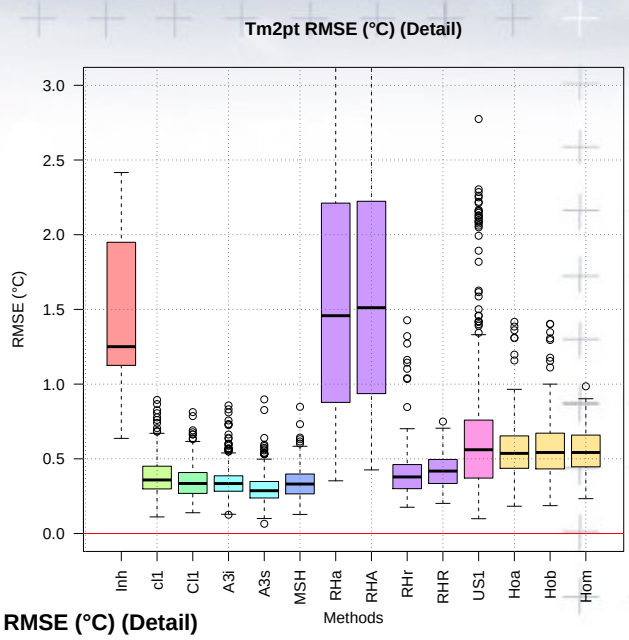


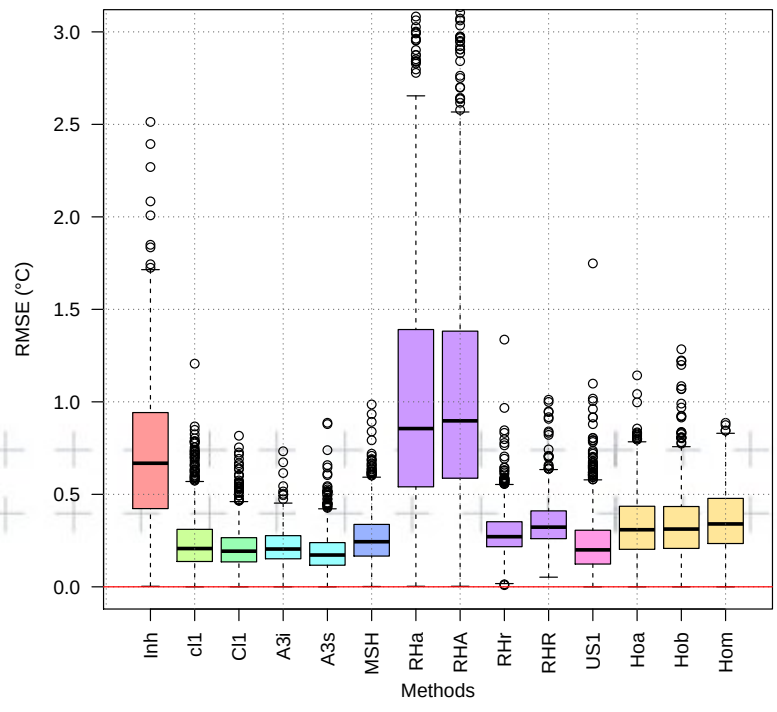


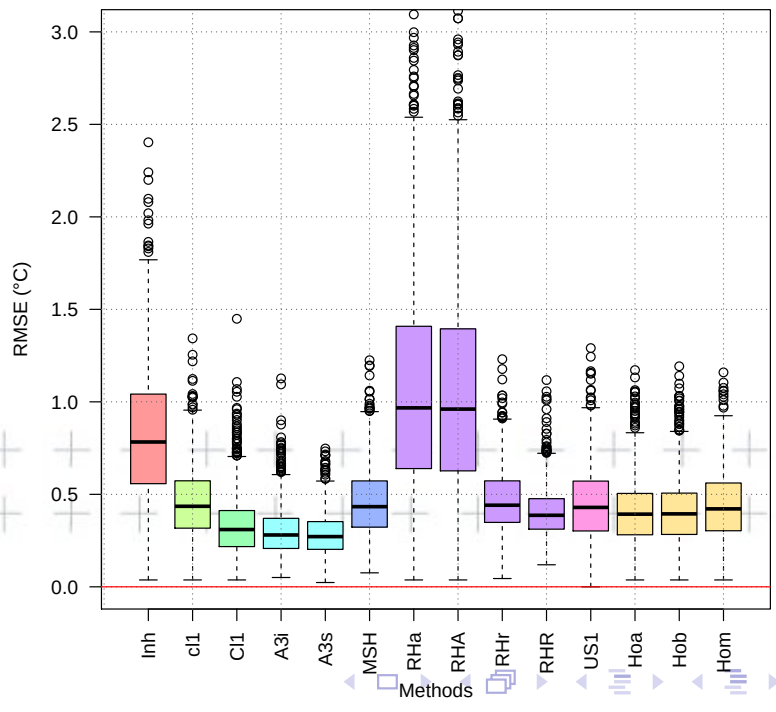
Temp(a): RMSE results Tm2



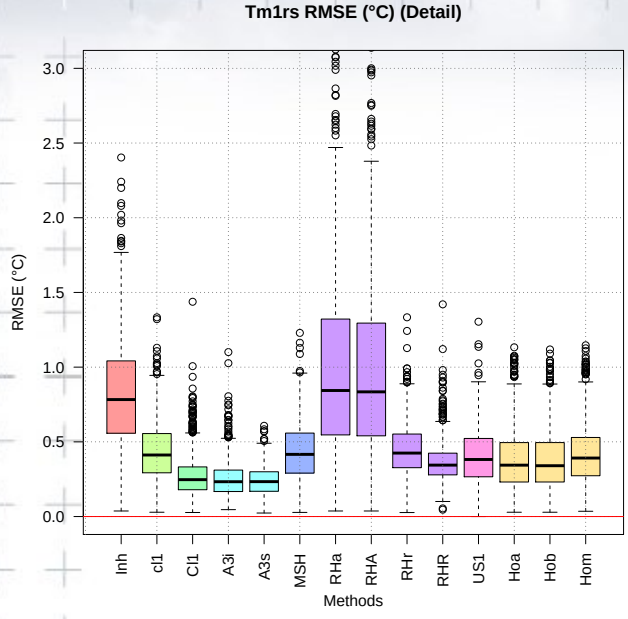


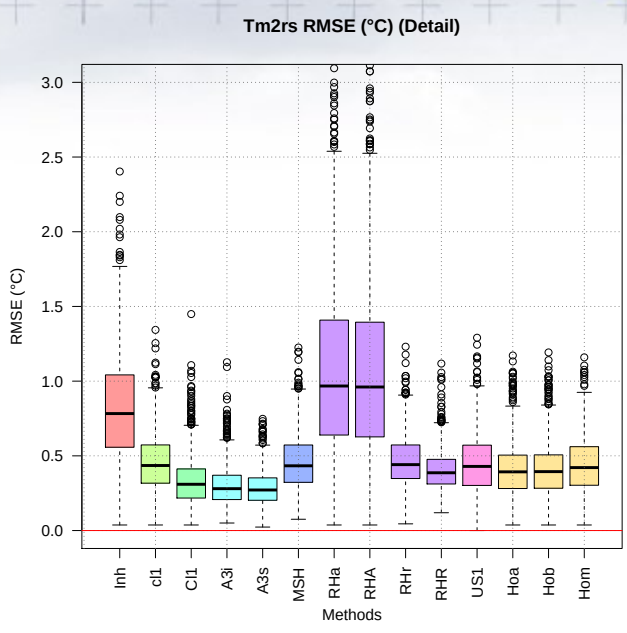


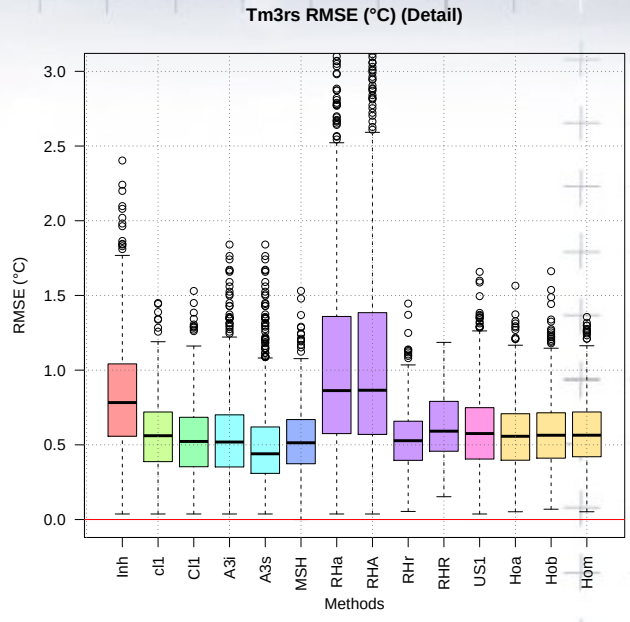


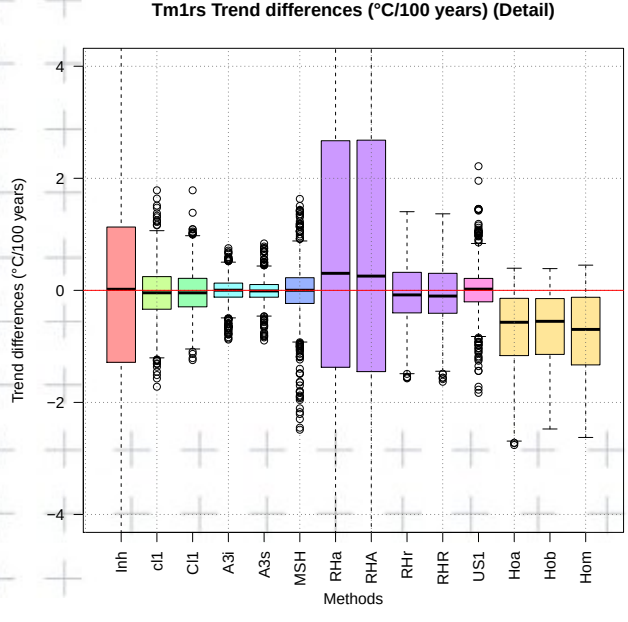


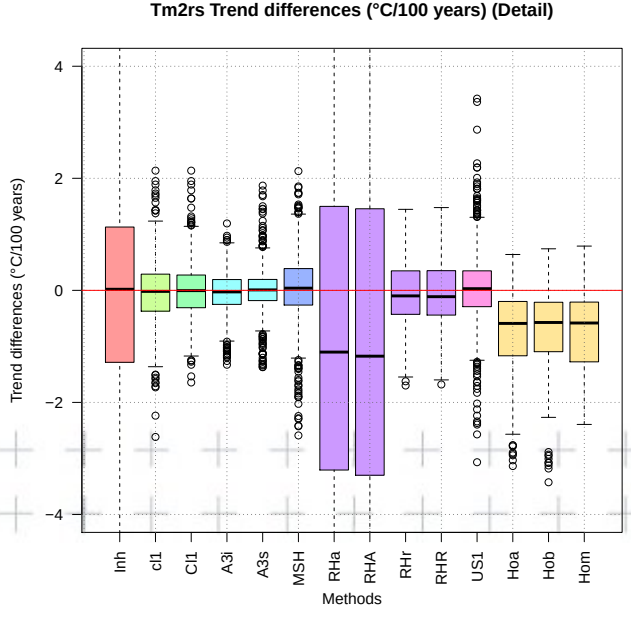
Temp(a): RMSE and trends 'rs'

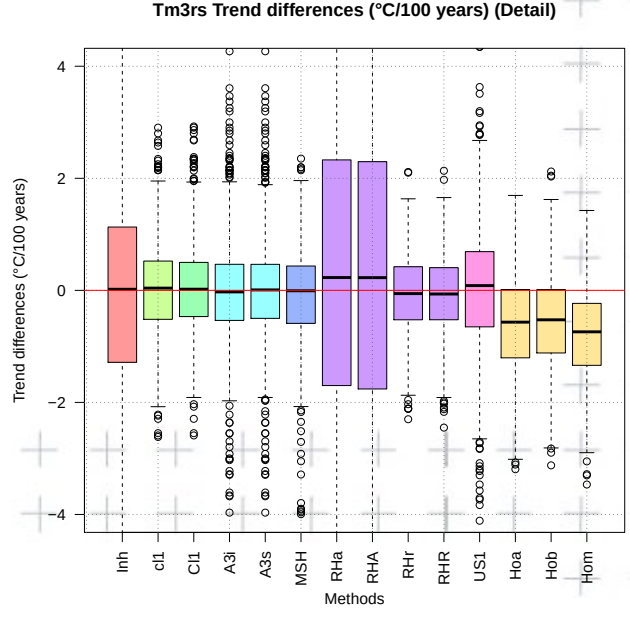










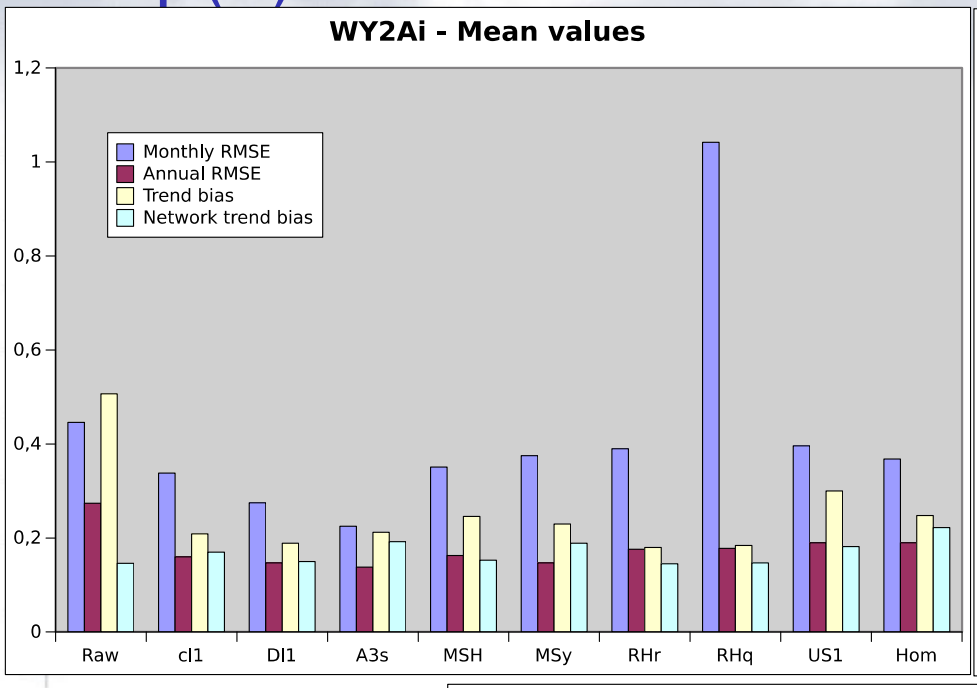


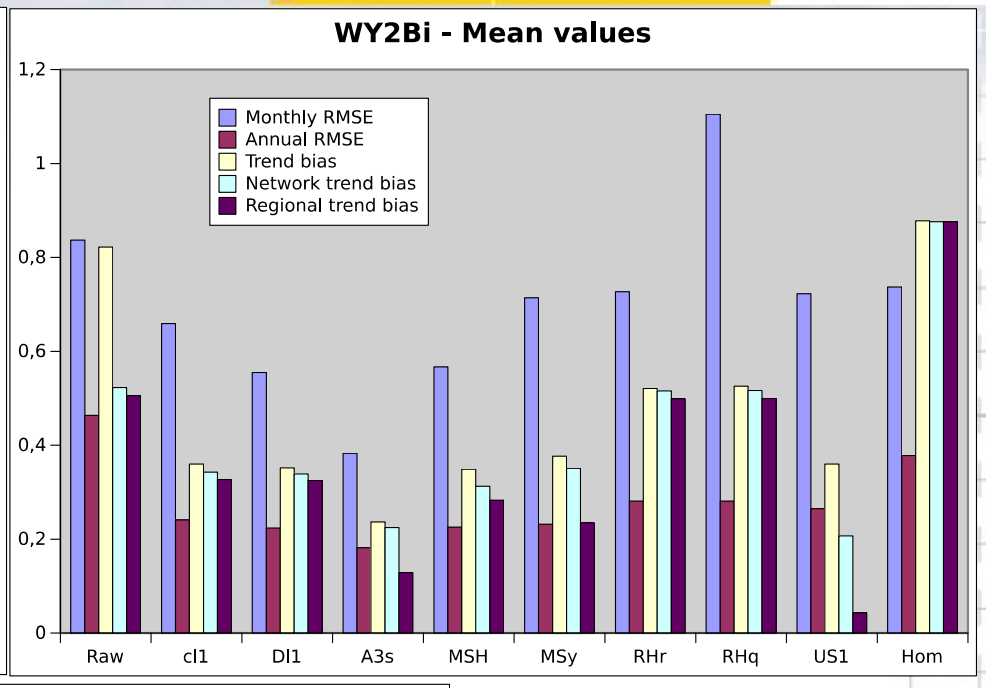
Temp(b) networks

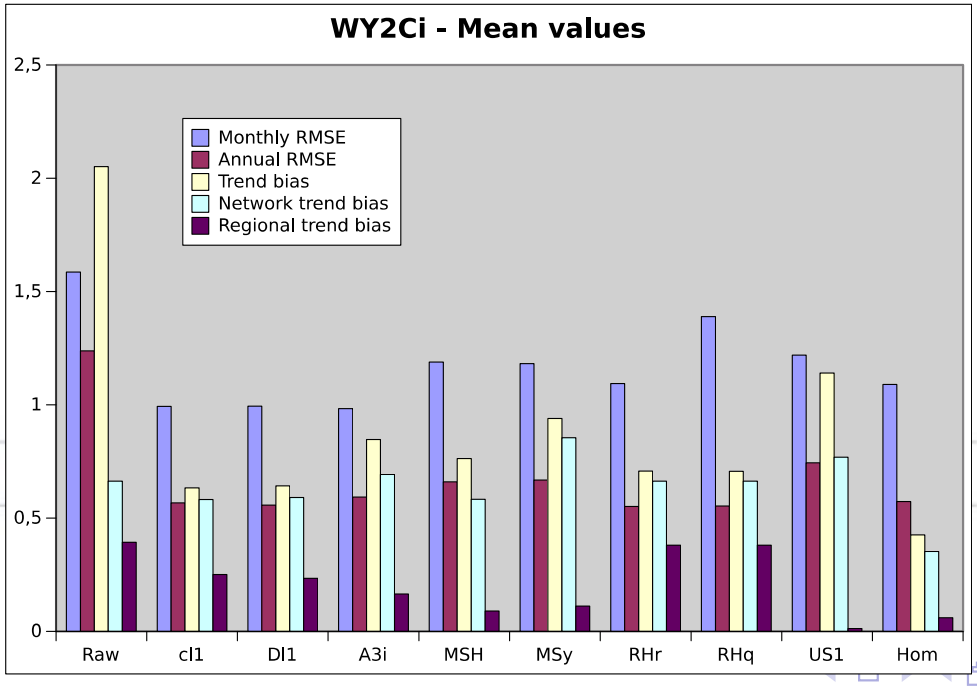
Networks WY2-Ai, WY2-Bi, and WY2-Ci derived from Rachel Killick's WY world 2, without missing data. 100 tests based on 10 series samples of 60 years (720 terms). Inhomogeneities added with a limited random walk algorithm.

	WY2Ai	WY2Bi	WY2Ci
Mean r	0.85	0.85	0.65
Outliers/100yr	2	5	10
Shifts/100yr	4	4	6
Trends/100yr	1	1	1
Platforms/100yr	2	5	15
Inhomogeneities size (°C)	N(0,0.5)	N(0,0.8)	N(0,1.5)
Bias seasonality	Mid. Lat.	Mid. Lat.	Monsoonal
Climatic trend sudden jump (°C)	0	0.8	0
Added network-mean bias	No	Yes	Yes

Temp(b) results







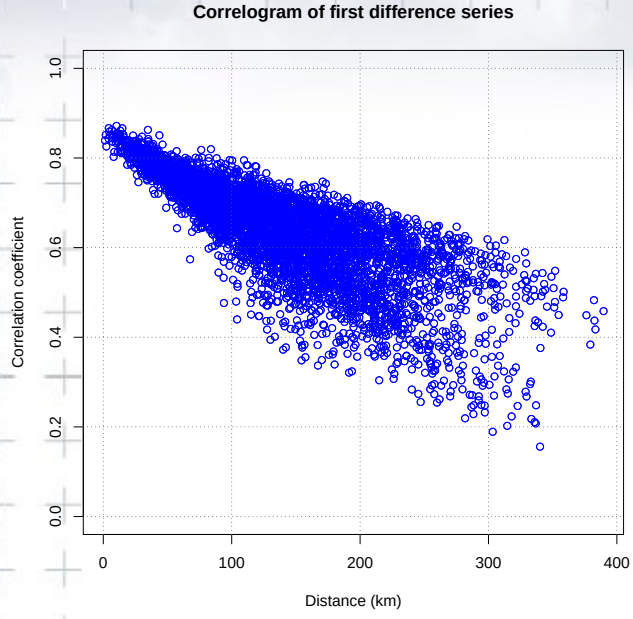
Precipitations

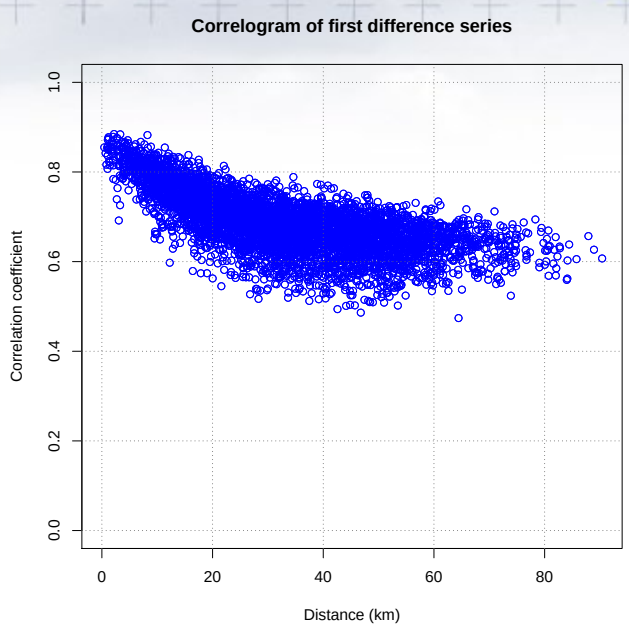
Three monthly precipitation networks were built simulating three different climates: Atlantic temperate (PEir), Mediterranean (PMca) and Monsoonal (PInd).

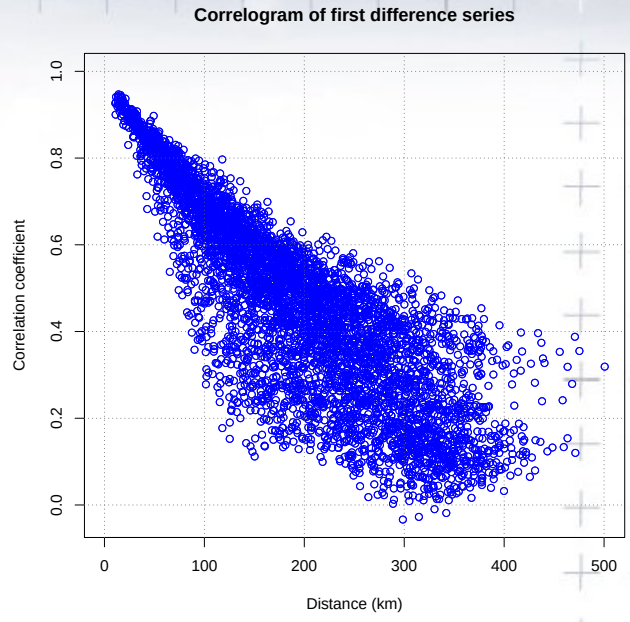
Real series from Ireland, Majorca and SW India (gridded) were respectively used to derive variograms, gamma coefficients and frequency of zeroes, which were used to compute their synthetic series by means of the R package gstat, preserving the spatial correlation structure.

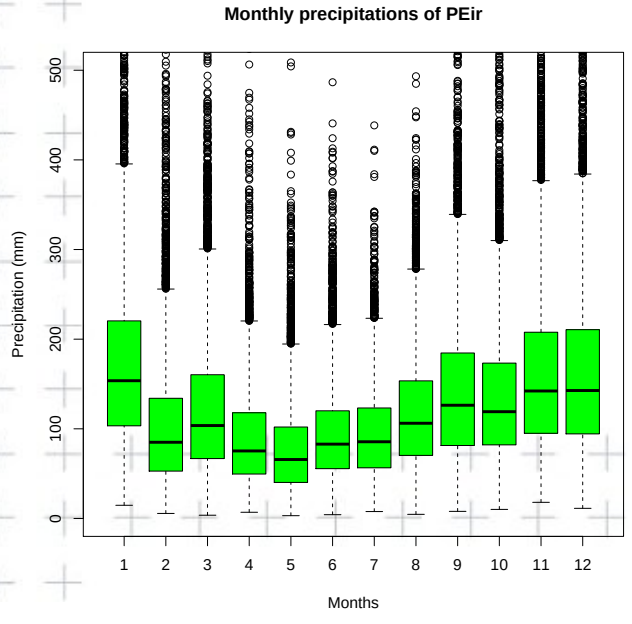
A random number of shifts (5/100yr) were introduced as **factors** drawn from $\mathbf{N}(0, 0.2)$ (in mode 'r': no seasonal perturbation were applied to these factors)

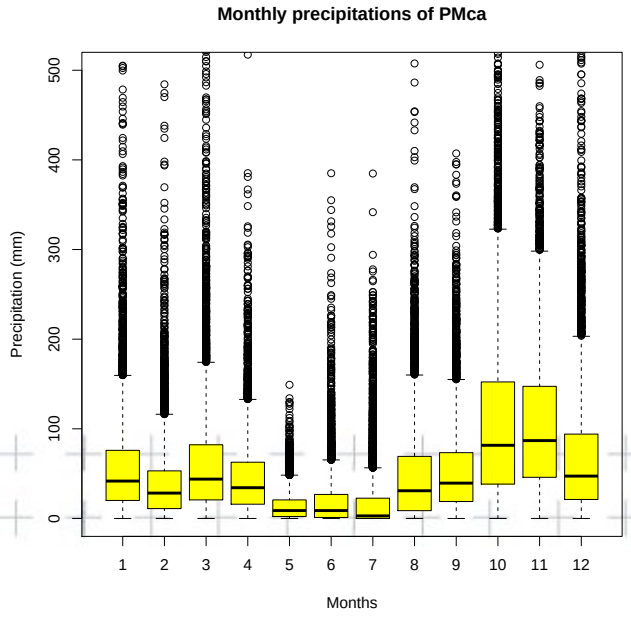
Precip. characteristics

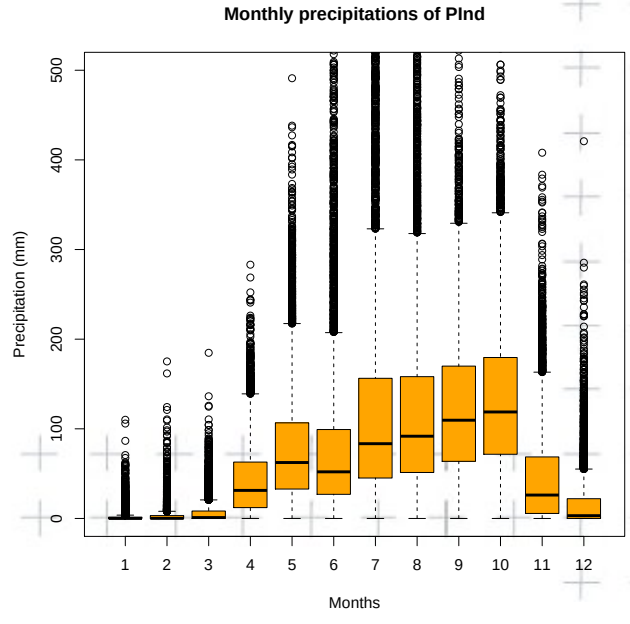




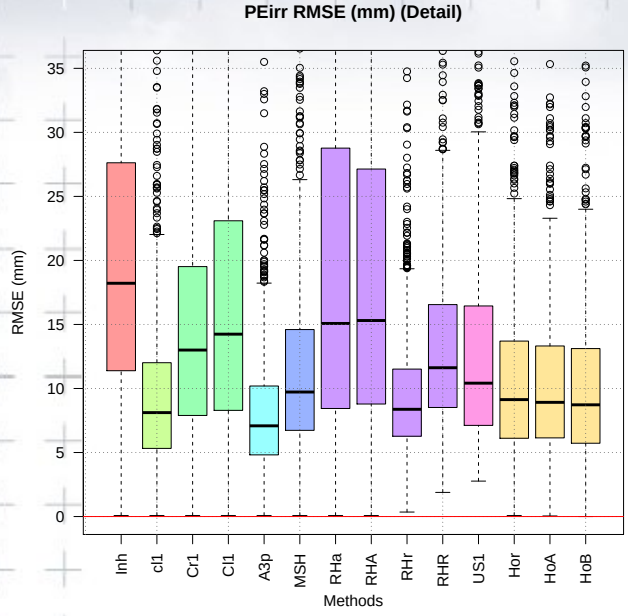


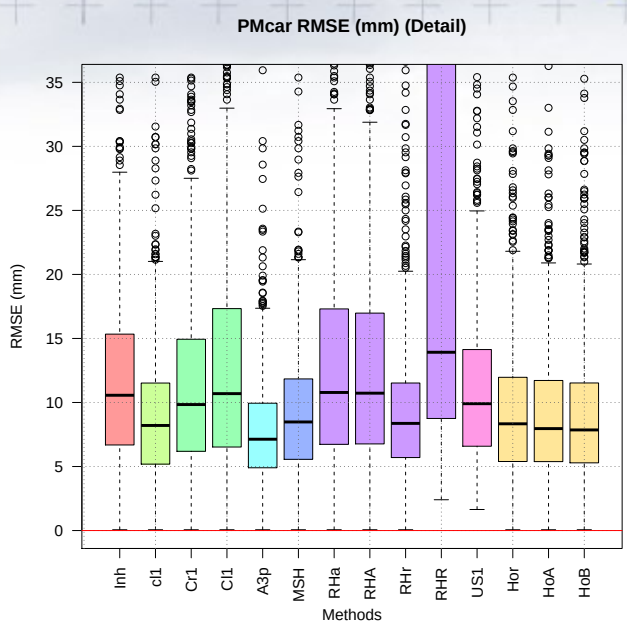


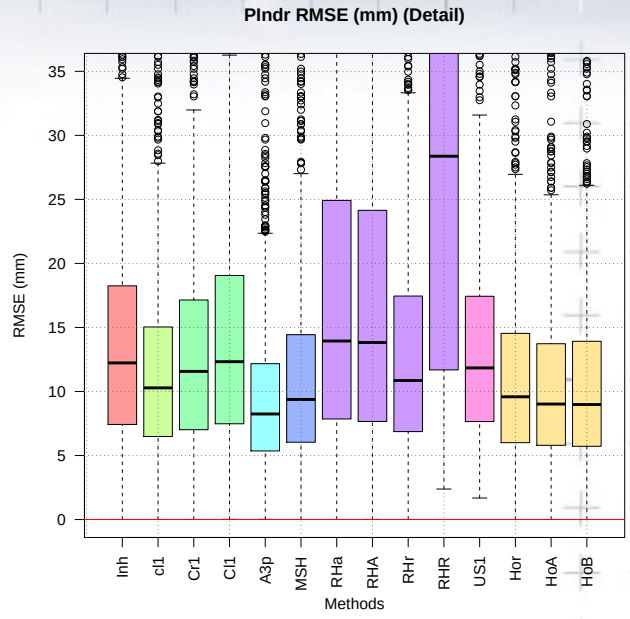


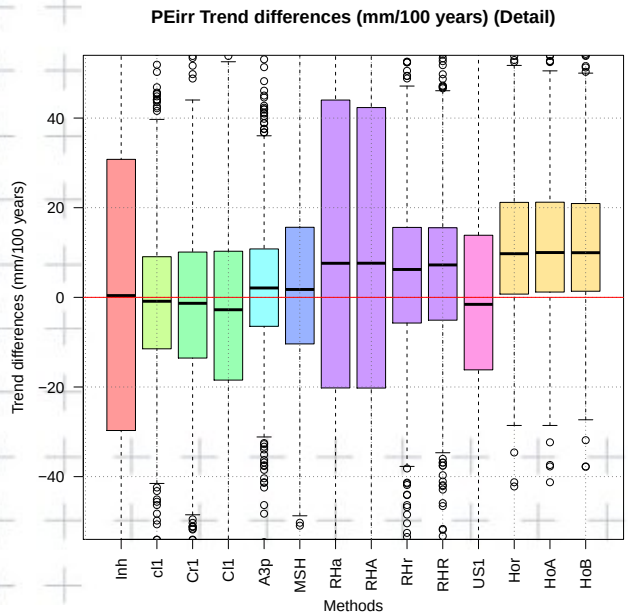


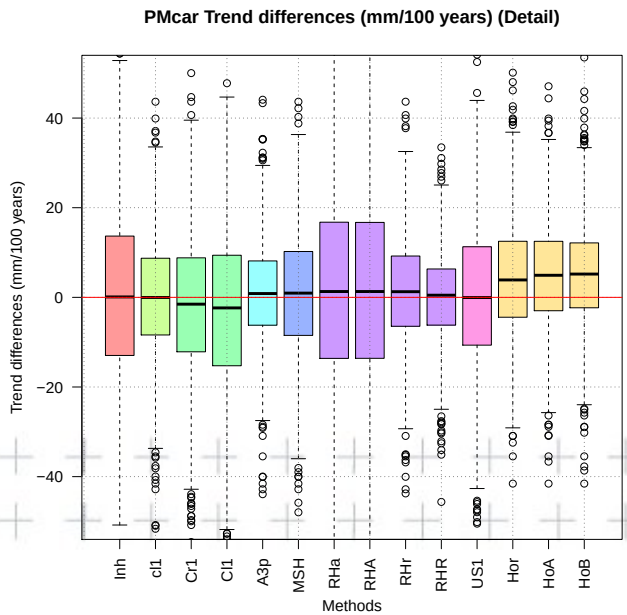
Precip: RMSE and trends ('r')

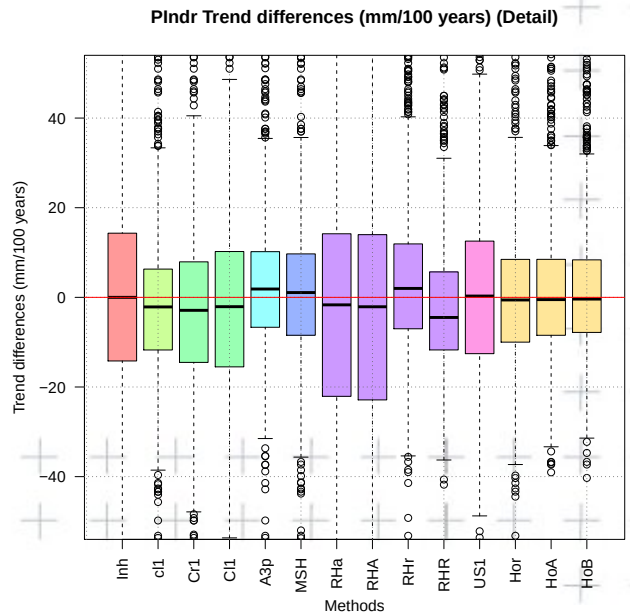






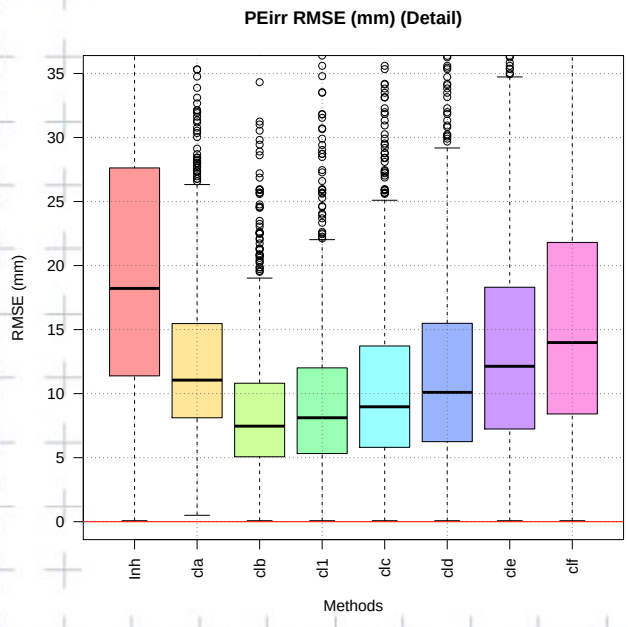


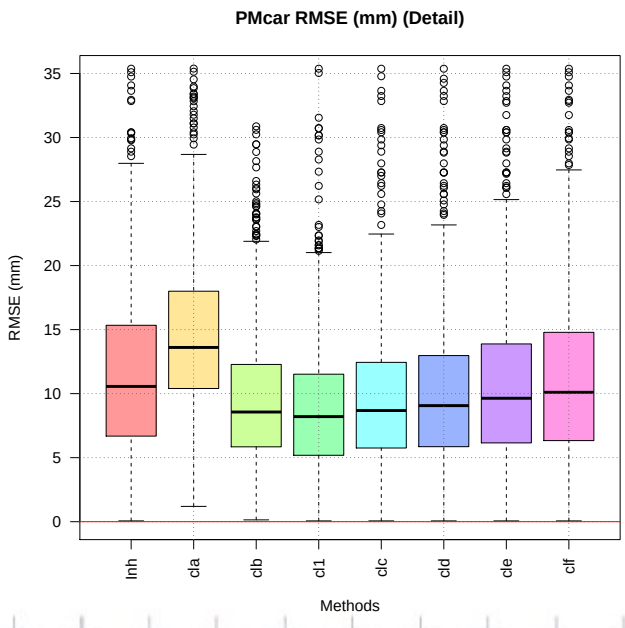


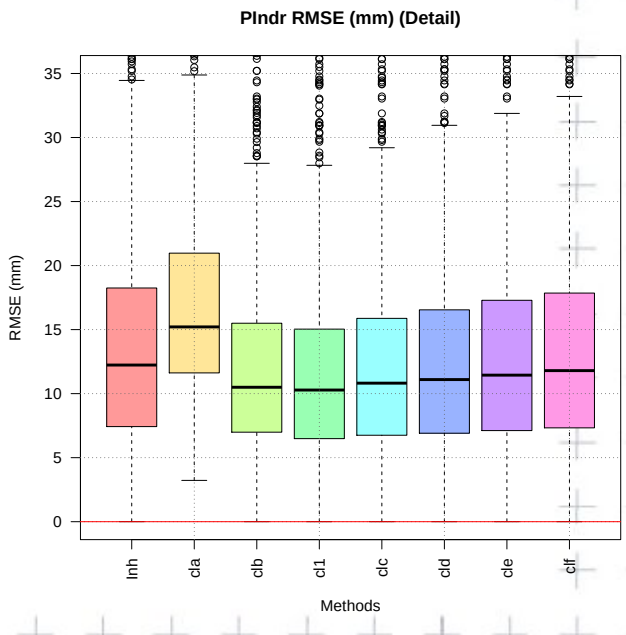


Precip: RMSE vs SNHT thresholds

RMSE obtained by ClimatoI (with rate normalization) on the precipitation tests with thresholds of SNHT = 5, 10, 15, 20, 25, 35 and 50 :







Conclusions

- The performance of the methods can vary depending on the characteristics of both the network and the inhomogeneities
- Hence the importance of showing results from different networks that can be representative of different real climates
- Unrealistically designed experiments also help in detecting the strengths and weaknesses of the methods
- Precipitation appears as probably being the most difficult variable to homogenize (many zeroes and a very biased PDF)
- The graphics displaying the results of the tests, as well as other characteristics of the software packages shown in http://www.climatol.eu/tt-hom/index.html, will facilitate the user to choose the method that better suits his needs

Future work

Future work includes:

- Test missing data tolerance of new packages
- Try longer series with missing data mimicking those in the HOME benchmark
- Put all results and scripts in a web page to allow reproducibility

Last remark: We are open to apply other settings to the methods if required by their developers

Acknowledgements

Project MULTITEST (Multiple verification of automatic softwares homogenizing monthly temperature and precipitation series, code CGL2014-52901-P) is funded by the Spanish Ministry of Economy and Competitiveness.

Thanks to Rachel Killick for providing the WY2 temperature series and to Met Éireann for doing the same with the Irish monthly precipitation series, that served as model to synthesize the network of Atlantic Temperate precipitations. Monthly precipitations from SW India, gridded at 0.5° resolution, were obtained from the Global Precipitation Climate Center (GPCC).

Many thanks to all the colleagues that made valuable suggestions to this project in general (Victor Venema) and on the most appropriate settings to run MASH (Tamás Szentimrey) and HOMER (Gregor Vertacnik, John Coll and Stefanie Gubler).