Comparison of correction methods for inhomogeneities in daily time series of air temperature

Petr Štěpánek, Pavel Zahradníček

Czech Hydrometeorological Institute, Regional Office Brno, Czech Republic

EGU 2011, Vienna, 5th April

E-mail: petr.stepanek@chmi.cz

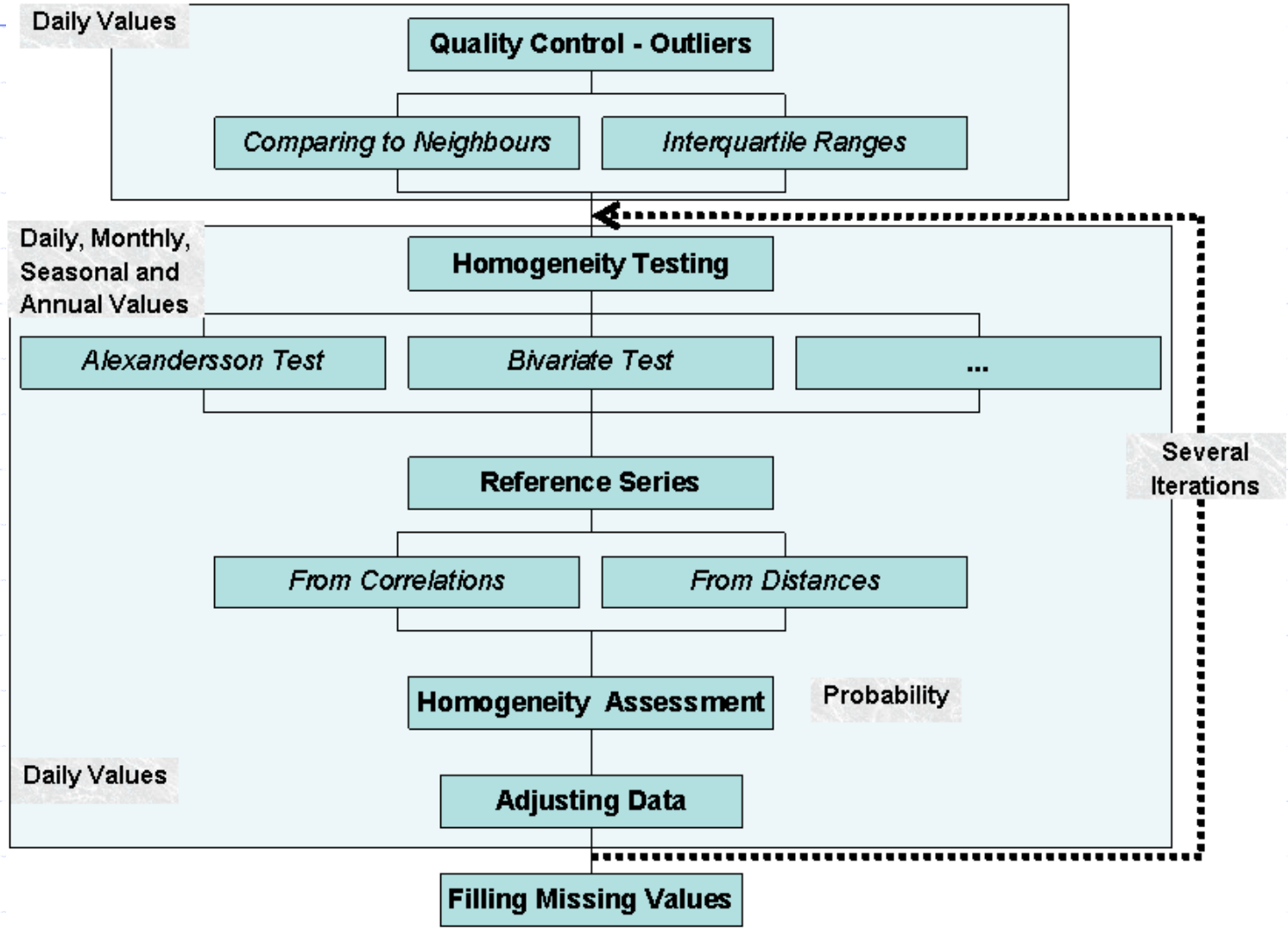
Outline

- Introduction
- Methods for correction of inhomogeneities in daily data
- Comparison of the selected methods
- Conclusions

Background

- The work contributes to the effort carried out by the COST action ESO601 (HOME) activity, WG4 – daily data
- Comparison of the correction methods in „blind testing“ on the surrogate Benchmark dataset (Victor Venema, air temperature and precipitation for 15 networks: 5 surrogates of each of the 3 original networks, each containing 9 stations, in the period 1908-2007, having statistical properties of measured data - based upon three networks from the Czech Republic in the period 1/1/1961-31/12/2007), Just released, please apply your method if you have any
- In this work, the breaks are simulated by combining base and neighbouring stations

Processing before any data analysis, General scheme



Software AnClim, ProClimDB

Daily data correction methods

- „Delta“ methods
- Variable correction methods – one element
- Variable correction methods – several elements

Daily data correction methods

- Interpolation of monthly factors
  - MASH
  - Vincent *et al* (2002) - cubic spline interpolation
- Nearest neighbour resampling models, by Brandsma and Können (2006)
- Higher Order Moments (HOM), by Della Marta and Wanner (2006)
- Two phase non-linear regression (O. Mestre)
- Distribution Adjusting by Percentiles (DAP), by Stepanek, 2009
- Using weather types classifications (HOWCLASS), by I. Garcia-Borés, E. Aguilar,
- Quantile matching algorithm (QM) by Xiaolan L. Wang (2010)

Daily data correction methods

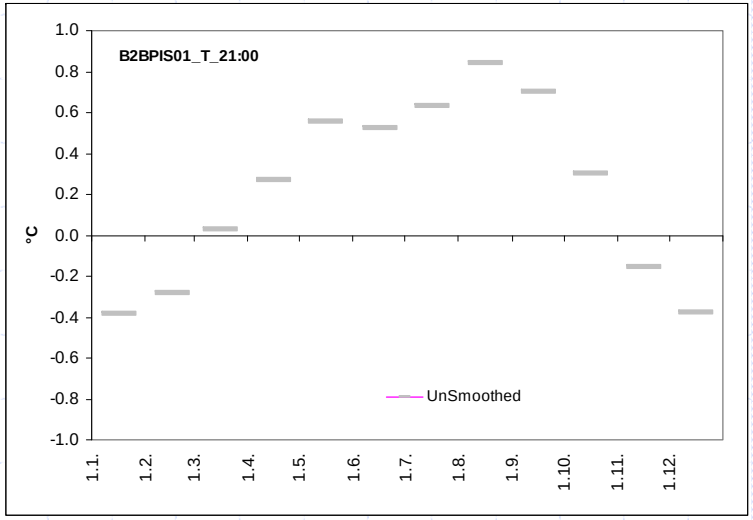
Modifications, Further activities,

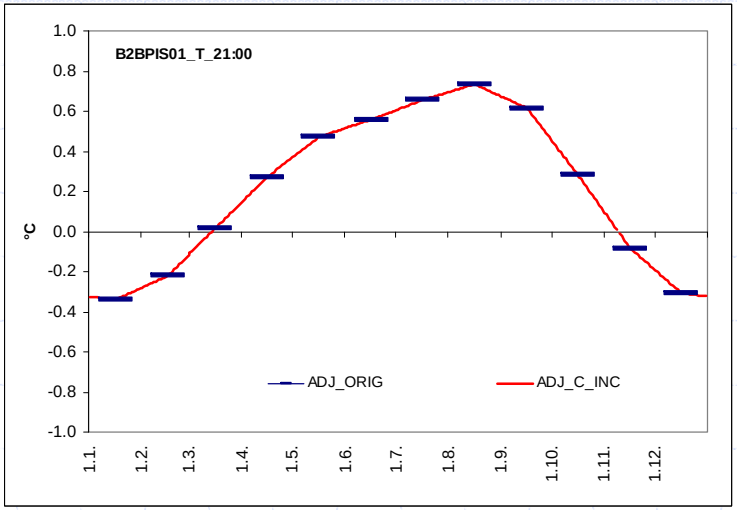
- Modification of HOM method (HOMAD) by A. Toreti et al.
- Applying bootstrapping technique for improving HOM and SPLIDHOM correction methods at ZAMG
- Predrag Petrovic, ReDistribution method for wind measurements . He also tries to process only a selection of data in order to detect inhomogeneities, e.g. selection of wind data by direction intervals or by wind speed intervals

The „delta“ method: Adjusting from monthly versus daily adjustments

Adjusting from **monthly** data

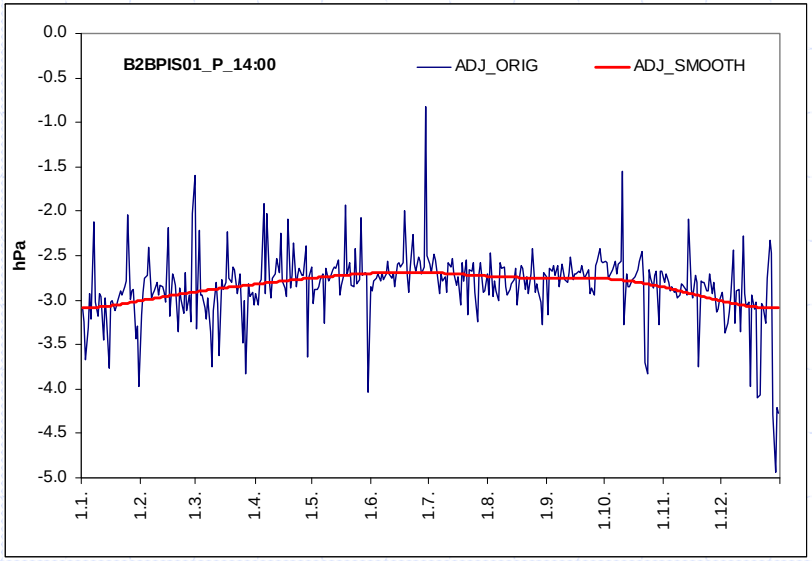
- monthly adjustments smoothed with Gaussian low pass filter (weights approximately 1:2:1)
- smoothed monthly adjustments are then evenly distributed among individual days





Adjusting straight from **daily** data

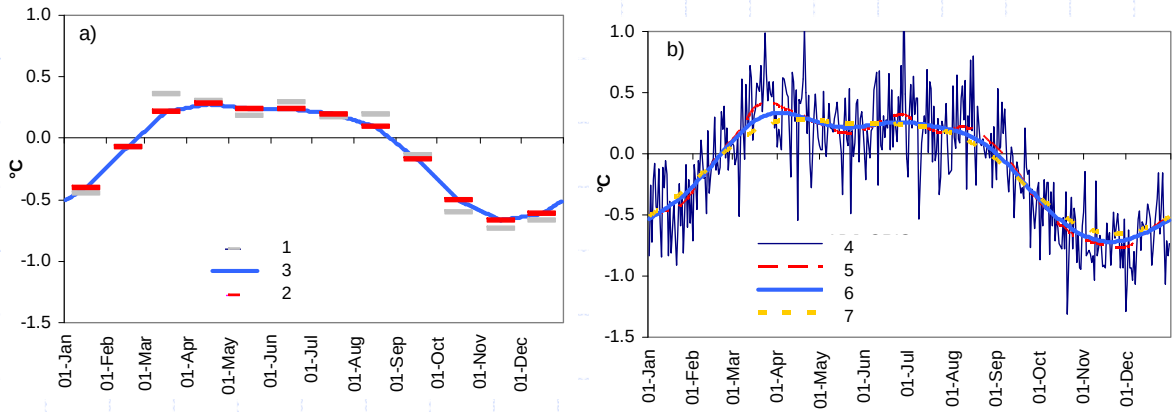
- Adjustment estimated for each individual day (series of 1st Jan, 2nd Jan etc.)
- Daily adjustments smoothed with Gaussian low pass filter for 90 days (annual cycle 3 times to solve margin values)



Adjustments (Delta method)

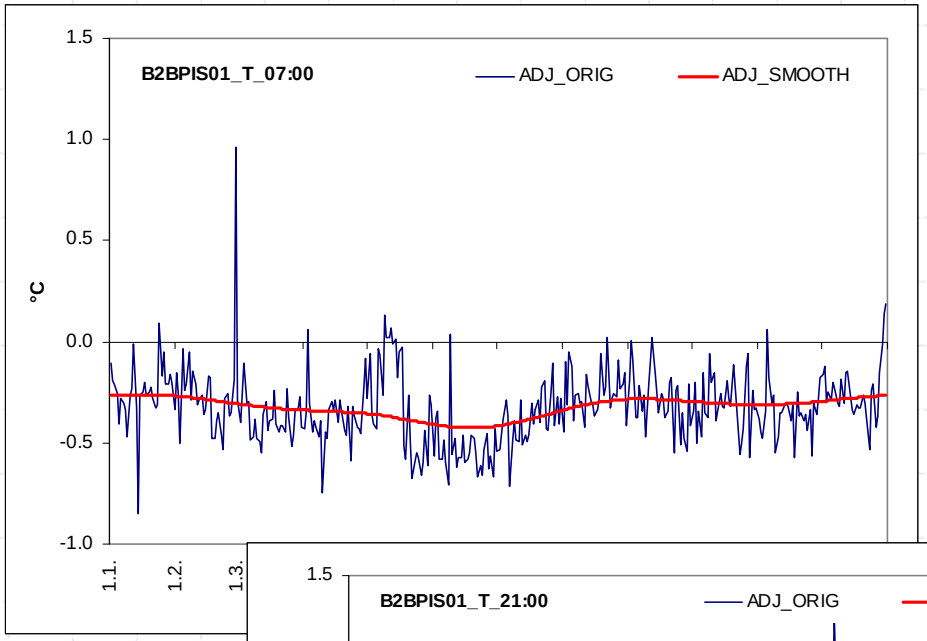
- **The same final adjustments may be obtained from either monthly averages or through direct use of daily data**

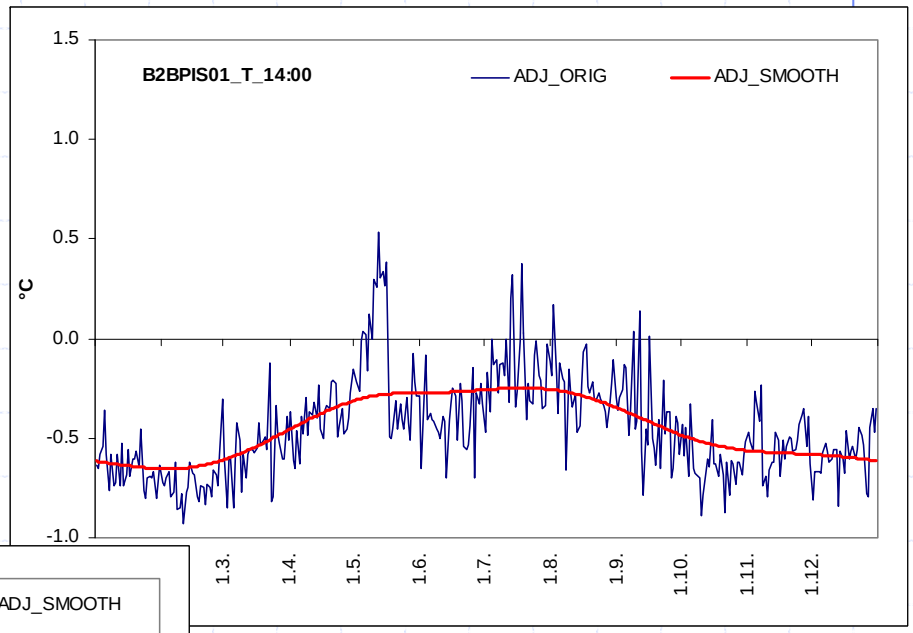
(for the daily-values-based approach, it seems reasonable to smooth with a low-pass filter for 60 days. The same results may be derived using a low-pass filter for two months (weights approximately 1:2:1) and subsequently distributing the smoothed monthly adjustments into daily values)

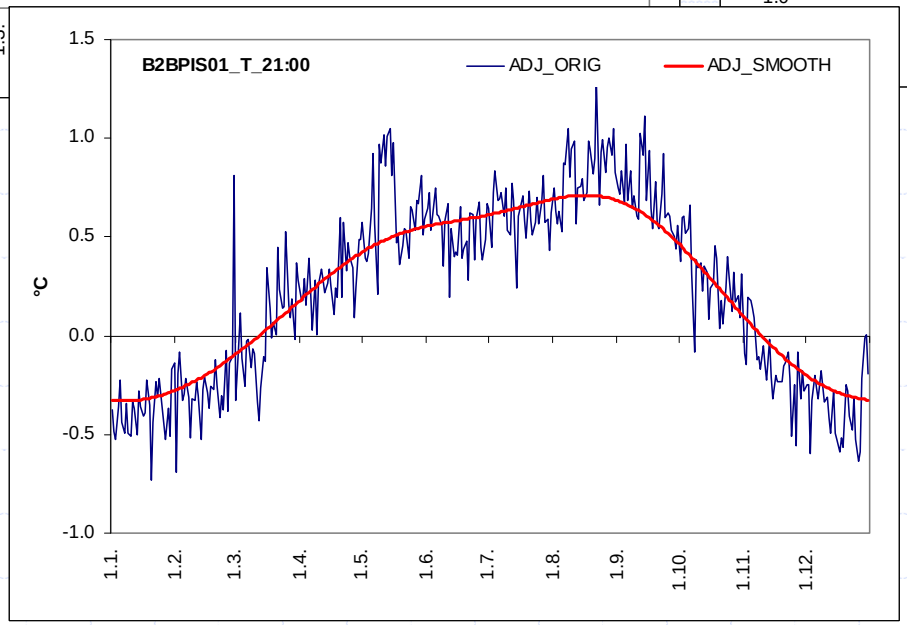


(1 – raw adjustments, 2 – smoothed adjustments, 3 – smoothed adjustments distributed into individual days), b) daily-based approach (4 – individual calendar day adjustments, 5 – daily adjustments smoothed by low-pass filter for 30 days, 6 – for 60 days, 7 – for 90 days)

Remark: daily data not sufficient (need to work with subdaily scale)







Air temperature.: adjustments applied to individual observation hours

be careful when applying straight to daily AVG, it is a mixture of sub-scale corrections

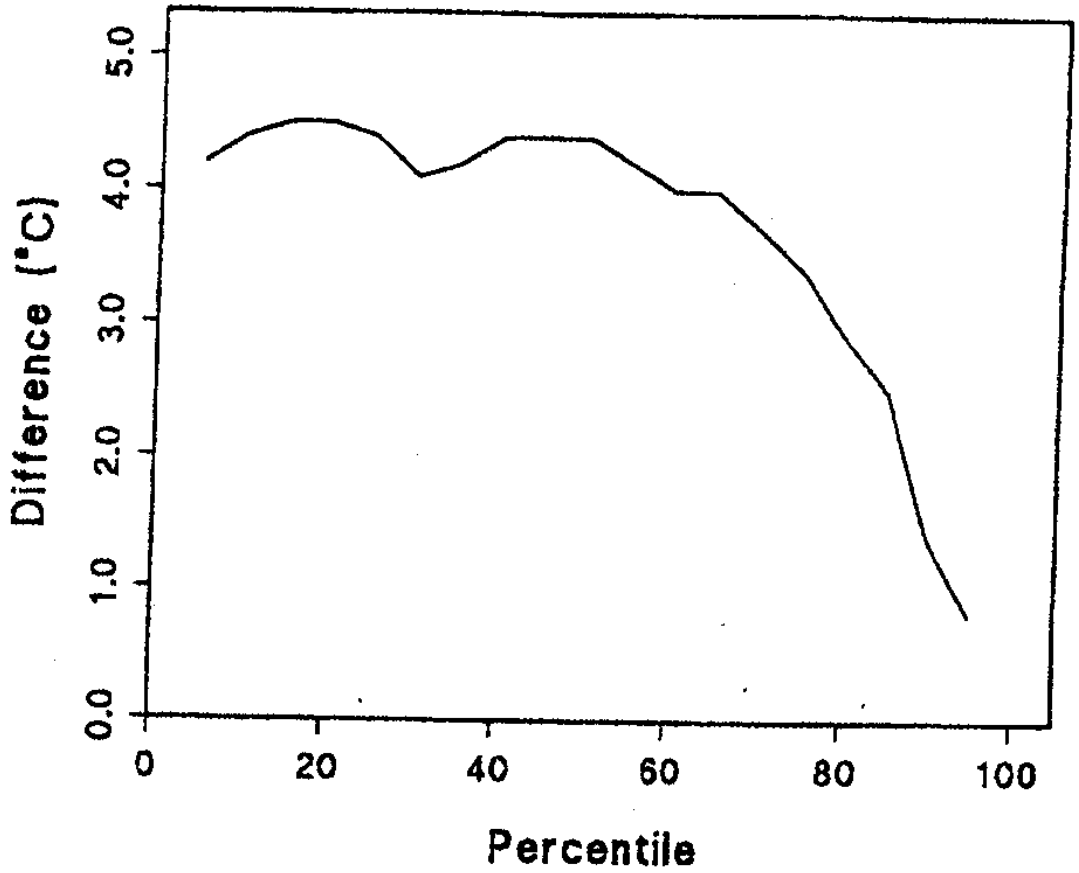
Variable correction

- **$f(C(d)|R)$, function build with the reference dataset R, d – daily data**
- **cdf, and thus the pdf of the adjusted candidate series C*(d) is exactly the same as the cdf or pdf of the original candidate series C(d)**

Variable correction

B. C. TREWIN AND A. C. F. TREVITT

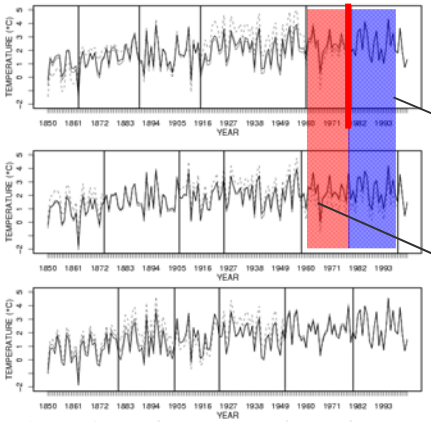
1996

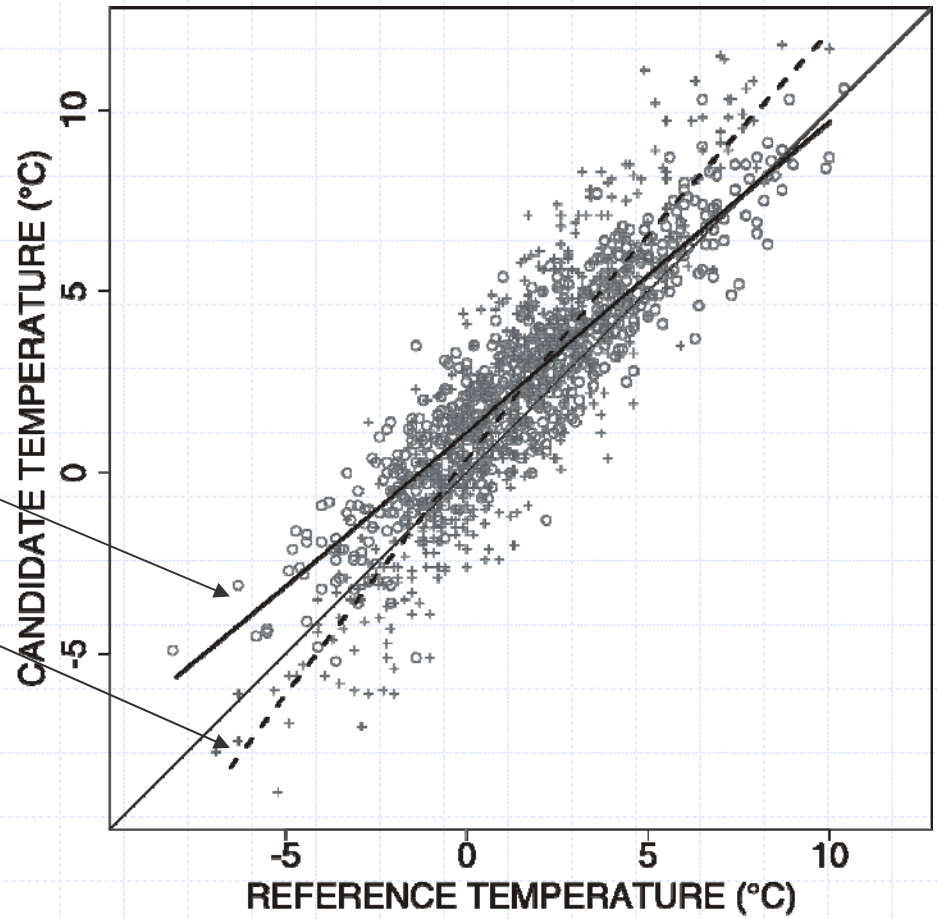


The HOM method concept: Fitting a model

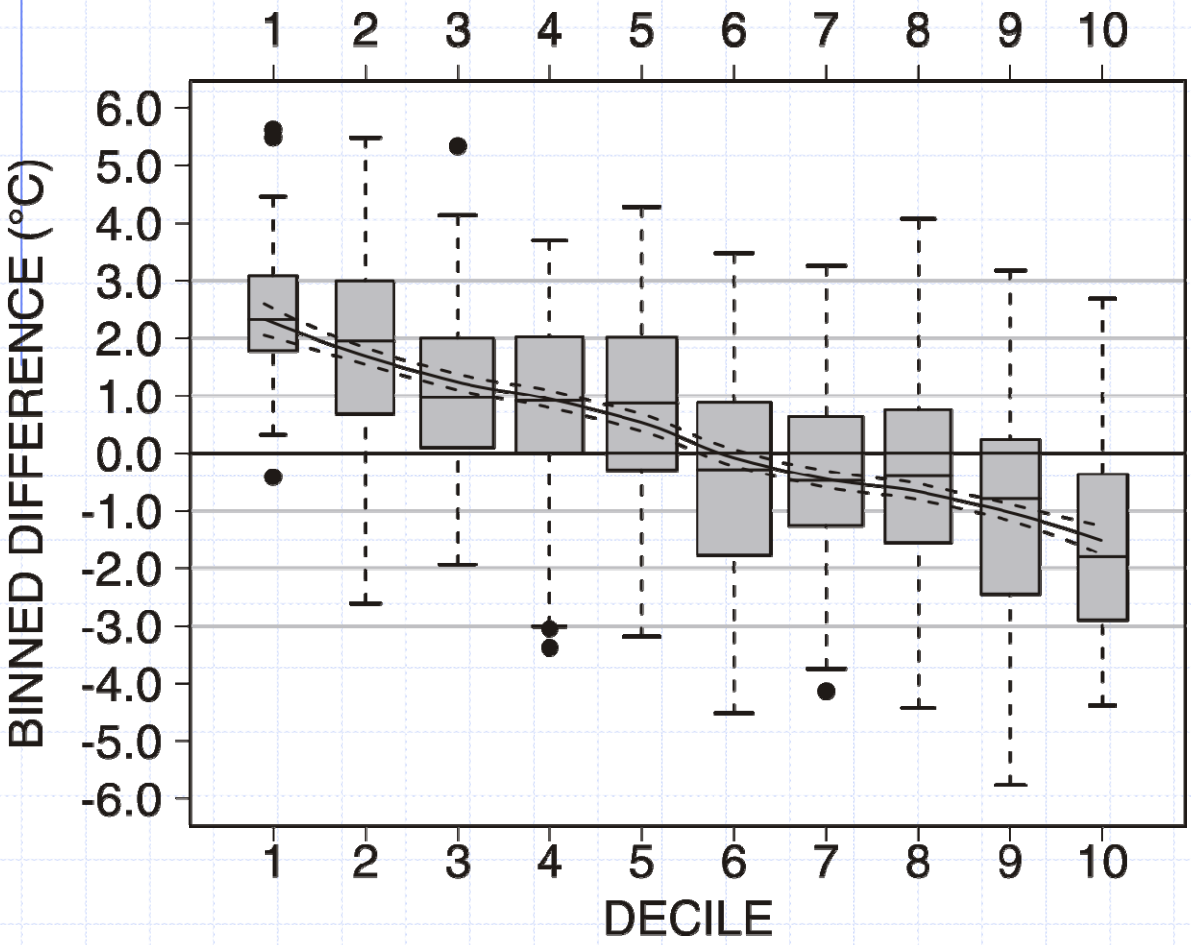
Locally weighted regression (LOESS) (Cleveland & Devlin,1998)

HSP2 *HSP1*





The HOM method

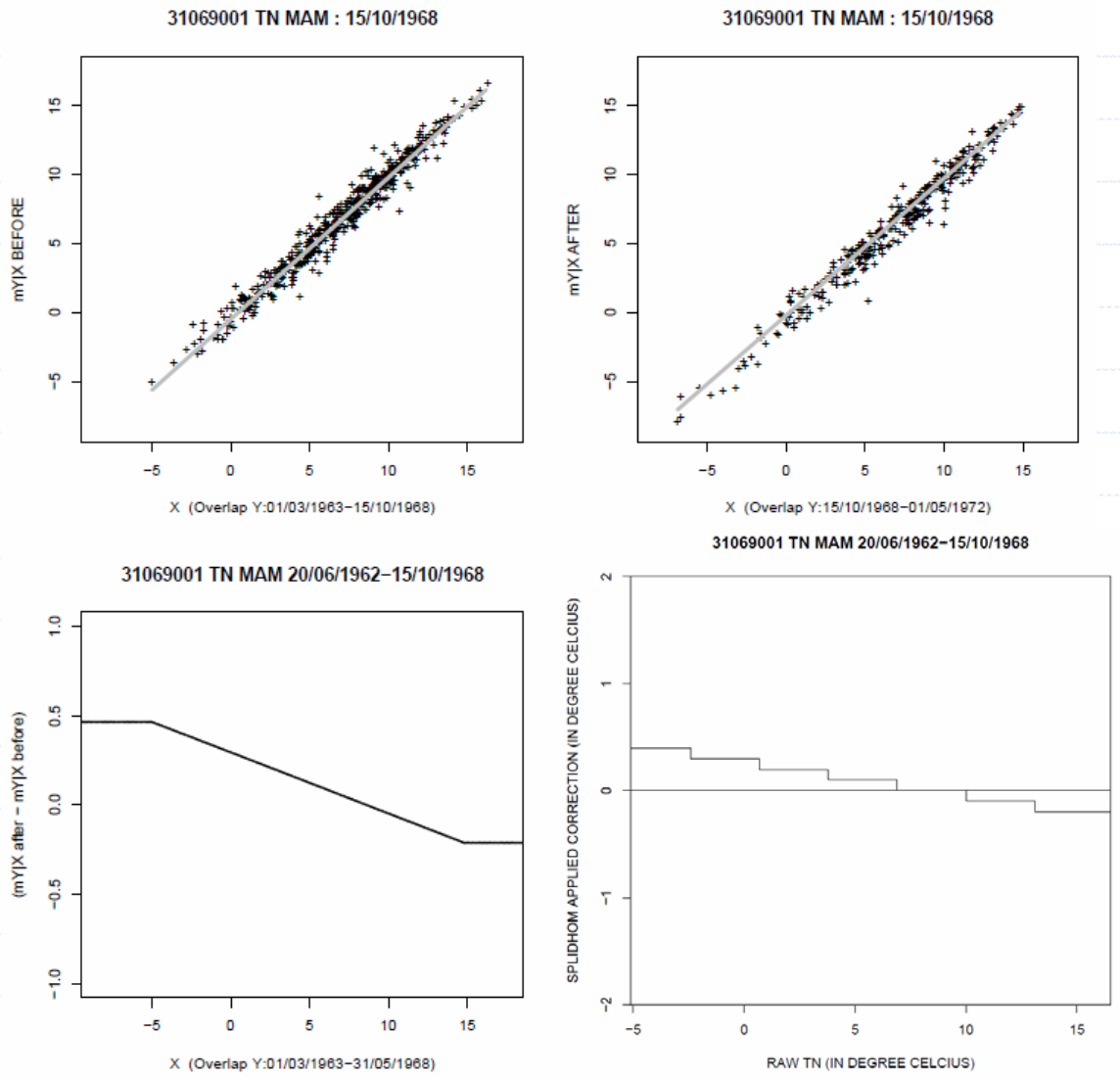


DELLA-MARTA AND WANNER, JOURNAL OF CLIMATE 19 (2006) 4179-4197

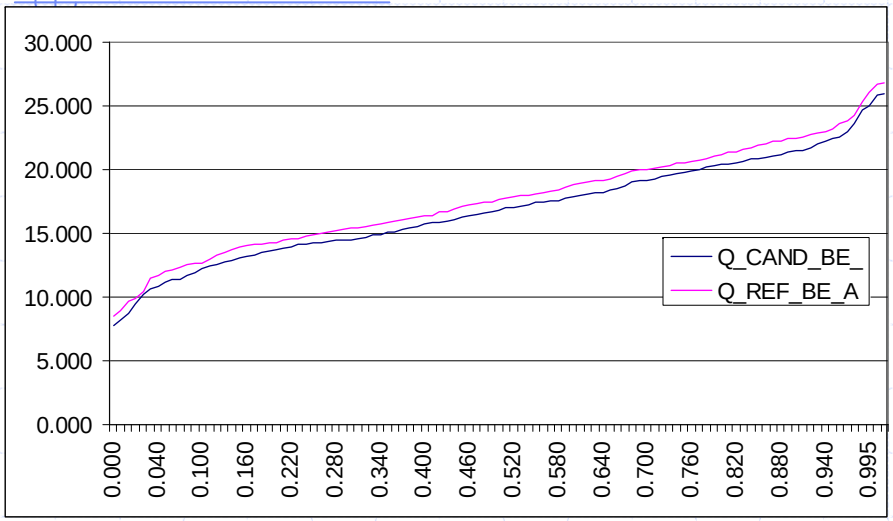
SPLIDHOM (SPLIne Daily HOMogenization), Olivier Mestre

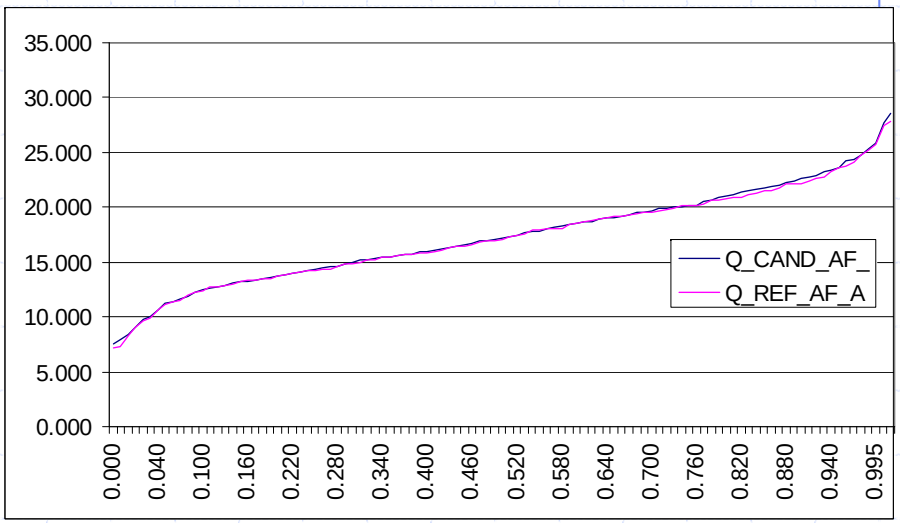
- **direct non-linear spline regression approach (x rather than a correction based on quantiles),**

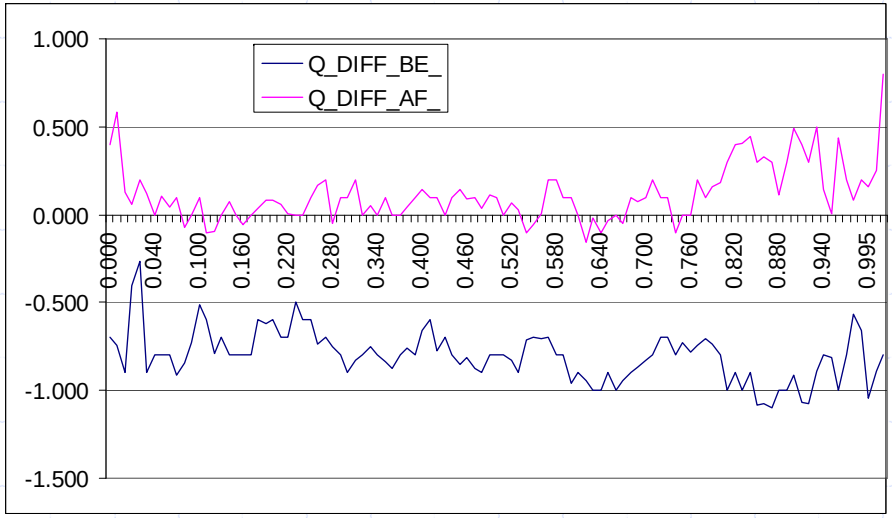
cubic smoothing splines for estimating regression functions



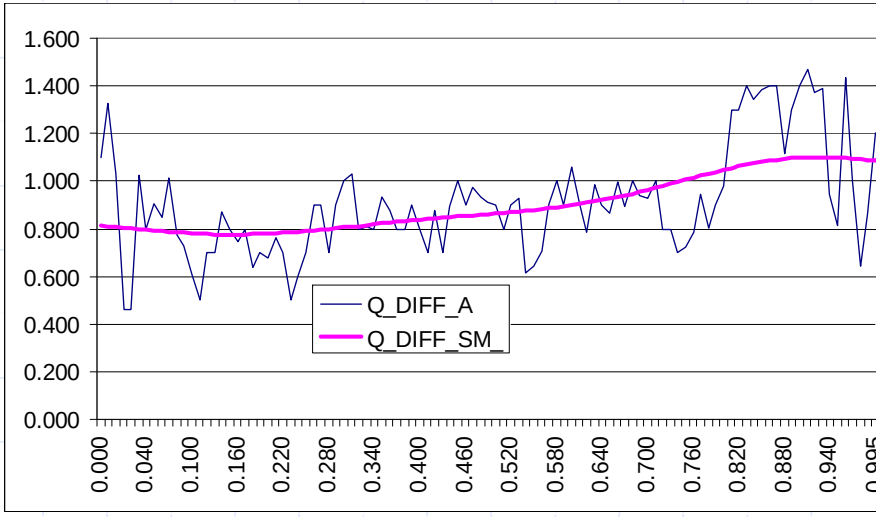
Distribution Adjusting by Percentiles (DAP)



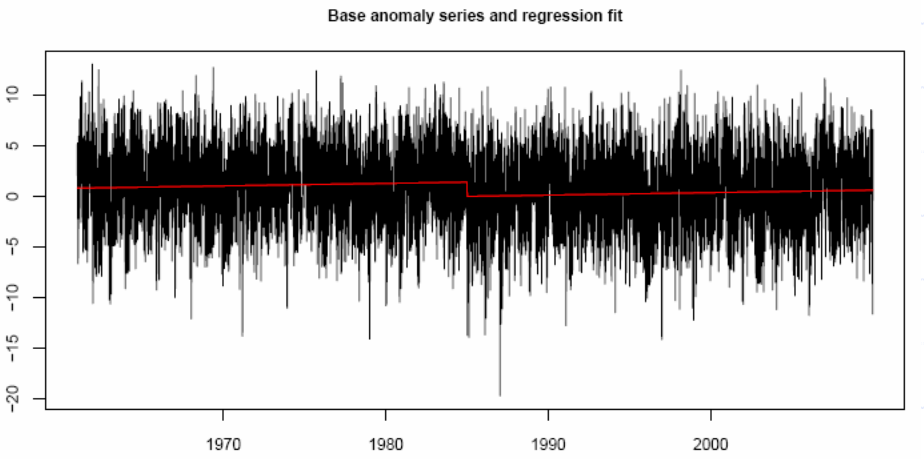


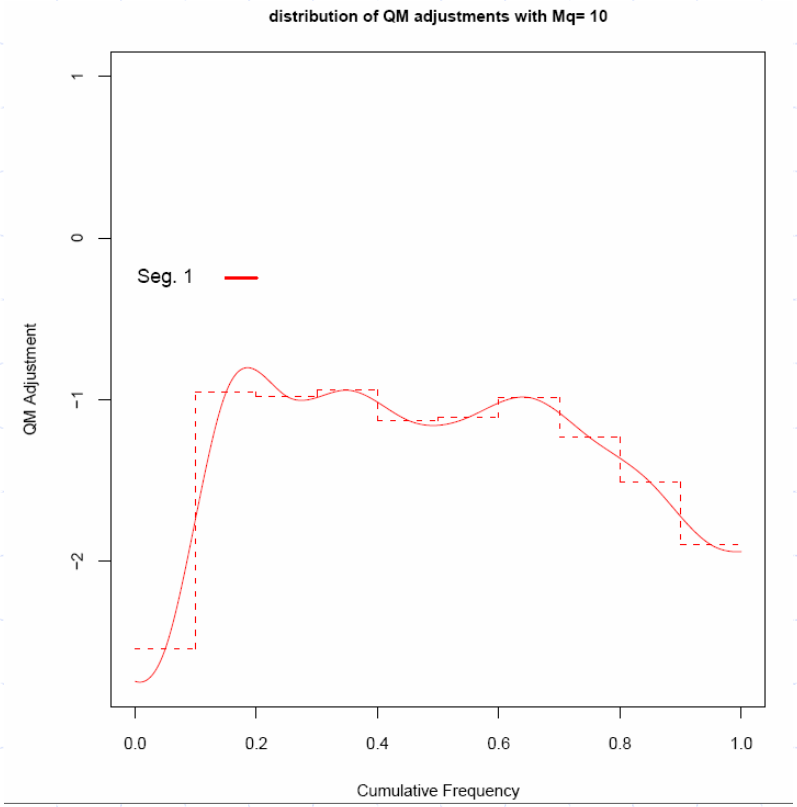


percentiles



QM algorithm, Xiaolan L. Wang



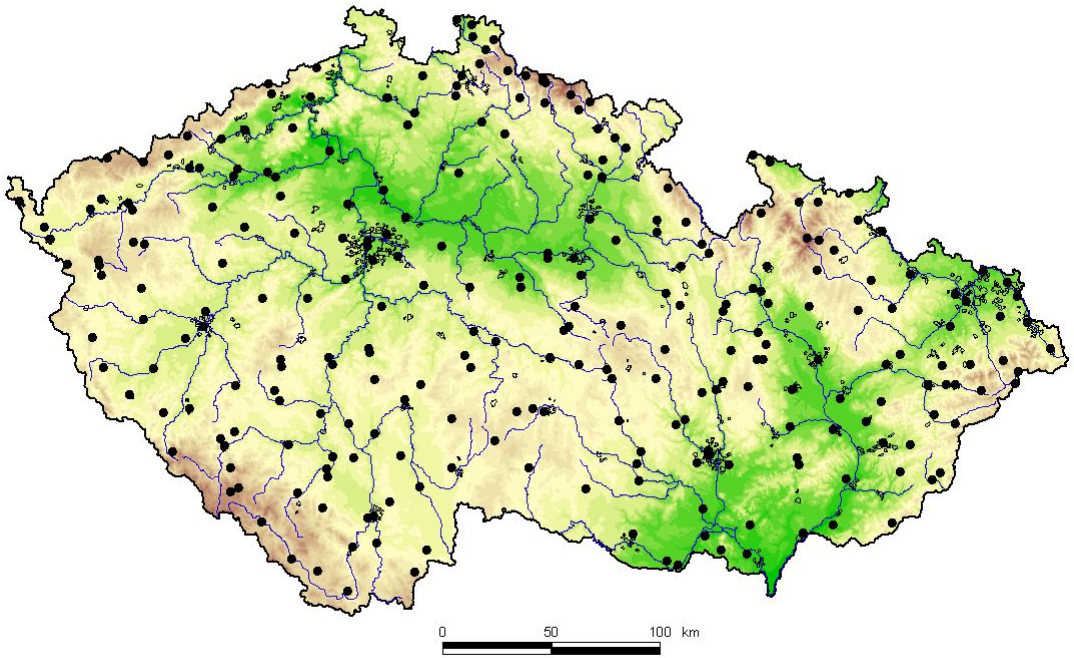


The F(l) Ds(l) relationship for each segment that needs to be adjusted (step lines) and the corresponding fitted natural cubic splines (i.e., distribution of the QM adjustments over empirical cumulative frequency) for two different Mq values.

distribution of QM adjustments with Mq = 10

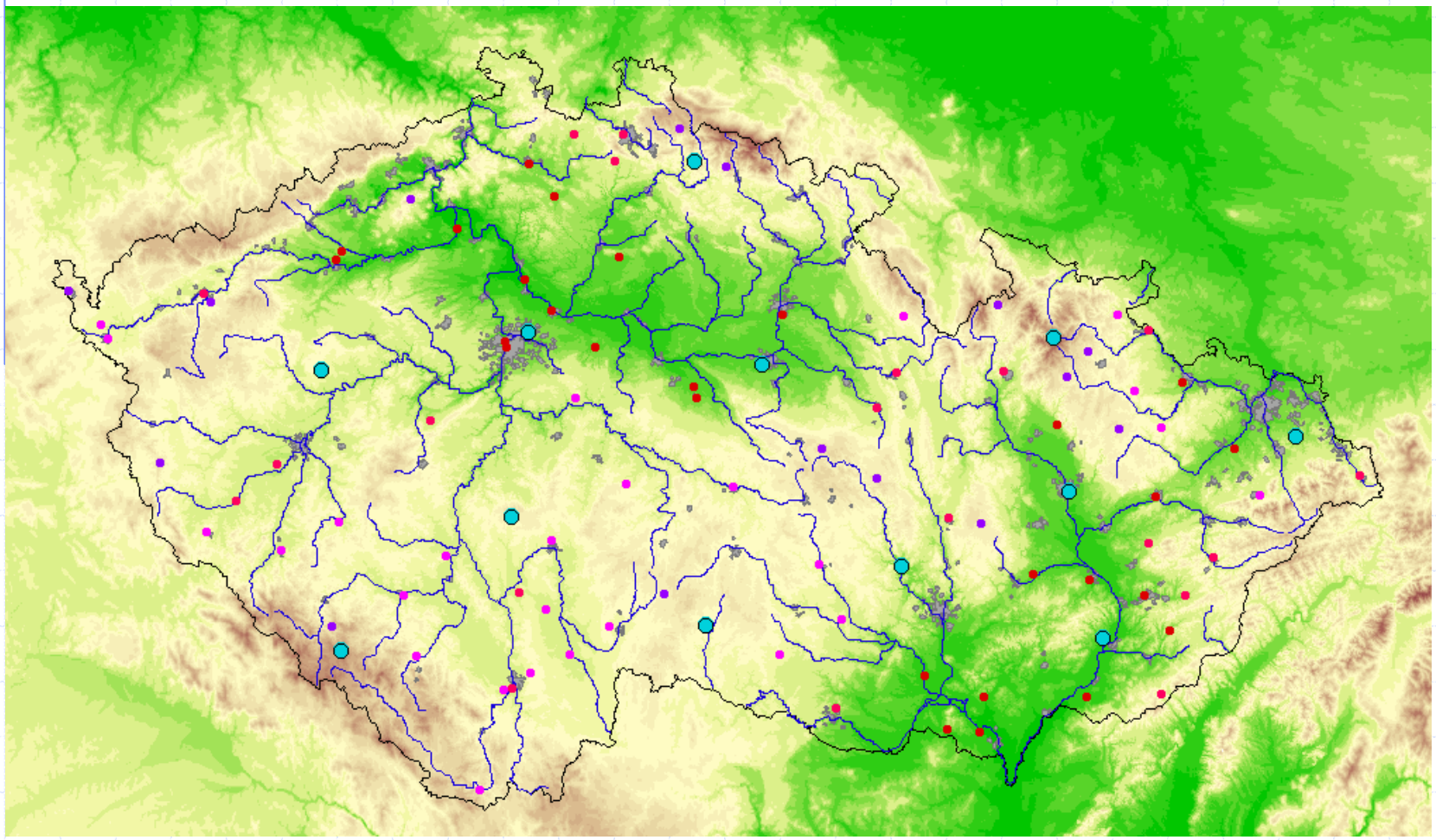
Data for the methods comparison

- Technical series for the Czech republic (267 climatological stations, quality controlled, homogenized, gaps filled)



100 stations selected,

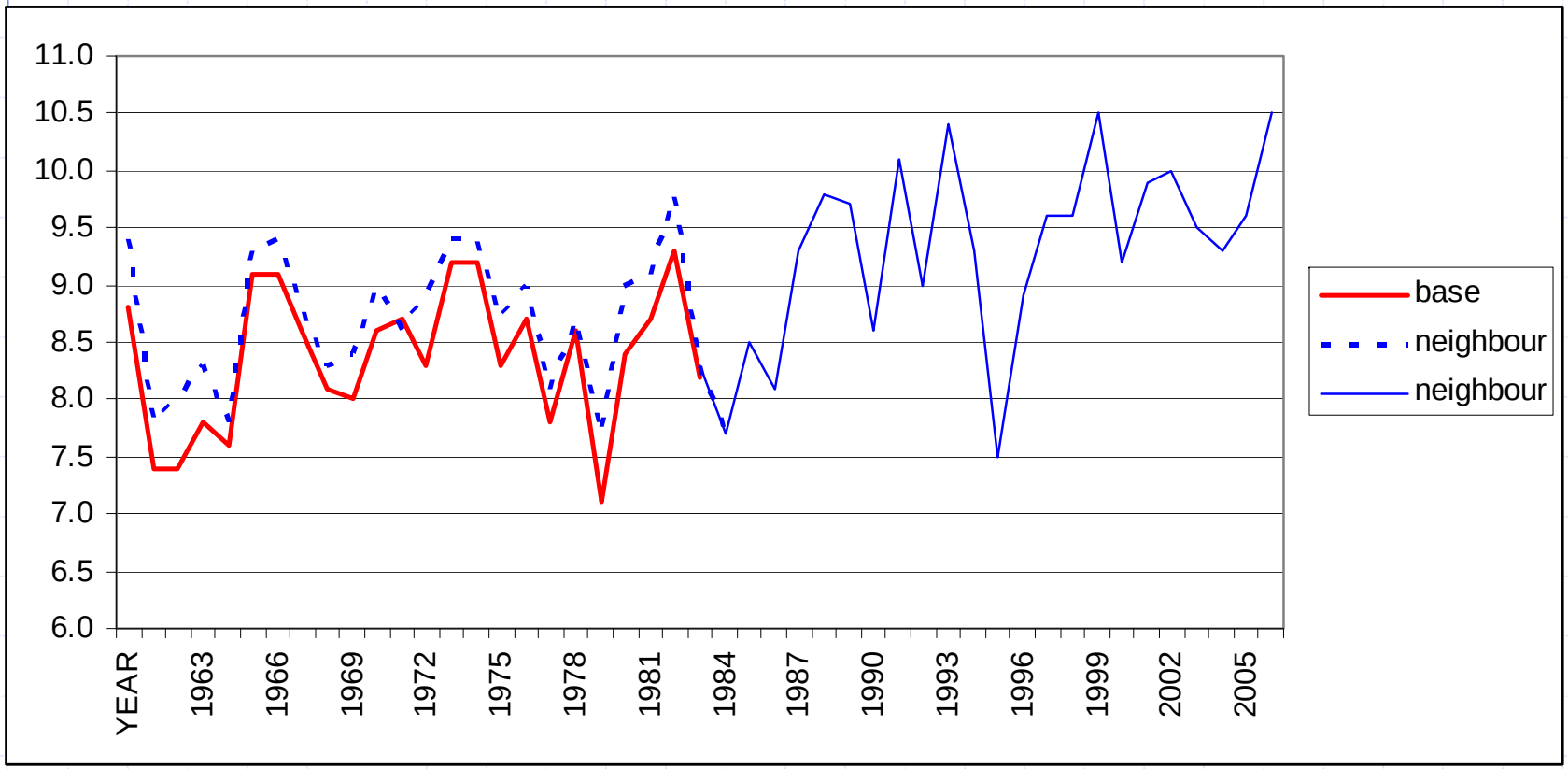
here only results for 12 stations and air temperature are presented



Creating the testing dataset

- Combining base station (1961 – 1984) with the first nearest neighbour (1985-2009)
- The „break“: 1/1/1985
- Reference series calculated from the other nearest neighbours

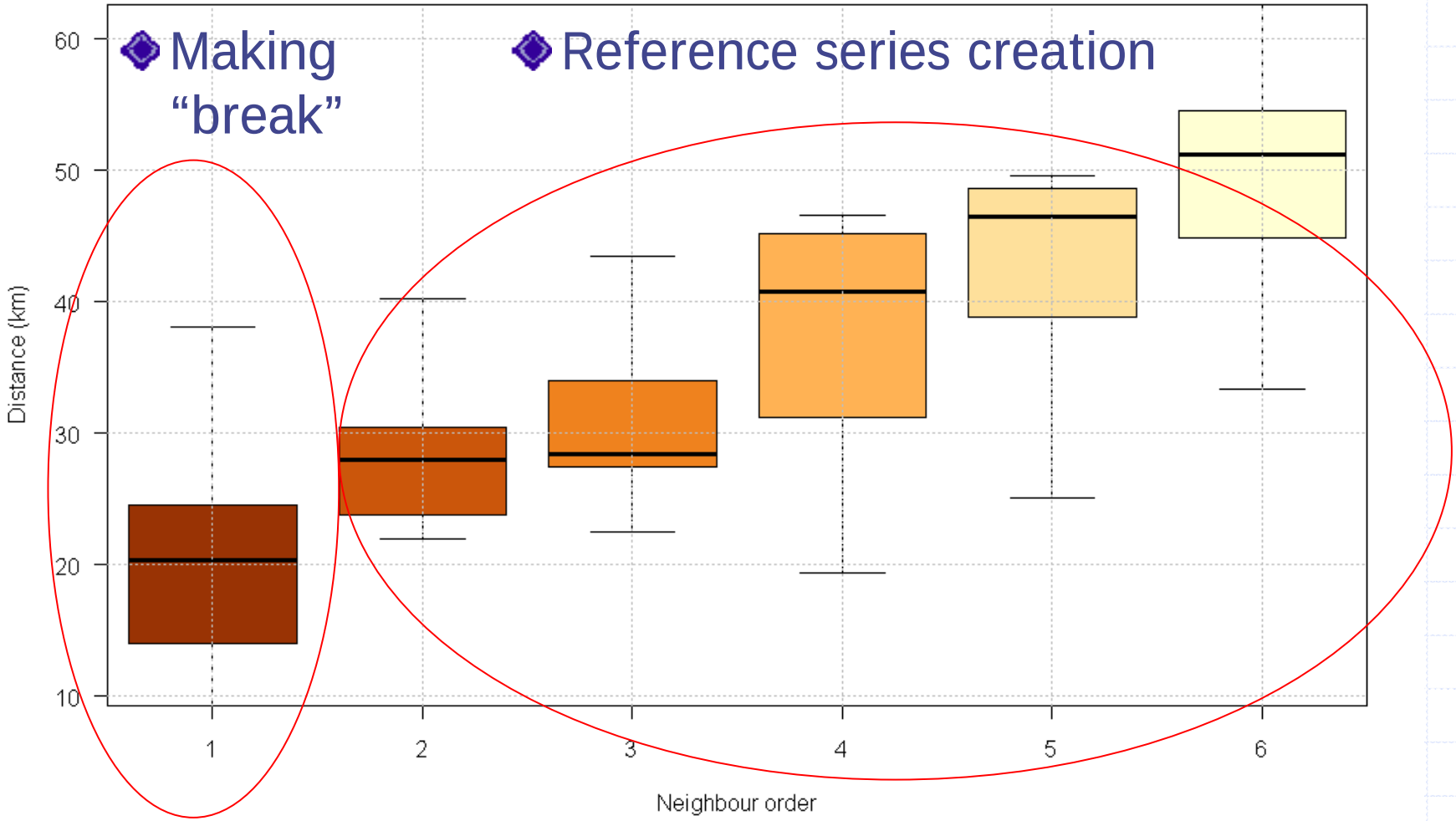
Creating the testing dataset



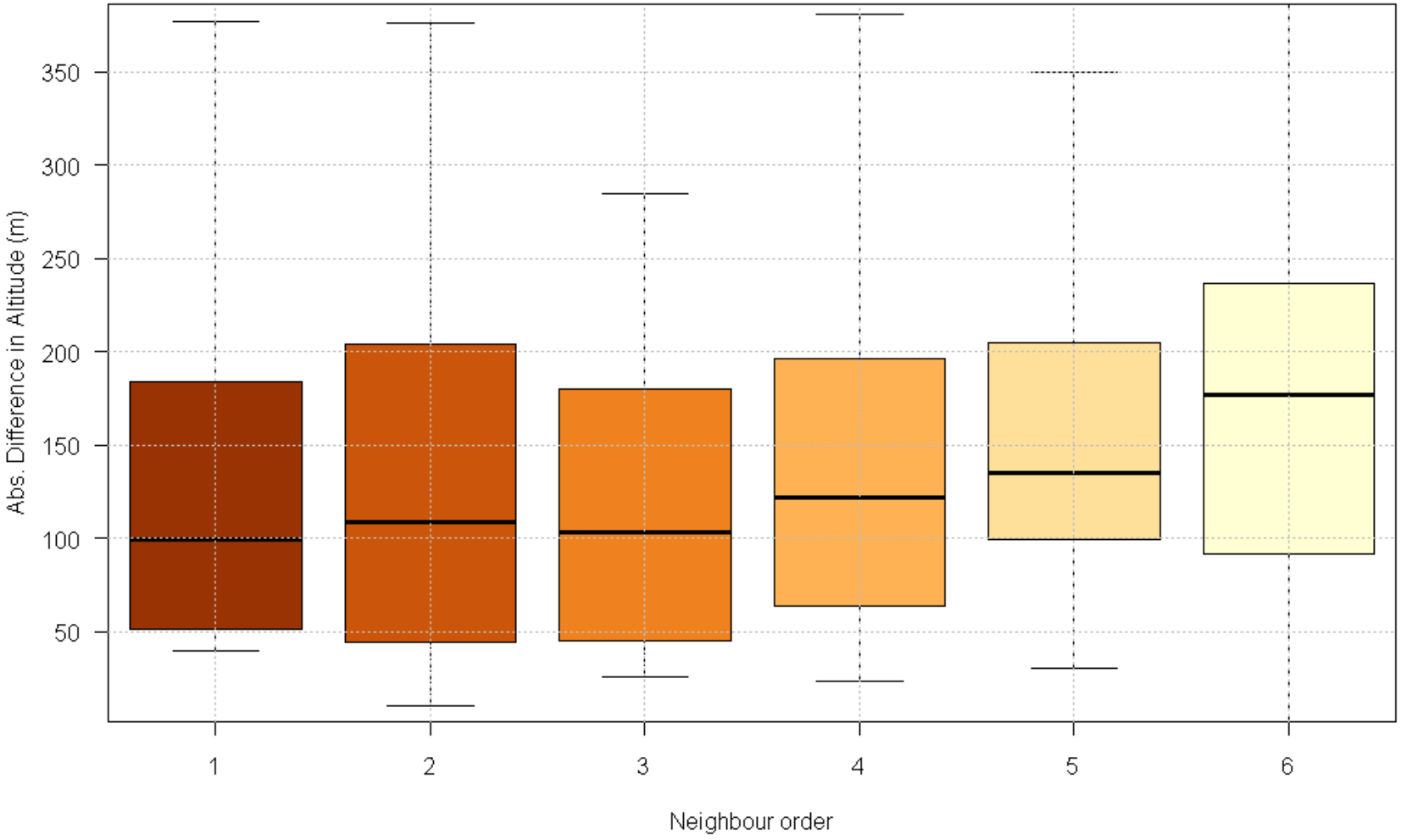
Fingind neighbours

Id_1	Id_2	Region	Begin	End	Date_inhom	Length	Remark	Correl	Distance	Azimuth	Alt_1	Alt_2	K1	K2
			. .	. .	. .									
B1BYSH01			1.1.1961	31.12.2009	2.7.1985	17897	5 st. (l:25.40)		41.97	-68.8	314			
B1BYSH01	O3HRAN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		17.23	79.5	314	248	-66.00	-1.00
B1BYSH01	B1ZLIN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		18.75	-96.1	314	261	-53.00	-1.00
B1BYSH01	B1VIZO01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		23.02	-56.9	314	313	-1.00	-1.00
B1BYSH01	O3VSET01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		24.30	-13.7	314	387	73.00	1.00
B1BYSH01	B1KROM01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		25.40	-150.8	314	233	-81.00	-1.00
			. .	. .	. .									
B1IVAN01			1.1.1961	31.12.2009	2.7.1985	17897	5 st. (l:41.19)		70.41	12.8	245			
B1IVAN01	B1KROM01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		20.21	-7.6	245	233	-12.00	-1.00
B1IVAN01	O2OLOM01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		32.89	64.7	245	210	-35.00	-1.00
B1IVAN01	B1STME01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		34.71	-44.0	245	221	-24.00	-1.00
B1IVAN01	B2BOSK01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		37.35	145.1	245	400	155.00	1.00
B1IVAN01	B1ZLIN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		41.19	-12.3	245	261	16.00	1.00
			. .	. .	. .									
B1KROM01			1.1.1961	31.12.2009	2.7.1985	17897	5 st. (l:32.93)		29.20	43.1	233			
B1KROM01	B1IVAN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		20.21	172.2	233	245	12.00	1.00
B1KROM01	B1ZLIN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		21.12	-17.0	233	261	28.00	1.00
B1KROM01	B1STME01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		22.02	-77.3	233	221	-12.00	-1.00
B1KROM01	B1BYSH01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		25.40	29.4	233	314	81.00	1.00
B1KROM01	O2OLOM01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		32.93	100.2	233	210	-23.00	-1.00
			. .	. .	. .									
B1LUHA01			1.1.1961	31.12.2009	2.7.1985	17897	5 st. (l:31.55)		31.68	114.3	254			
B1LUHA01	B1VIZO01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		14.35	64.9	254	313	59.00	1.00
B1LUHA01	B1ZLIN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		16.07	121.9	254	261	7.00	1.00
B1LUHA01	B1STRN01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		22.98	-99.6	254	383	129.00	1.00
B1LUHA01	B1STME01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		23.98	-176.2	254	221	-33.00	-1.00
B1LUHA01	O3VSET01		1.1.1961	31.12.2009	. .		17897 y. comm.per.		31.55	57.1	254	387	133.00	1.00

Distances from neighbours



Differences in Altitude with neighbours

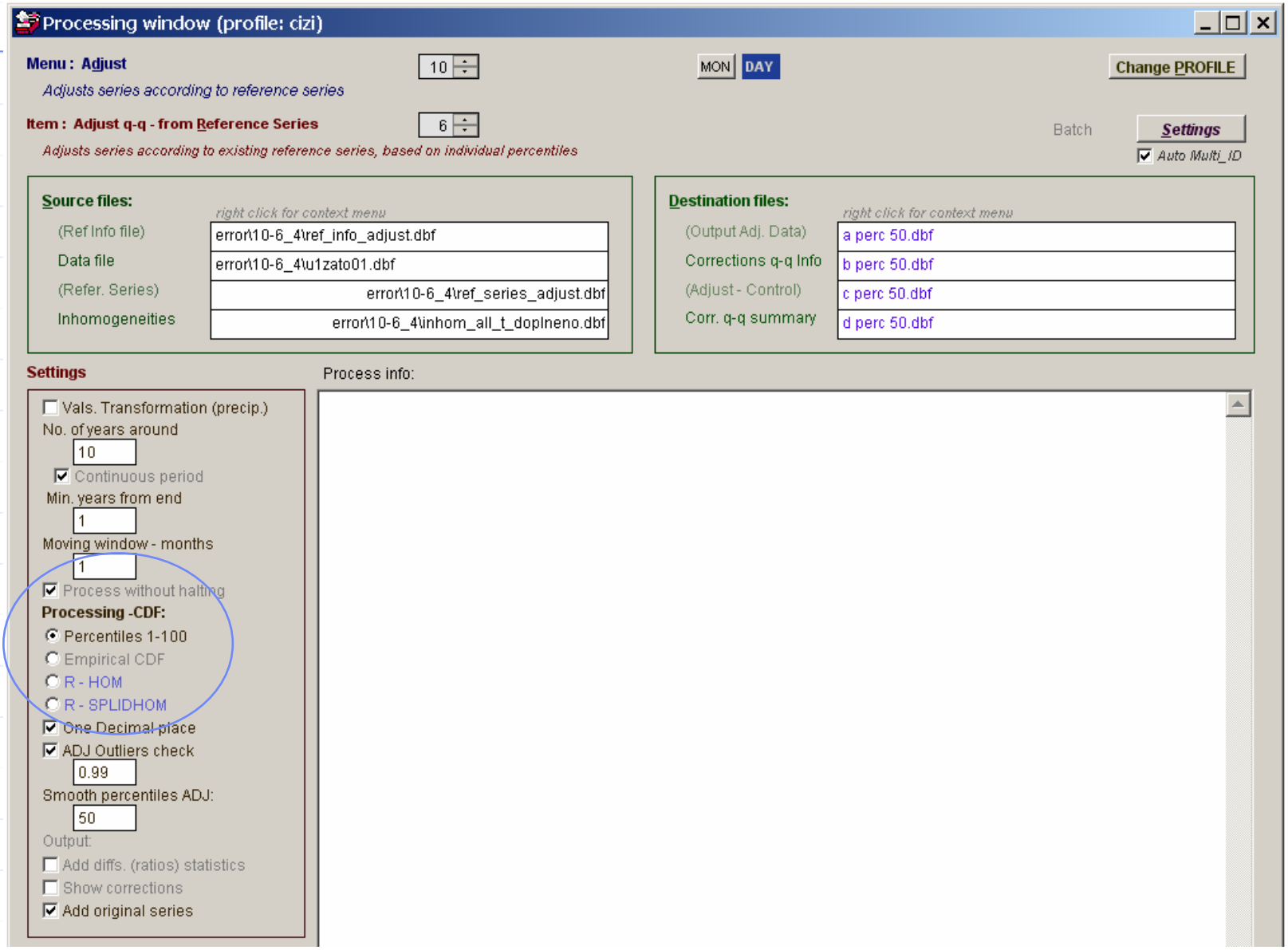


Tools used

- Using ProClimDB software
- Various methods included (DAP, HOM, SPLIDHOM, QM, Delta method)

Comparison of the methods, ProClimDB software

Variable correction methods



Adjusting data

- using reference series based on distances (5 nearest neighbours)
- adjustment: from differences (ratios), *x* years before and after a change, monhtly

Validation and statistical characteristics

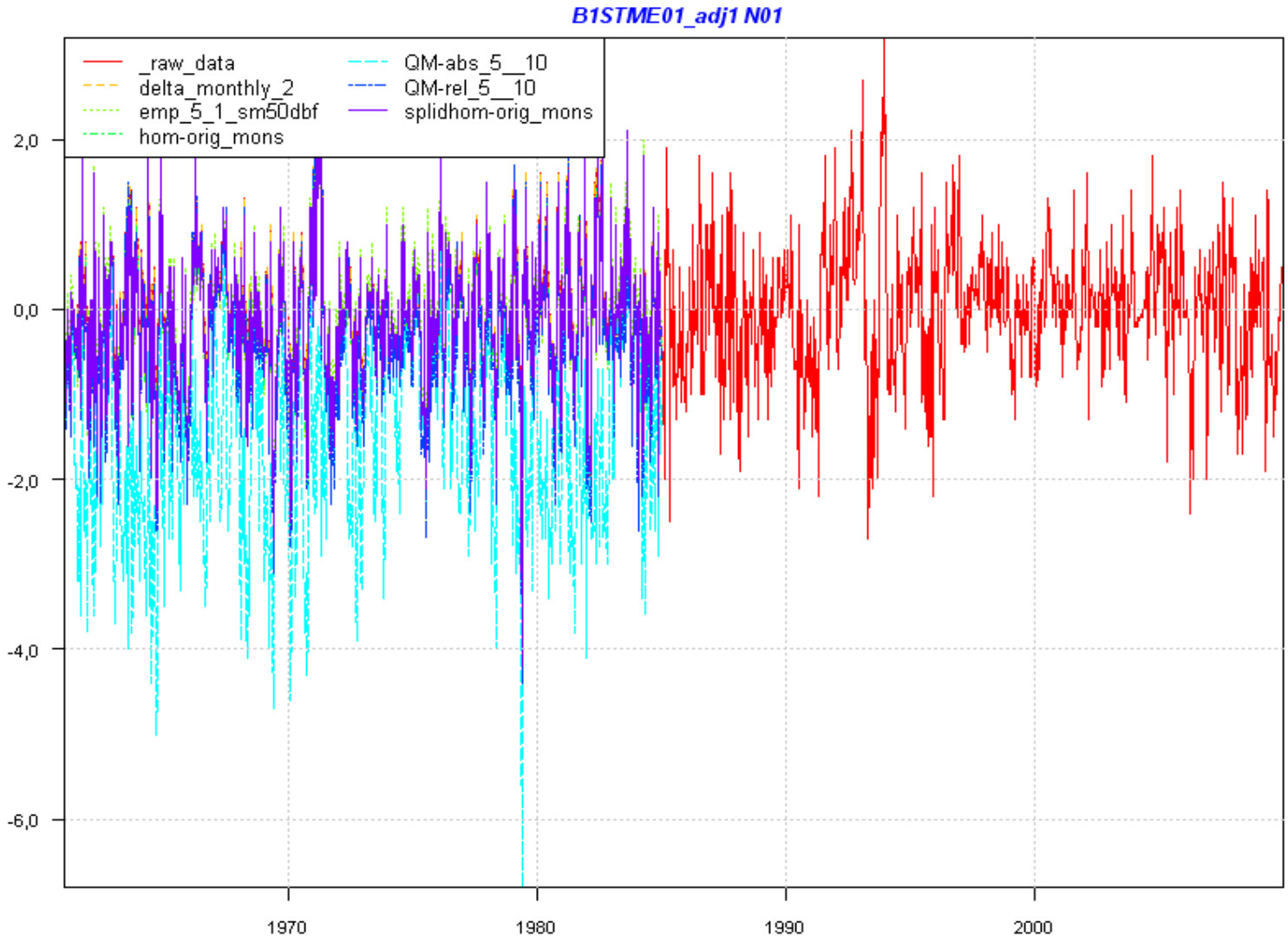
(original x adjusted neighbour station)

- Spearman rank correlation
- Min. difference (Bias)
- Max. difference (Bias)
- Abs. range of difference (Bias)
- Bias (mean)
- MAE (mean absolute error)
- RMSE (root mean squared error)
- NRMSE (normalized RMSE by the range of
- STD of differences
- p-value of t-test for Paired Values
- F-test (STDs)
- p-value of F-test
- t-test (AVGs, INDEPENDENT! samples)
- p-value of t-test
- … K-S test

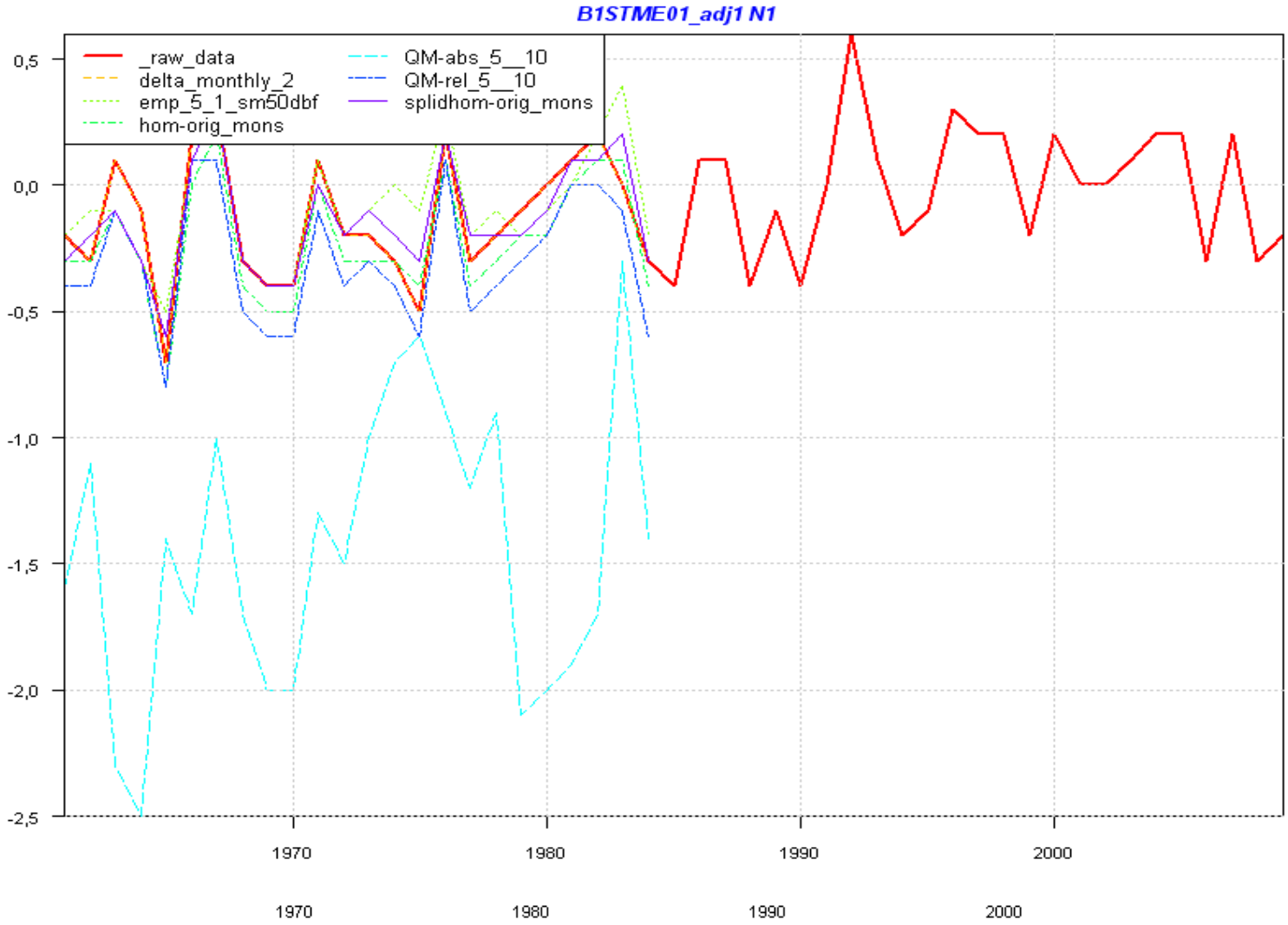
For each series:

- Count
- AVG
- STD
- Skew
- Kurtosis
- Minimal value
- Maximal value
- b1 coeff. (lin. regr)
- 01st percentile
- 05th percentile
- 10th percentile
- 25th percentile
- 50th percentile
- 75th percentile
- 90th percentile
- 95th percentile
- 99th percentile
- Return periods

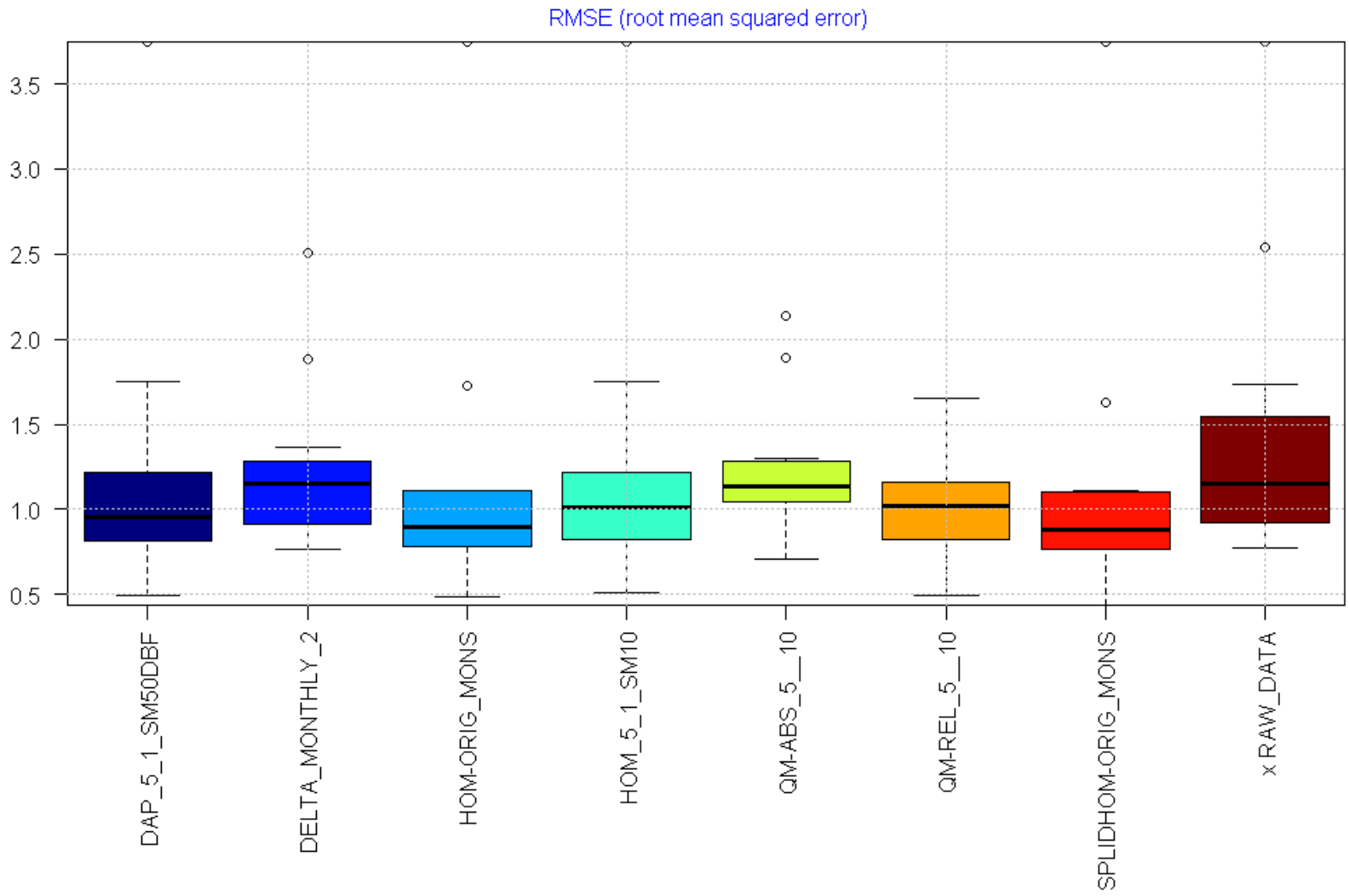
Differences with the base station



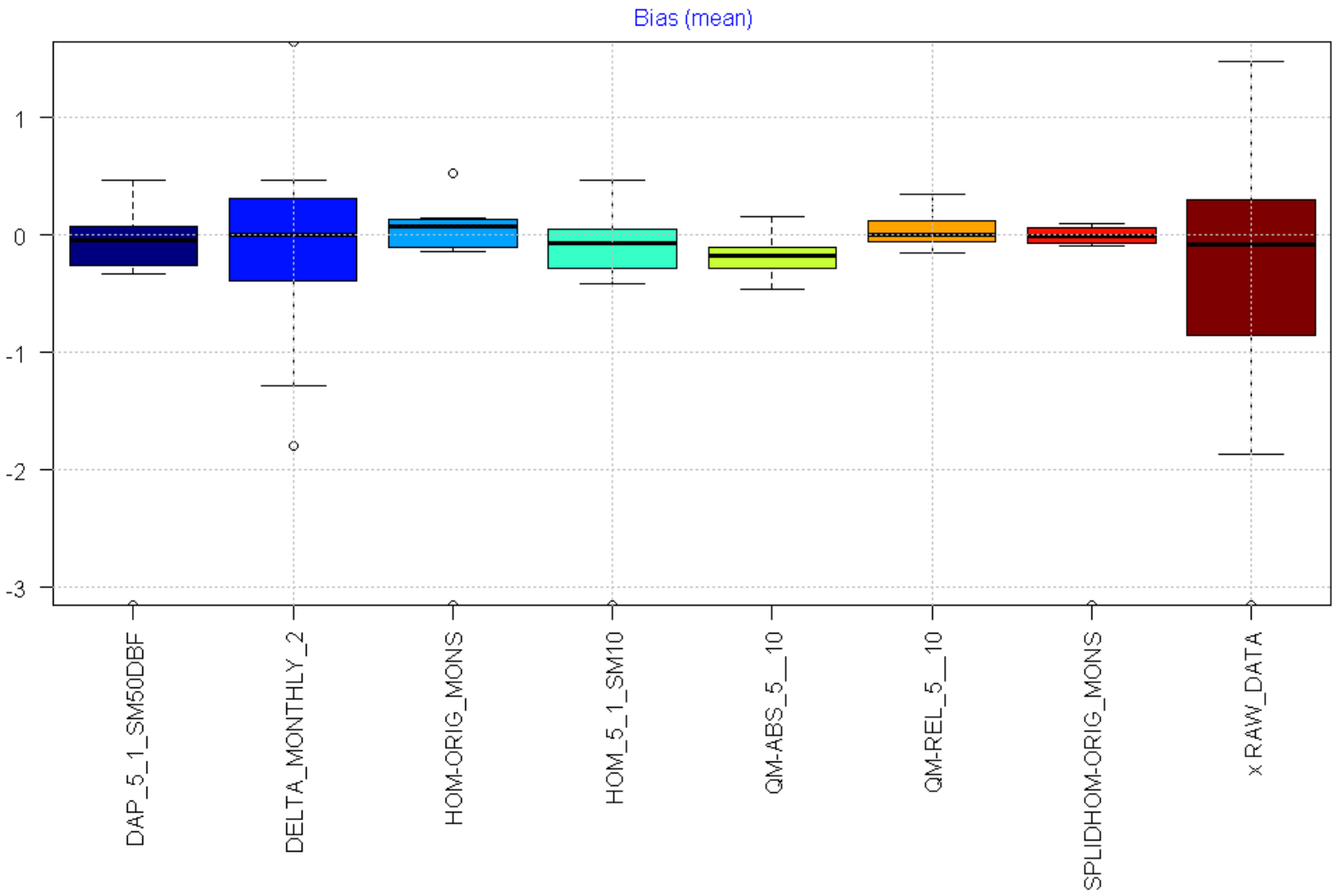
Differences with the base station, monthly



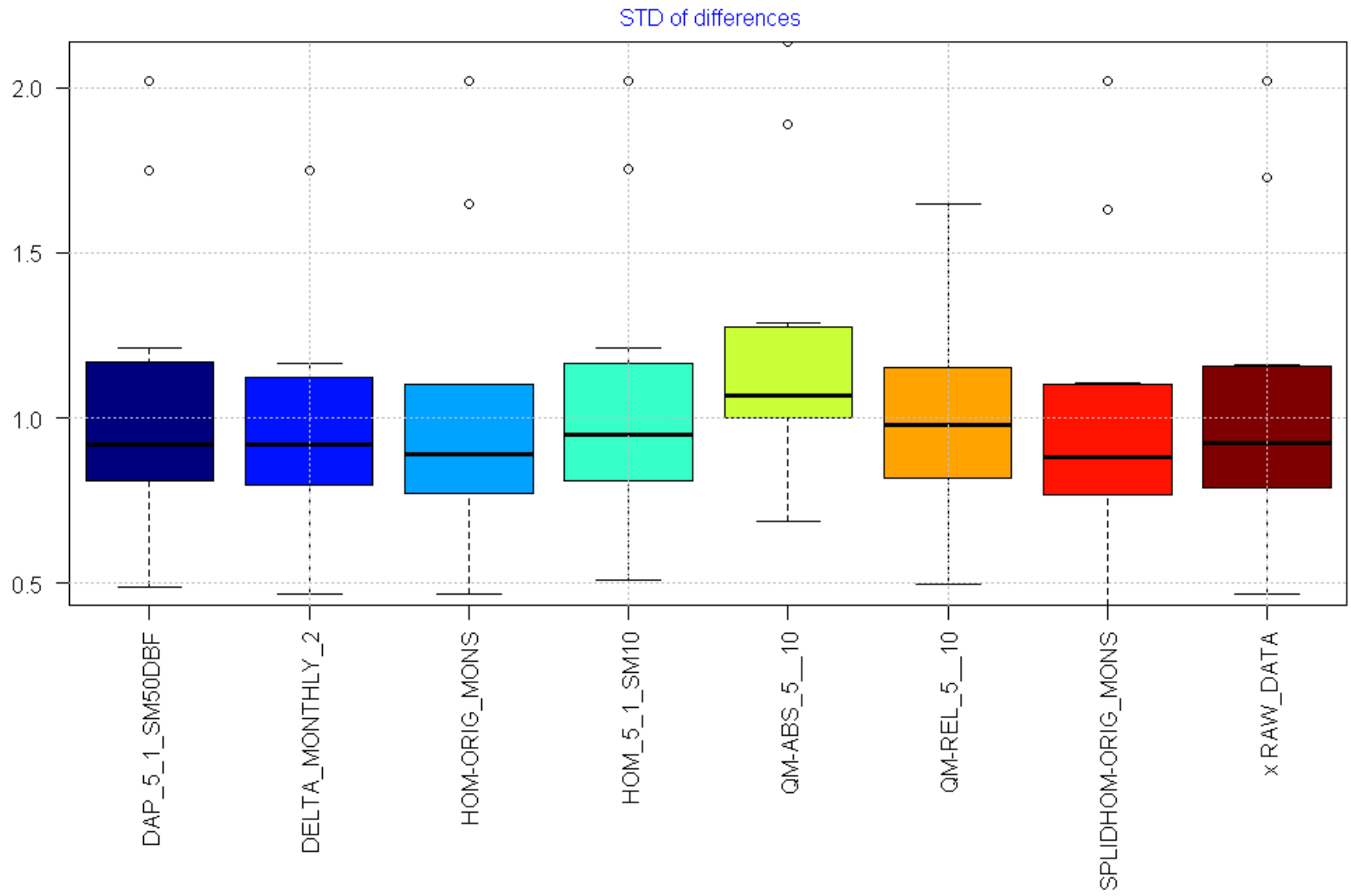
Evaluation of the results (1961-1984 period, true and adjusted series comparison), RMSE



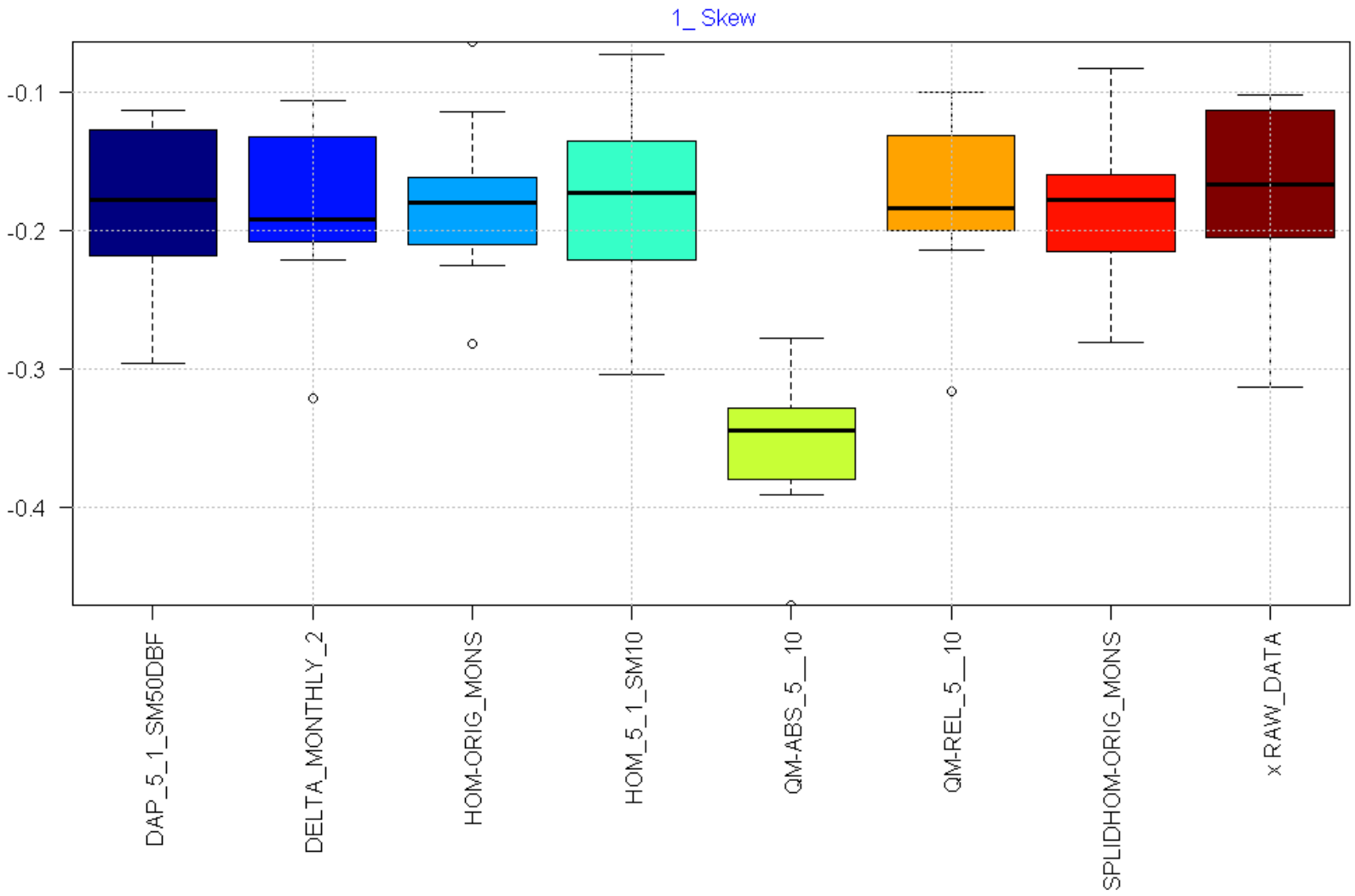
Evaluation of the results (1961-1984 period, true and adjusted series comparison), Bias



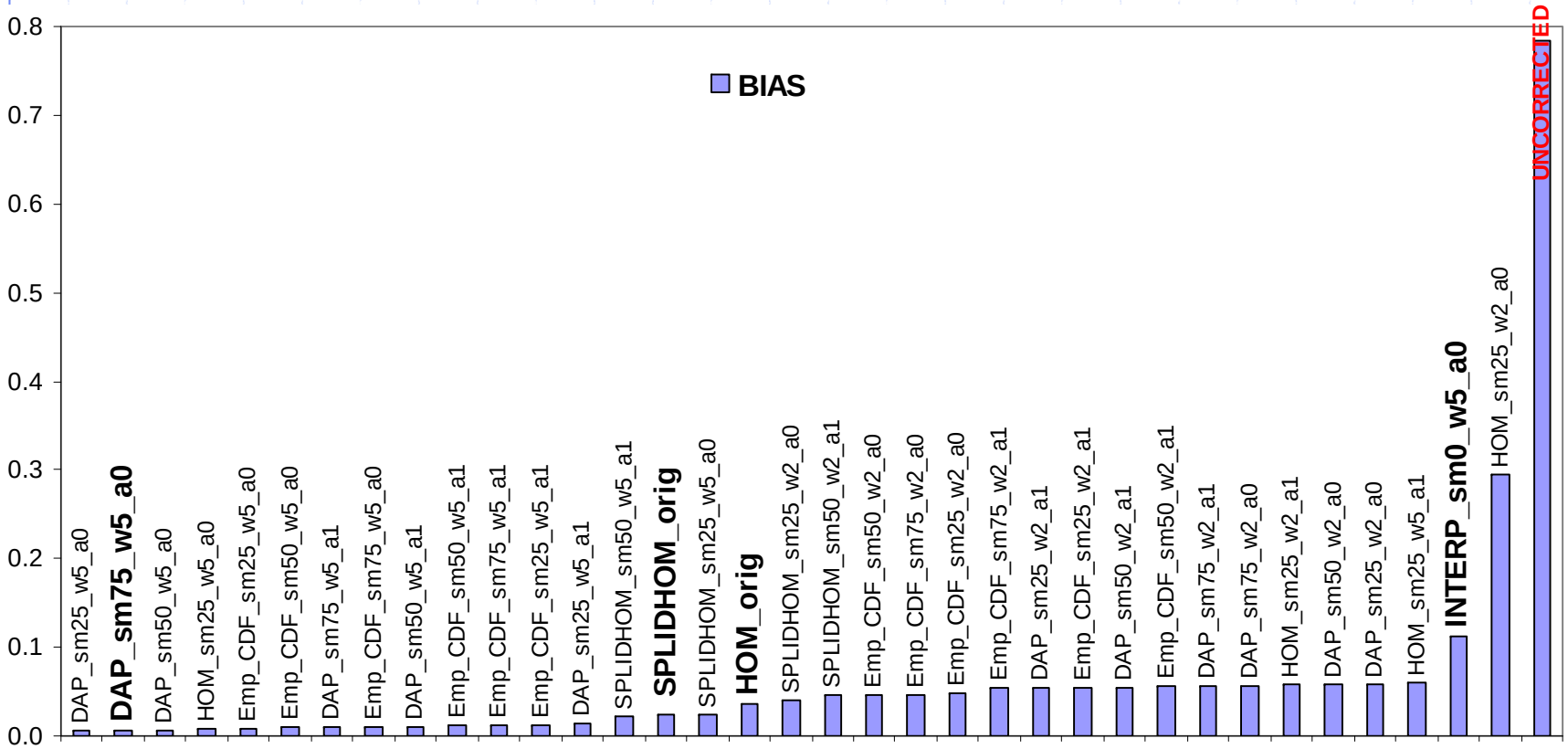
Evaluation of the results (1961-1984 period, true and adjusted series comparison), STD of differences



Evaluation of the results (1961-1984 period, true and adjusted series comparison), Skew of adjusted data

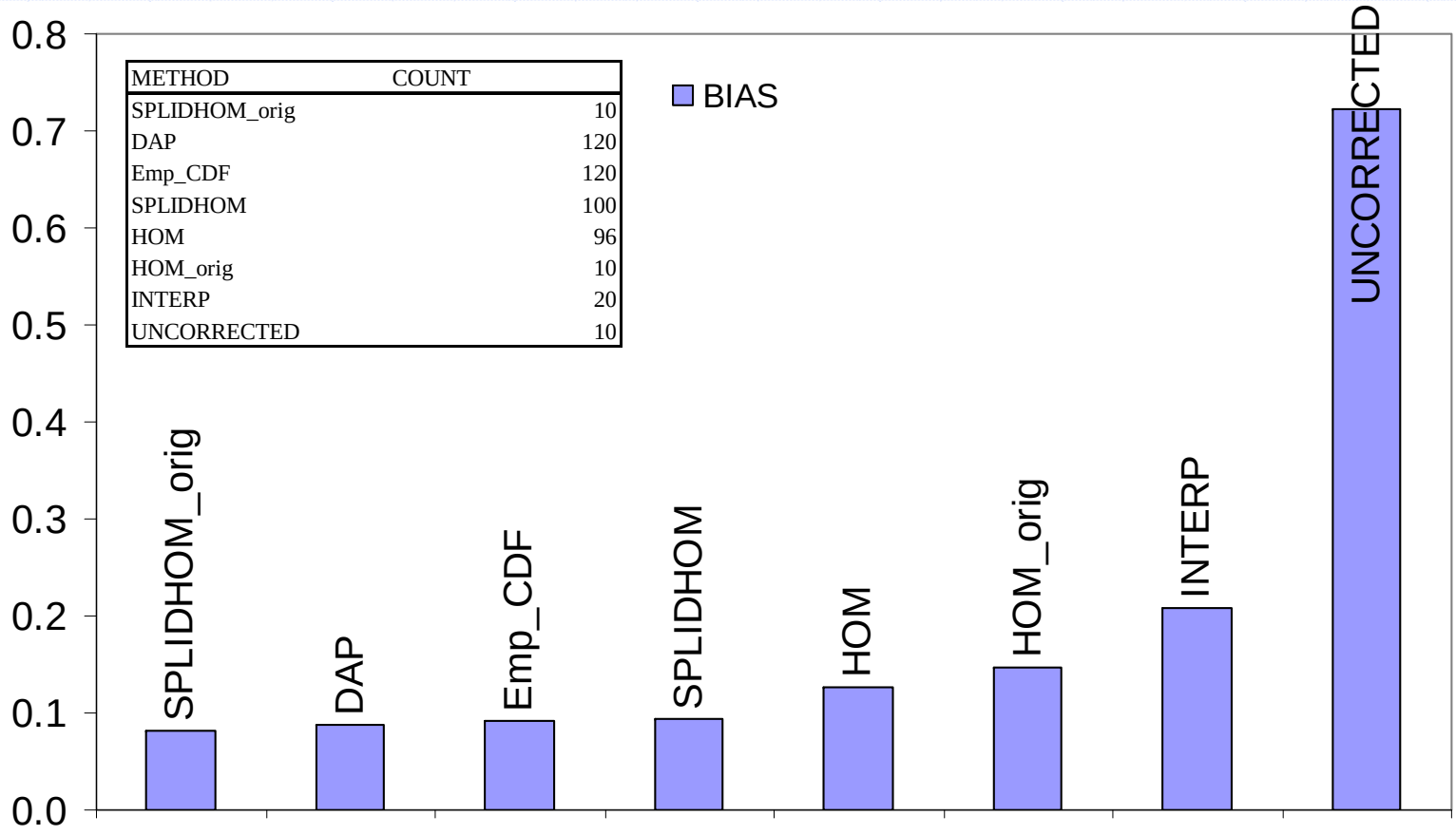


Comparison of correction methods for daily air temperature, ***PREVIOUS RESULTS***



Comparison of correction methods for daily air temperature,

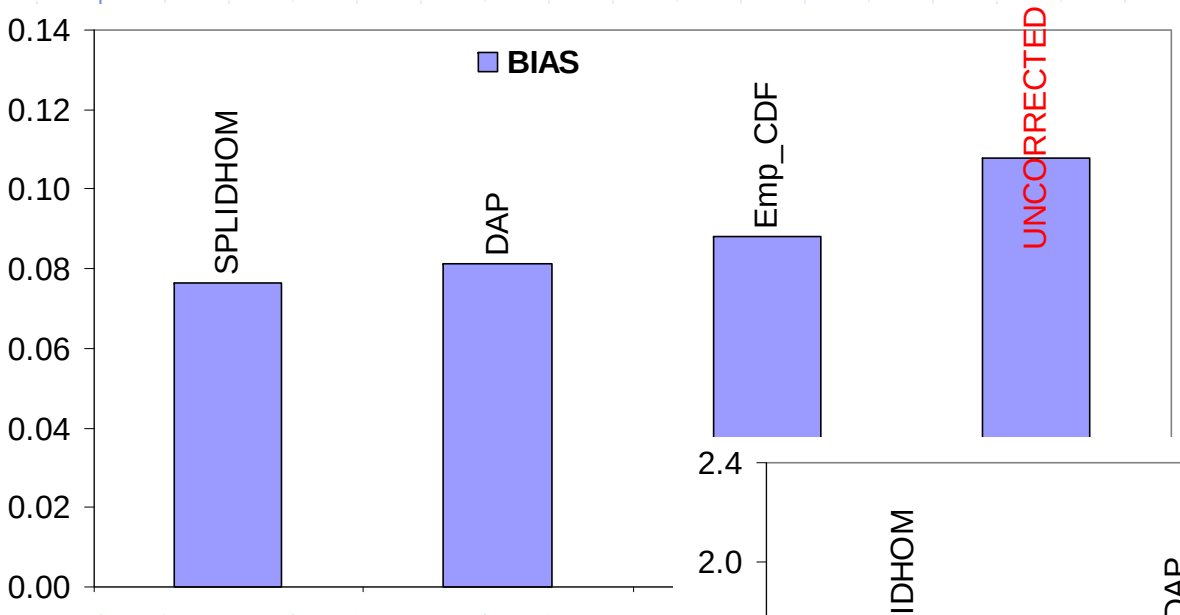
PREVIOUS RESULTS



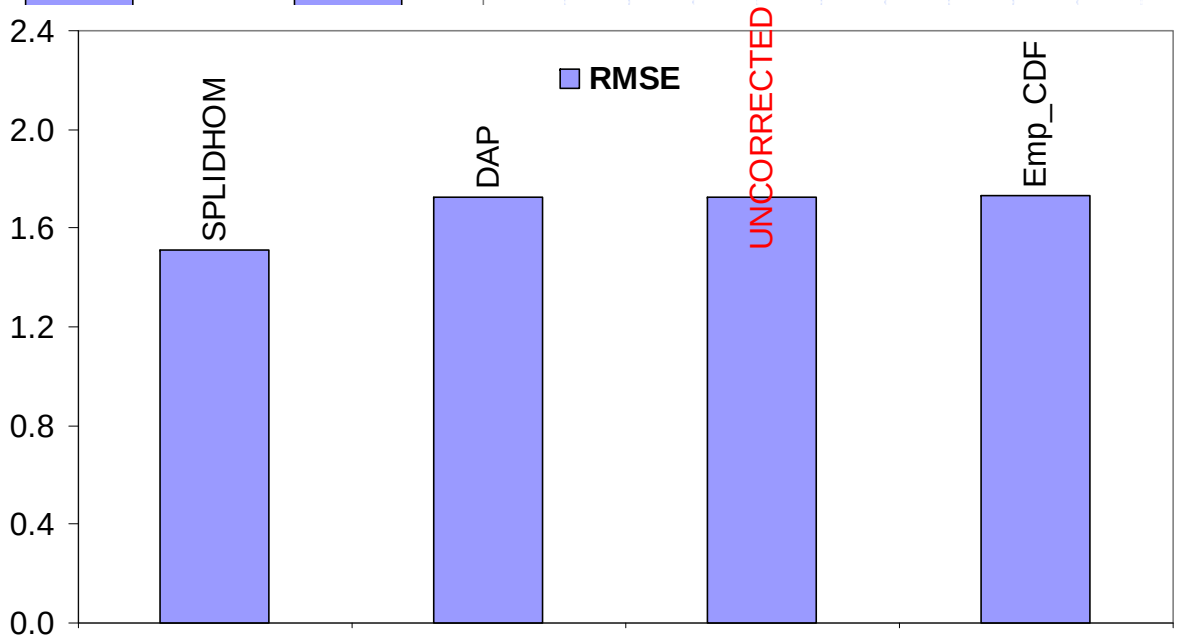
METHOD	COUNT
SPLIDHOM_orig	10
DAP	120
Emp_CDF	120
SPLIDHOM	100
HOM	96
HOM_orig	10
INTERP	20
UNCORRECTED	10

Comparison of correction methods for daily Precipitation,

- annually, average ***PREVIOUS RESULTS***

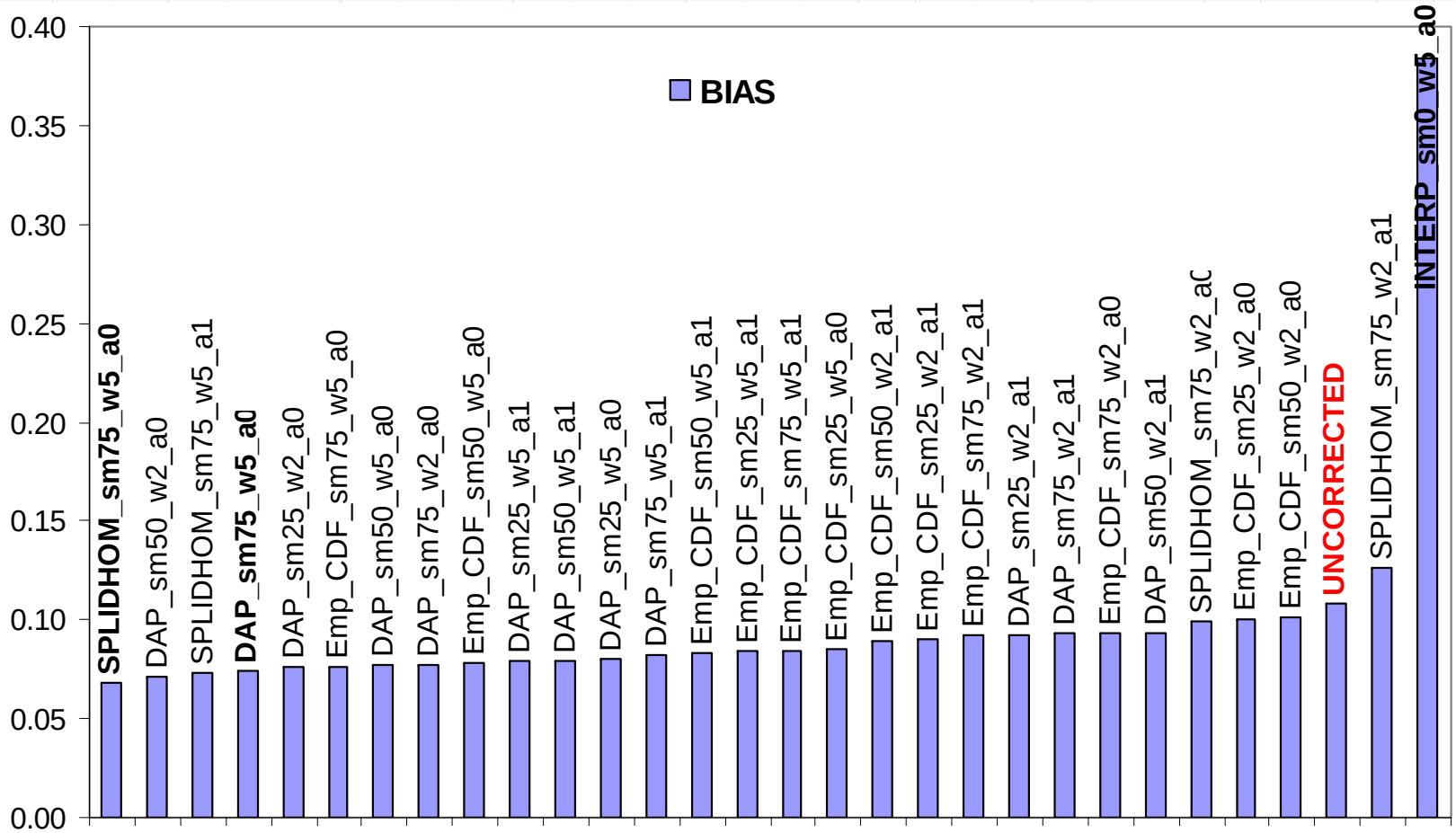


METHOD	COUNT
SPLIDHOM	48
DAP	120
Emp_CDF	120
UNCORRECTED	10
INTERP	19



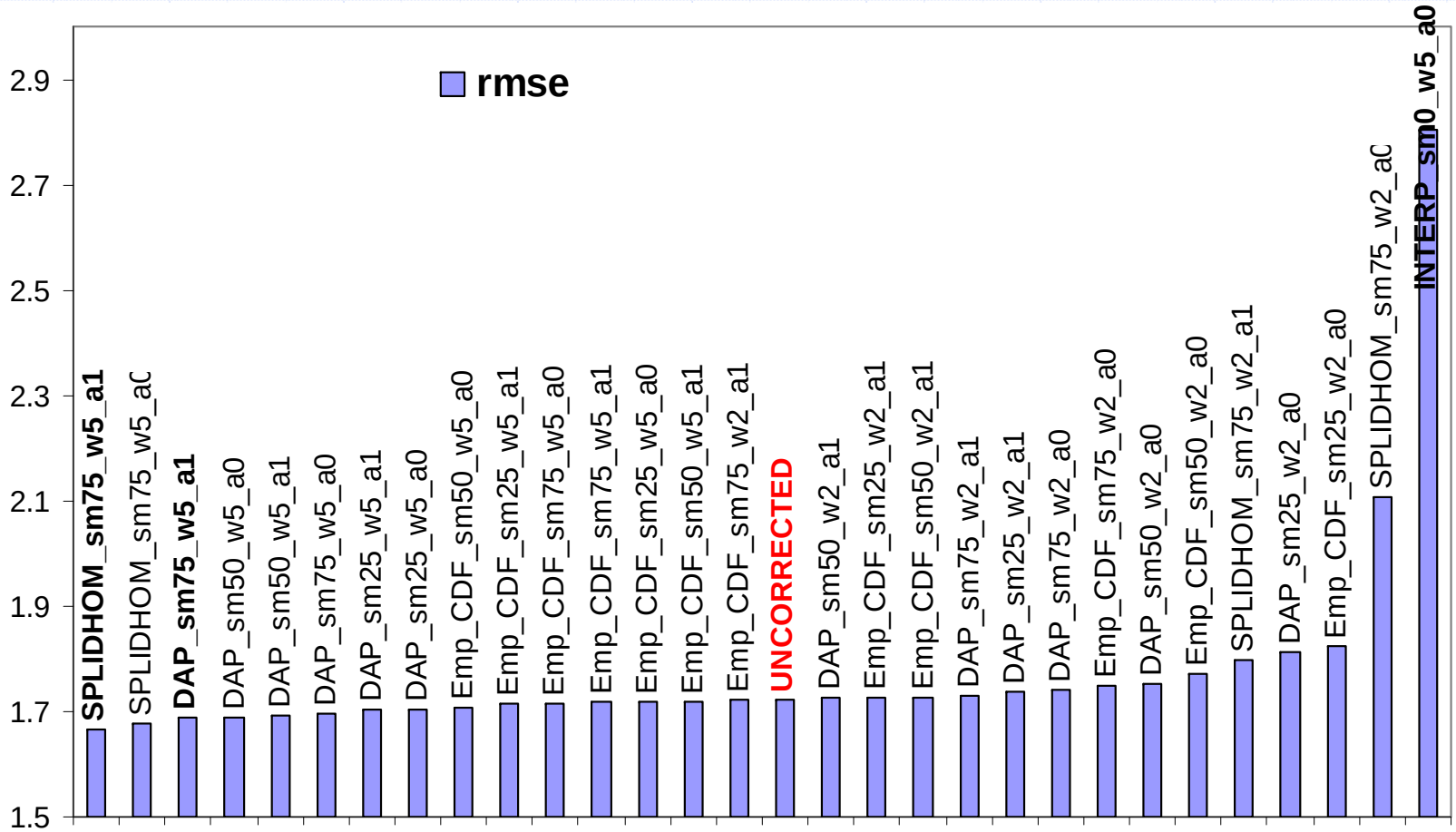
Comparison of correction methods for daily Precipitation,

average from 10 testings ***PREVIOUS RESULTS***



Comparison of correction methods for daily Precipitation,

average from 10 testings, ***PREVIOUS RESULTS***



Multi-element approach
for the correction of inhomogeneites in daily data

- approach tested on 69 stations from SE of the Czech Republic, period 1990 – 2008
- question: does discriminating of days according to a certain weather types lead to better description of neighbours differences (e.g. for air temperature)
- results for
  - cloudiness, three categories
  - cyclone or anticyclone situations from catalogue of CHMI (about 30 types, subjective)
  - CKMEANSC09

Neighbours comparison, discriminating into categories according to **cloudiness**

- Three categories: C: cloudy <8/10, 10/10>, P: partially cloudy (2/10, 8/10), S: clear sky<0/10, 2/10>
- For each station individually
- Cloudy conditions prevailed in 41% of all days (from all the stations), partially cloudy in 46% and clear sky days occurred in 13% of all cases.
- only dicriminating temperatures according to days with **cloudy** condititons lead to higher correlations between candidate and neighbour stations and lower RMSE
- Conclusion: it does not make much sense to divide air temperature into days according to cloudiness

Neighbours comparison, discriminating into categories according to **cloudiness**

- Spearman rank correlation coefficient, median for 69 pairs of candidate and neighbour stations. Category: all – all days, C selection for cloudy days, P – partially cloudy, S – clear sky

Category	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	K_YEAI	K_DJF	K_MAM	K_JJA	K_SON
all	0.982	0.984	0.986	0.989	0.985	0.978	0.970	0.978	0.988	0.990	0.990	0.986	0.984	0.984	0.986	0.975	0.990
C	0.983	0.985	0.988	0.990	0.989	0.978	0.969	0.985	0.988	0.993	0.992	0.986	0.987	0.985	0.989	0.977	0.991
P	0.979	0.981	0.986	0.989	0.984	0.974	0.966	0.975	0.987	0.987	0.989	0.983	0.982	0.981	0.986	0.972	0.988
S	0.967	0.983	0.987	0.984	0.977	0.969	0.943	0.967	0.985	0.980	0.988	0.967	0.976	0.973	0.983	0.962	0.984

- RMSE, median for 69 pairs of candidate and neighbour stations. Category: all – all days, C selection for cloudy days, P – partially cloudy, S – clear sky

Category	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	K_YEAI	K_DJF	K_MAM	K_JJA	K_SON
all	1.006	1.004	1.003	1.017	0.994	1.013	1.002	0.979	0.994	0.888	0.878	0.931	0.976	0.988	1.003	1.002	0.893
C	0.839	0.824	0.789	0.764	0.766	0.766	0.743	0.724	0.705	0.699	0.766	0.802	0.766	0.824	0.766	0.740	0.722
P	1.145	1.143	1.050	1.058	1.049	1.047	1.041	1.031	1.040	1.020	0.985	1.098	1.058	1.124	1.050	1.041	1.016
S	1.496	1.391	1.331	1.221	1.221	1.241	1.268	1.282	1.280	1.239	1.290	1.443	1.312	1.443	1.273	1.251	1.279

Neighbours comparison, discriminating into categories according to **circulation types, C and A categories**

- Classification of CHMI (CZ), about 30 types
- Two groups: cyclone or anticyclone situations (69%, 31% resp.)
- one common classification for all the stations
- only discriminating temperatures according to days with **cyclone situations** lead to higher correlations between candidate and neighbour stations and lower RMSE

Neighbours comparison, discriminating into categories according to **circulation types, C and A categories**

- Spearman rank correlation coefficient, median for 69 pairs of candidate and neighbour stations. Category: all – all days, A anticyclone situations, C – cyclone situations

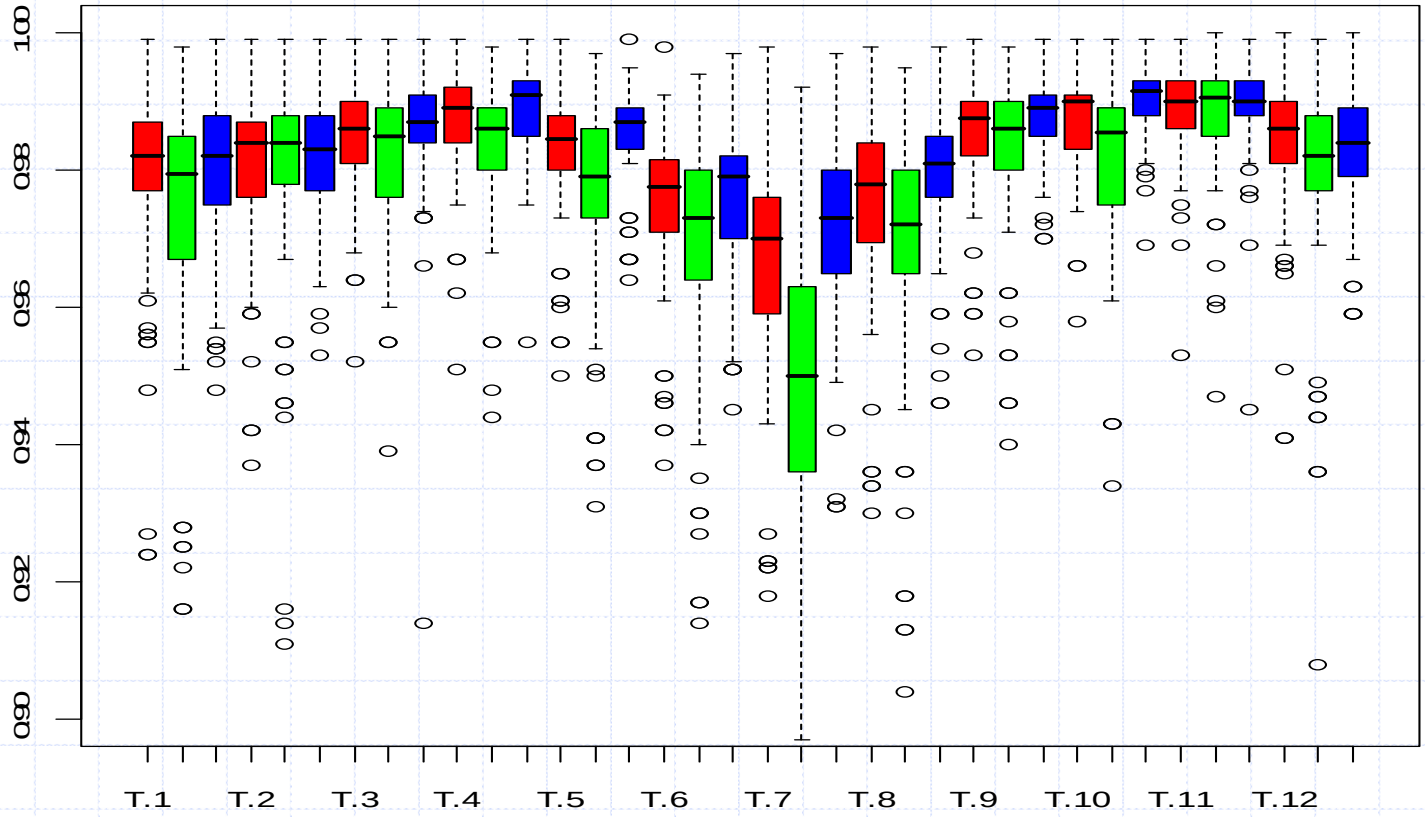
Category	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	K_YEAI	K_DJF	K_MAM	K_JJA	K_SON
all	0.982	0.984	0.986	0.989	0.985	0.978	0.970	0.978	0.988	0.990	0.990	0.986	0.984	0.984	0.986	0.975	0.990
A	0.980	0.984	0.985	0.986	0.979	0.973	0.950	0.972	0.986	0.986	0.991	0.982	0.981	0.983	0.983	0.968	0.988
C	0.982	0.983	0.987	0.991	0.987	0.979	0.973	0.981	0.989	0.992	0.990	0.984	0.985	0.983	0.988	0.978	0.990

- RMSE, median for 69 pairs of candidate and neighbour stations. Category: all – all days, A anticyclone situations, C – cyclone situations

Category	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	K_YEAI	K_DJF	K_MAM	K_JJA	K_SON
all	1.006	1.004	1.003	1.017	0.994	1.013	1.002	0.979	0.994	0.888	0.878	0.931	0.976	0.988	1.003	1.002	0.893
A	1.169	1.169	1.153	1.132	1.181	1.239	1.217	1.164	1.117	1.003	1.003	1.140	1.153	1.161	1.158	1.201	1.010
C	0.919	0.931	0.917	0.958	0.968	0.957	0.988	0.904	0.914	0.856	0.832	0.867	0.914	0.908	0.960	0.937	0.856

Neighbours comparison, discriminating into categories according to **weather types, C and A categories**

- Spearman rank correlation coefficient, three categories (all – red, A – green, C – blue), 12 months (Jan. to Dec.)



Further investigation

- First results for including co-variables (during correction)
- Circulation patterns do not contribute to better understanding of neighbors differences
- Weather types, based on PCA – is promising – will be analyzed
- other met. elements (air pressure, ...)

Conclusions

- Preliminary results of the daily correction methods comparison were presented
- Some of the methods / applied settings can worsen the results (especially for precipitation!)
- Benchmark dataset testing for evaluation of the methods ... please contribute to the COST Action ESO601 effort!

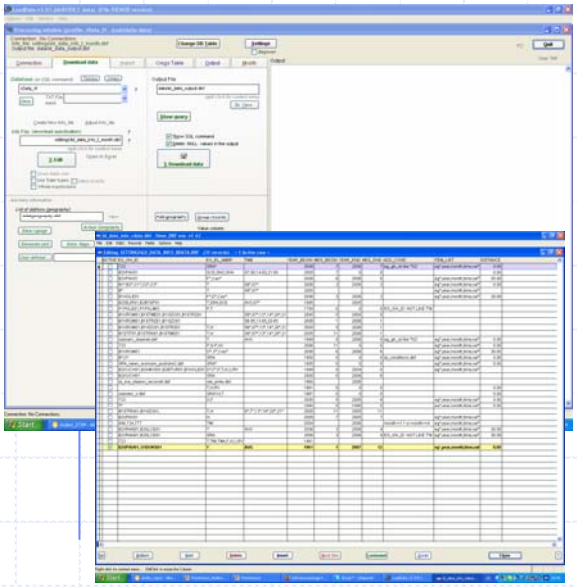
(www.homogenization.org)

Thank you for your attention ...

http://www.climahom.eu

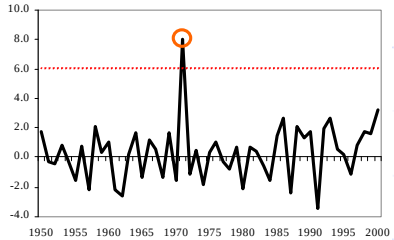
Download data from database (e.g. Oracle)

(LoadData)



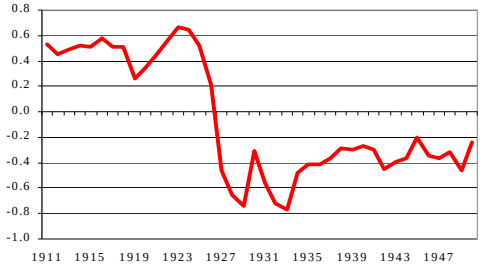
Quality control

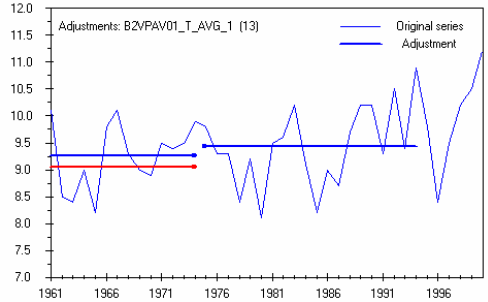
(ProClimDB)



Homogenization

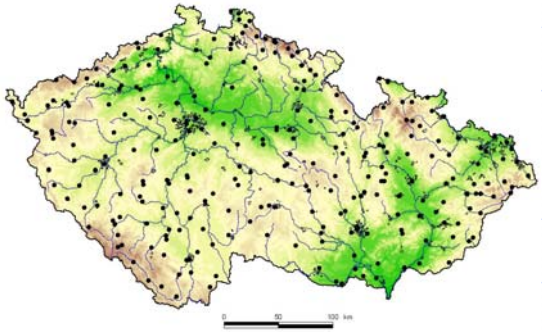
(ProClimDB/AnClim)

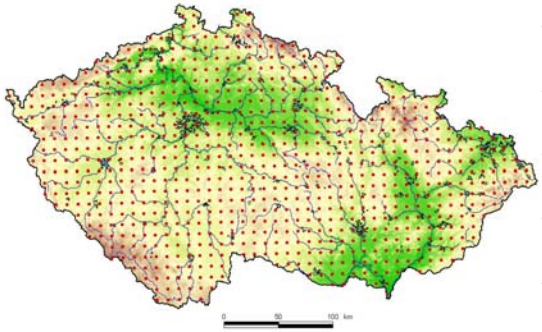




„Technical“ series and grid points calculation

(ProClimDB)





Statistical analysis

...

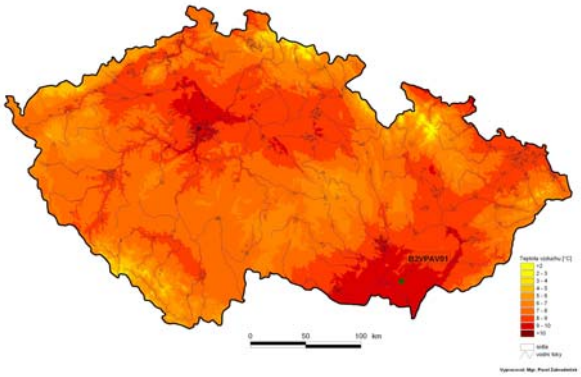
Validation of RCM outputs

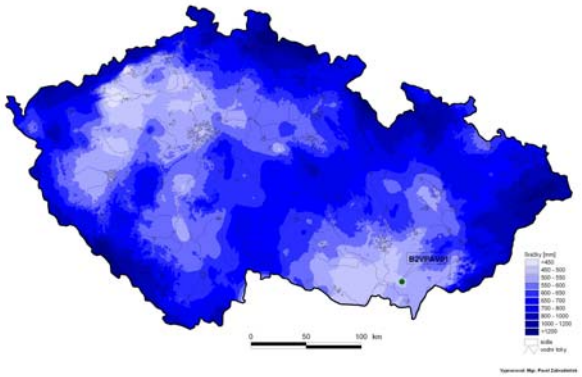
Extreme value analysis

Correction of RCM outputs

Spatial analysis

(connection ProClimDB - ArcView)





Further tools:

(connection ProClimDB - R)