Uncertainty of Climatol's adjustment algorithm for daily time series of additive climate variables

Oleg Skrynyk[1,2], Enric Aguilar[1], José A. Guijarro[3], Sergiy Bubin[4]

[1] *Center for Climate Change (C3), Universitat Rovira i Virgili, Tarragona, Spain*
[2] *Ukrainian Hydrometeorological Institute, Kyiv, Ukraine*
[3] *AEMET (State Meteorological Agency), Palma de Mallorca office, Spain*
[4] *Department of Physics, Nazarbayev University, Nur-Sultan, Kazakhstan*

EGU General Assembly 2020
4-8 May 2020

Introduction

Motivation

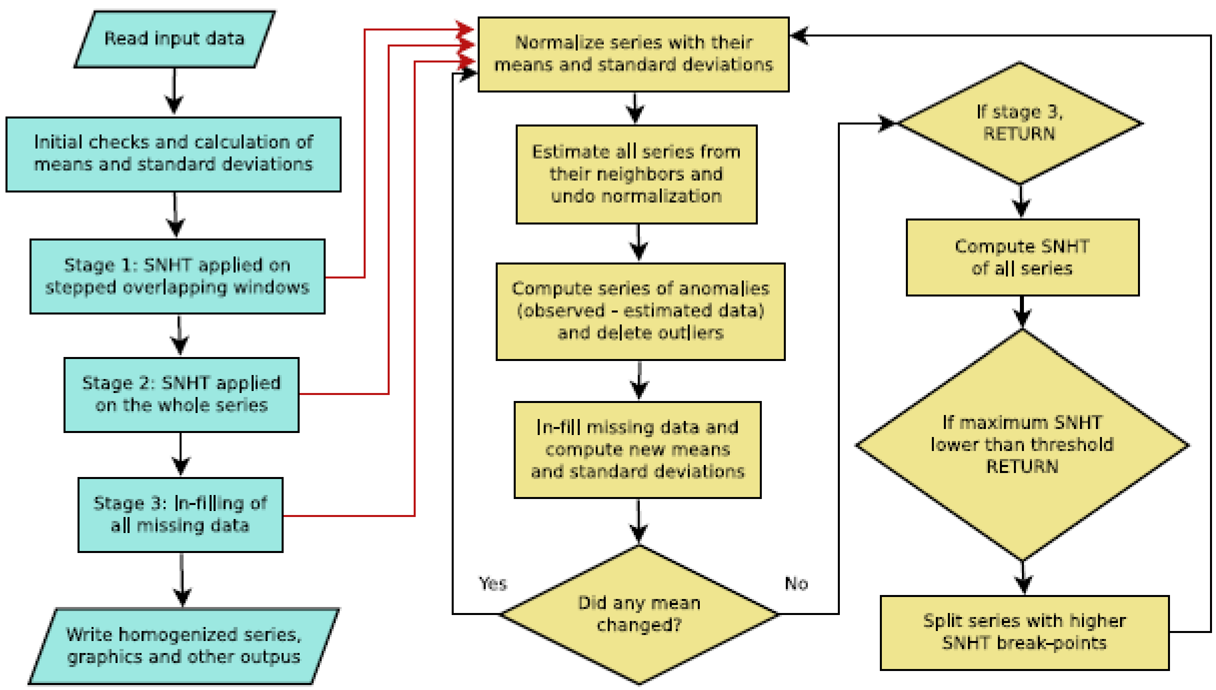
- Homogenization increases consistency of climatological data. However, it is not clear which residual errors could still be present in homogenized/adjusted data.
- The problem is particularly important when dealing with daily time series as they are the basis for many modern climatological studies (e.g. monitoring, detection and attribution of changes in climate extremes).

Objective

- To evaluate the uncertainty associated to the adjustment of daily minimum, TN, and maximum, TX, air temperature series using Climatol (Guijarro, 2018). We restrict our studies to the case a perfect detection in order to focus on Climatol’s adjustment algorithm.

Methods and data

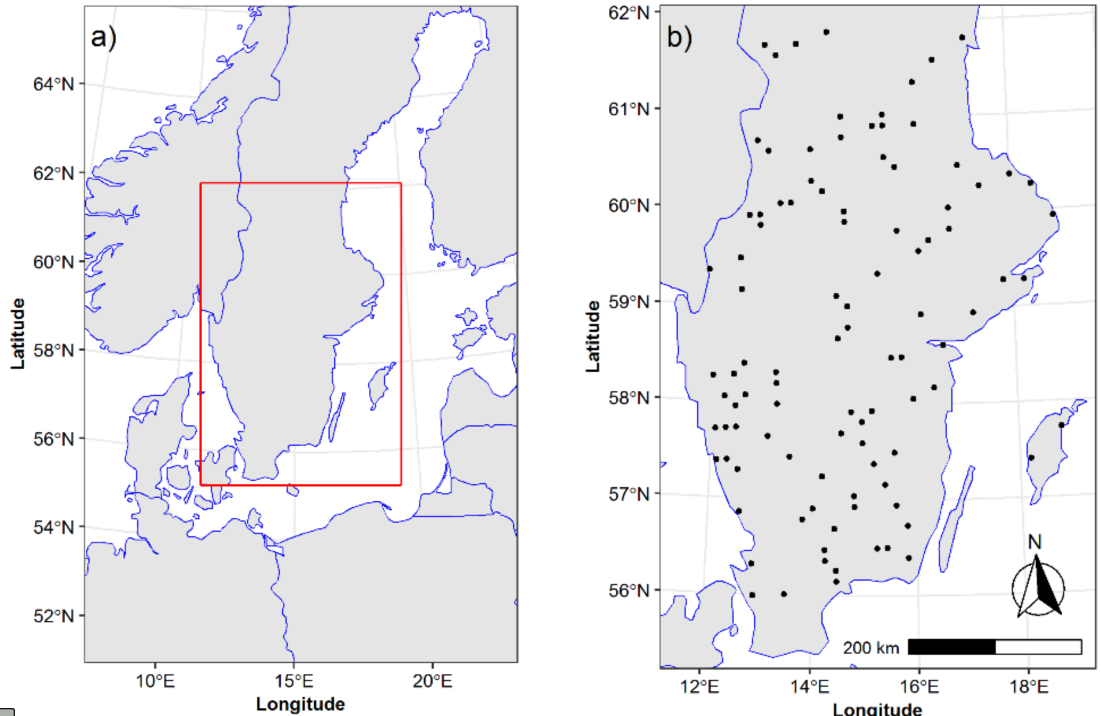
- **The Climatol homogenization software (Climatol 3.1.1)**
  - R package (https://cran.r-project.org/web/packages/climatol/index.html)
  - Detection is based on the standard normalized homogeneity test (SNHT) (Alexandersson, 1986, doi:10.1002/joc.3370060607)
  - Adjustment terms are computed from the interpolation with orthogonal (type II) linear regression model



Flow-chart of the Climatol operation (from the Climatol guide)

- **The benchmark TN and TX data sets**

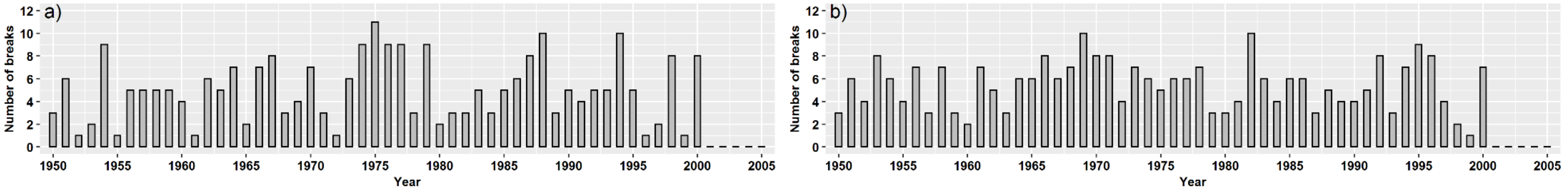
  - Created in the scope of the INDECIS project (www.indecis.eu).
  - Consist of **clean** data (100 series over the period of 1950-2005) extracted from an output of KNMI RACMO v2 (driven by MOHC-HadGEM2-ES), and **inhomogeneous** data, created by introducing realistic breaks and errors.
  - Cover the area of Southern Sweden.



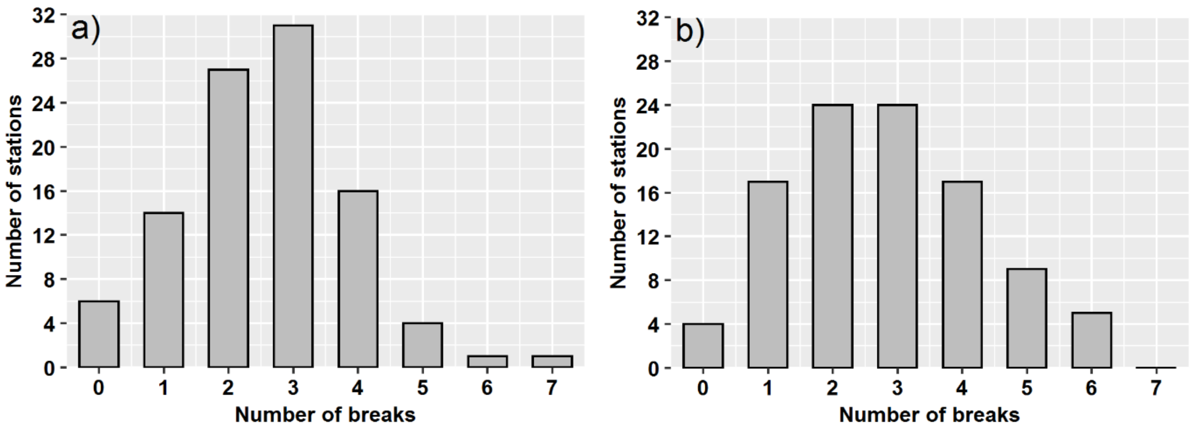
(a) The domain of the Southern Sweden (inside the red frame)

(b) Locations of the ‘stations’ (the subset of the RACMO grid points, shown as black dots)

- **Peculiarities of station signals introduced into clean data**
  - The introduction of station signals was done by simulating relocations
  - The total numbers of break points inserted into TN and TX clean time series are **258** and **280** respectively

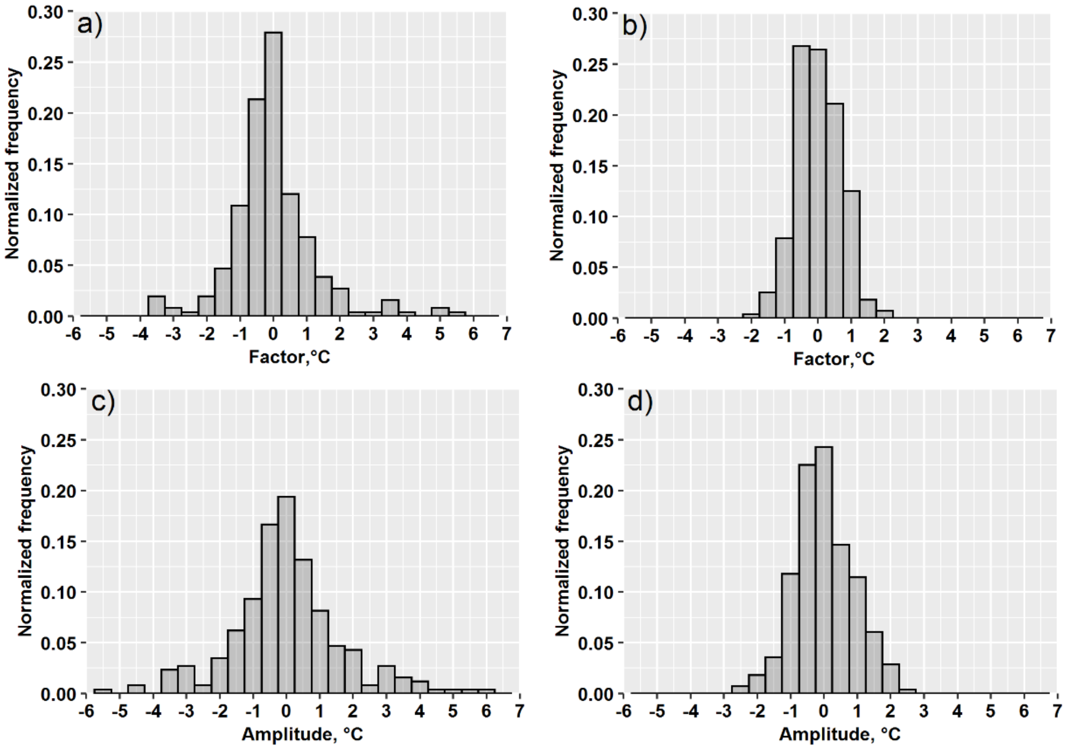


Number of break points per year introduced to clean (a) TN and (b) TX air temperature time series



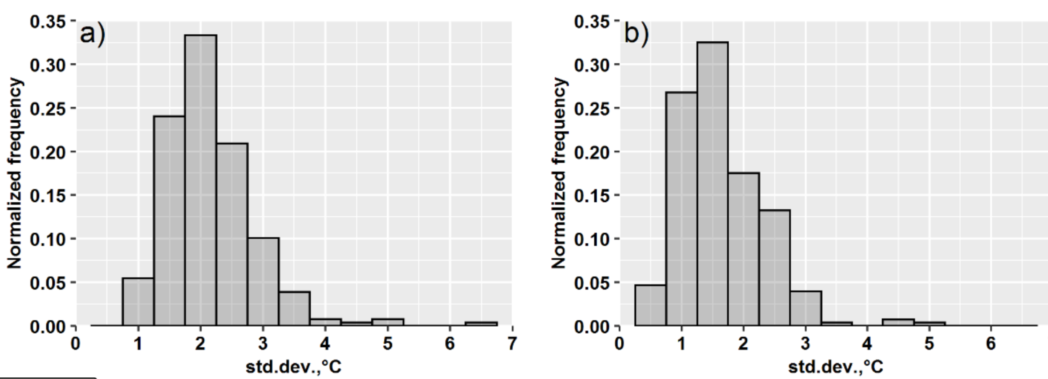
Distribution of the number of stations/time series with respect to the number of break points in one time series: (a) TN, (b) TX

• Peculiarities of station signals introduced into clean data



Histograms of the factors (a, b) and amplitudes (c, d) of the shifts at the break points in TN (a, c) and TX (b, d) raw data sets (differences between factors and amplitudes are made following HOMER's notations).

- The frequency/count was normalized by the total number of the breaks.
- The factors/amplitudes were estimated by averaging respective segments of the daily error time series

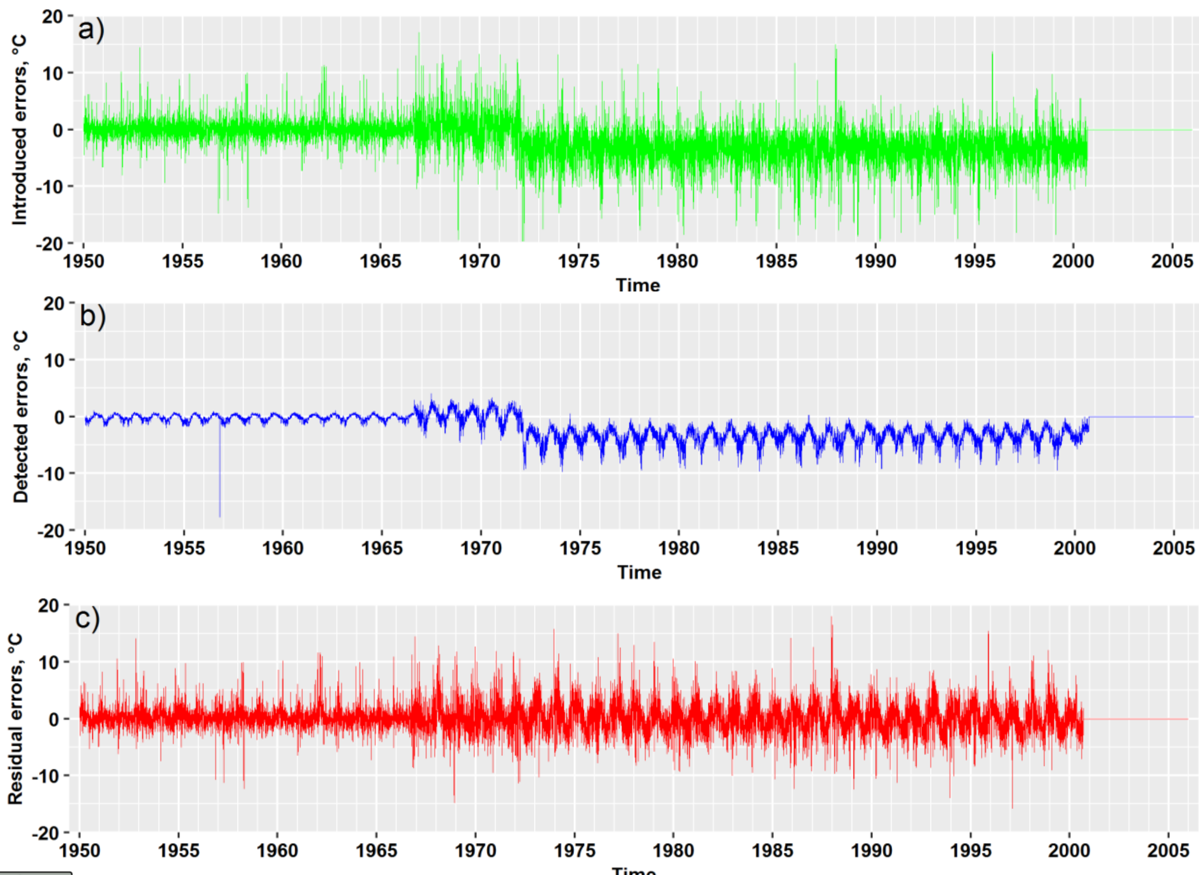


Histograms of standard deviations of the introduced daily errors at the inhomogeneity segments: (a) TN, (b) TX.

• Methodology applied to evaluate uncertainty of adjustment

- **I**nhomogeneous, **H**omogenized (by means of Climatol) and **C**lean data sets: $\boldsymbol{X}^I, \boldsymbol{X}^H$ and $\boldsymbol{X}^C$
- Each is a collection of time series: $\boldsymbol{X} = \{x_{ij}\}, i = 1, \dots, M\ j = 1, \dots, N,$ where M is number of stations, N is number of time steps/days
- **R**eal/introduced, **D**etected and **H**omogenization/residual error time series:

$$\boldsymbol{E}^R = \boldsymbol{X}^I - \boldsymbol{X}^C, \boldsymbol{E}^D = \boldsymbol{X}^I - \boldsymbol{X}^H, \boldsymbol{E}^H = \boldsymbol{X}^H - \boldsymbol{X}^C = \boldsymbol{E}^R - \boldsymbol{E}^D$$



Examples of time series of errors: $\boldsymbol{E}_k^R$ (a), $\boldsymbol{E}_k^D$ (b) and $\boldsymbol{E}_k^H$ (c) calculated for k-th ($i = k$) station

● Methodology applied to evaluate uncertainty of adjustment

- **'Uncertainty'** is 'any departure from the unachievable ideal' (Walker et al., 2003, doi:10.1076/iaij.4.1.5.16466). The **uncertainty** of homogenization adjustment is any departure of the model prediction, $\boldsymbol{X}^H$, from the etalon/reference result, $\boldsymbol{X}^C$

- To evaluate uncertainty usually means to define a **width** of the **error distribution**, which is created by considering the whole credible range of every uncertain input and parameter of a predicting model (adjustment algorithm in our case)

- **Complex approach** was used, quantifying uncertainty on different levels of detail and time resolution:
  - on daily scale, by calculating parameters of error distributions for each day of some period
  - on daily scales, through calculation of a set of metrics
  - on yearly scale, through calculation of a set of metrics for some climate extreme indices

• Methodology applied to evaluate uncertainty of adjustment

○ Metrics for the adjustment evaluation on the daily time scale

- $RMSE$ (the **main** metric to evaluate overall adjustment uncertainty; random error evaluation):

$$RMSE_i = \sqrt{\frac{\sum_{j=1}^{N_i}\left(x_{ij}^H - x_{ij}^C\right)^2}{N_i}}, \qquad i = 1, \dots, M$$

N_i $(N_i < N)$ is a number of pairs $\left(x_{ij}^C,\ x_{ij}^H\right)$ in an adjusted segment/segments

- B (bias; systematic error evaluation): $B_i = \frac{1}{N_i}\sum_{j=1}^{N_i}\left(x_{ij}^H - x_{ij}^C\right)$

- $FOEX$ (factor of excedance; systematic error evaluation): $FOEX_i = \left(\frac{N_{\left(x_{ij}^H > x_{ij}^C\right)}}{N_i} - 0.5\right)100$

$N_{\left(x_{ij}^H > x_{ij}^C\right)}$ is a number of pairs $\left(x_{ij}^C,\ x_{ij}^H\right)$ in an adjusted segment/segments when $x_{ij}^H > x_{ij}^C$

- $POD05$ and $POD2$ (percentage of days within ±0.5 (±2) ºC margin; random error evaluation):

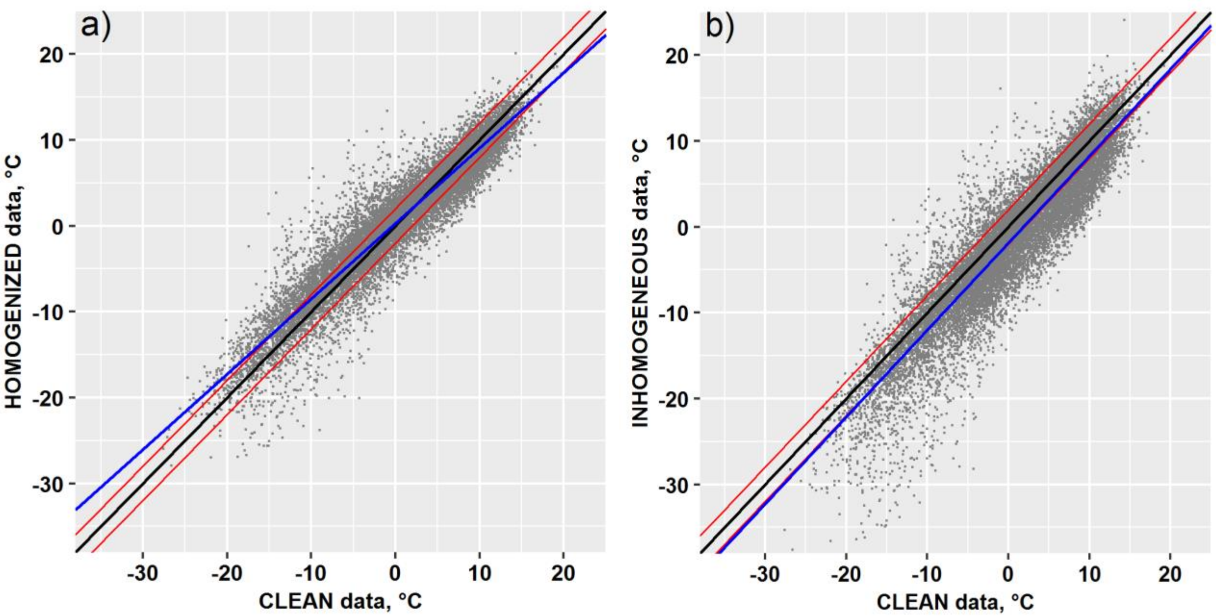
$$POD05_i = \frac{N_{\left|x_{ij}^H - x_{ij}^C\right| < 0.5}}{N_i}100, \quad POD2_i = \frac{N_{\left|x_{ij}^H - x_{ij}^C\right| < 2}}{N_i}100$$

$N_{\left|x_{ij}^H - x_{ij}^C\right| \leq a}$ $(\mathrm{a} = 0{,}5 \text{ or } 2)$ is a number of pairs $\left(x_{ij}^C,\ x_{ij}^H\right)$ in adjusted segments when $\left|x_{ij}^H - x_{ij}^C\right| \leq a$

- $SlopeD$ (difference in slopes ; systematic error evaluation): $SlopeD_i = b_i - 1$

b_i is a slope of a linear regression model $\boldsymbol{X}_i^H = a_i + b_i\boldsymbol{X}_i^C$ (see explanation on the next slide)

● Methodology applied to evaluate uncertainty of adjustment



Example of scatter diagrams. Homogenized $\boldsymbol{X}_k^H$ (a) and raw $\boldsymbol{X}_k^I$ (b) daily data are built against respective clean values $\boldsymbol{X}_k^C$.

- Blue lines

$$\boldsymbol{X}_k^H = a_k + b_k \boldsymbol{X}_k^C \text{ (a)}$$
$$\boldsymbol{X}_k^I = a_k + b_k \boldsymbol{X}_k^C \text{ (b)}$$

are used to calculate $SlopeD$

- $SlopeD$ is introduced to evaluate over/under-estimation of seasonal amplitude
- Red lines is related to $POD2$
- Black line is 'true' predictions

○ Quantifying the discrepancies between homogenized and clean data on the yearly scale

- Extreme indices: (**TN**) FD, TR, TN10p, TN90p; (**TX**) ID, SU, TX10p, TX90p (the thresholds for FD, TR, ID, SU were slightly shifted due to peculiarities of the Southern Sweden climate)
- Metrics: B (systematic error evaluation), $RMSE$ (random error evaluation), TrD (differences in the indices linear trends)

• Methodology applied to evaluate uncertainty of adjustment

- Potential sources of the uncertainty considered
  - input data: statistical properties of introduced station signals $\boldsymbol{E}^R$. The residual errors $\boldsymbol{E}^H$ should depend on the introduced errors $\boldsymbol{E}^R$
  - station density (correlation between time series)
  - length of the time series

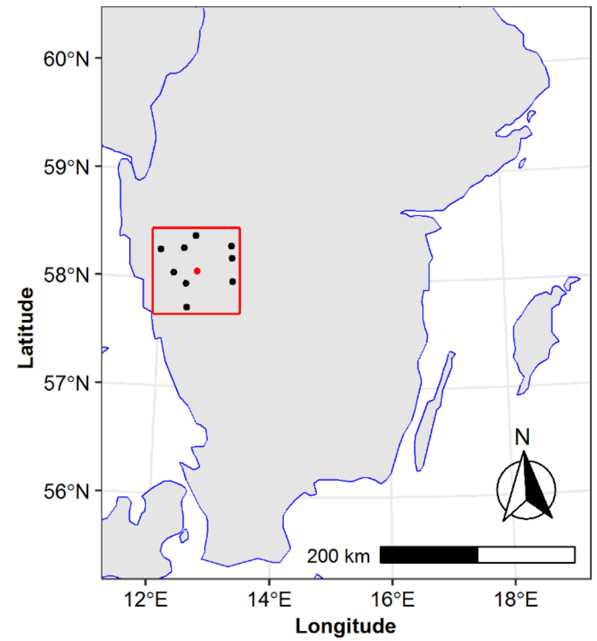
- Samples of Climatol’s adjustment outputs were created based on replacements and random permutations of time series in $\boldsymbol{E}^R$

- Uncertainty of Climatol’s adjustment algorithm was evaluated by means of several case studies (series of numerical experiments), which complexity was gradually increased

Results and Discussion

• Case study #1

- 10 stations, period 1971-1980 (10 years)
- Time series of 9 stations (black dots in Figure) left clean
- Time series of 10-th station (red dot in Figure) is corrupted: 1 break point on 01.01.1976
- Inhomogeneous segment 1971-1975 was created by adding errors (extracted from $\boldsymbol{E}^R$) to the clean data
- Total of 185/193 (TN/TX) 5-year segments with different statistical properties were found in the error time series $\boldsymbol{E}^R$
- The identified error segments were shifted to 1971-1975 and added to respective clean data of the 10-th time series
- Consequently, 185/193 Climatol adjustment exercises were performed, creating the corresponding number of homogenized versions of the 10-th series (a sample to evaluate error distribution)
- The same disturbed period along with unchanged system of reference series allows to obtain statistically reliable and justified evaluation of the residual errors

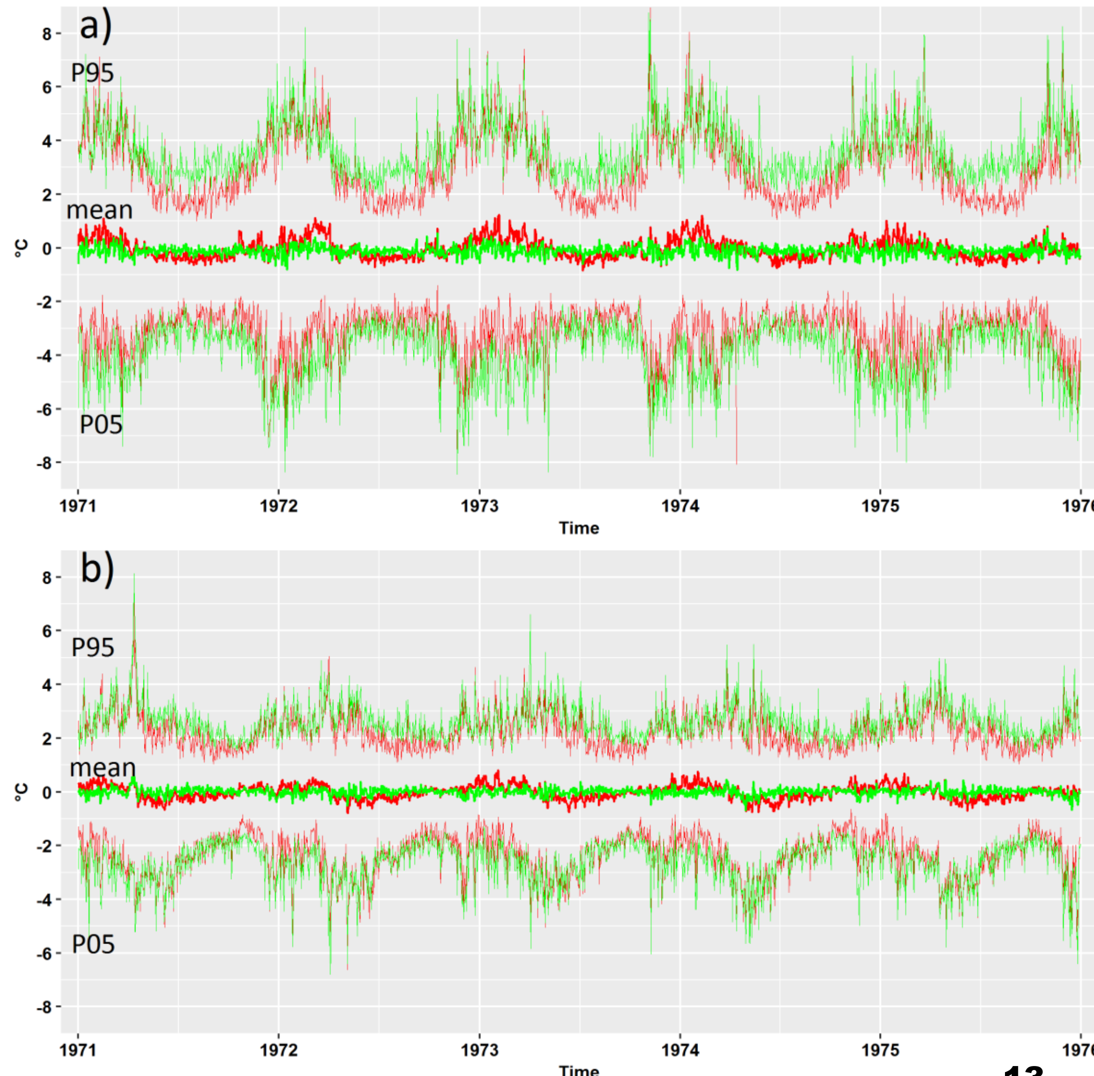


Meteorological stations in set #1 of numerical experiments

● Case study #1

○ Distribution of the residual error on day-to-day basis

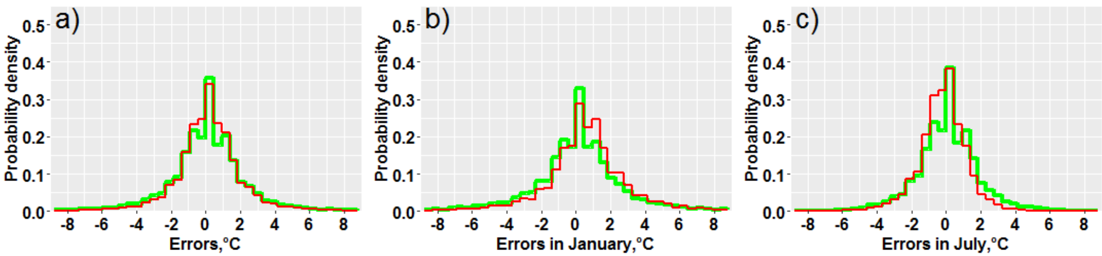
- In general, dispersion of the residual errors is less than dispersion of the introduced ones, meaning less uncertainty in Climatol output comparing to raw data
- Significant non-stationarity of the error distribution in 1-st and 2-nd moments
- Uncertainty of Climatol's adjustment is less in summer months
- For TN data: in winter negative errors are adjusted slightly better compared to positive errors
- Non-stationarity of means of the residual errors is related to underestimation of the seasonal cycle amplitude



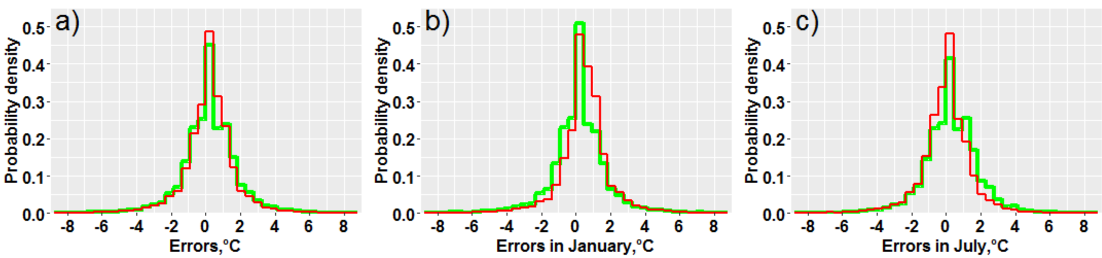
Mean, P05 and P95 of empirical error distributions, evaluated for every day of the disturbed segment: (a) TN, (b) TX. **Green** color refers to the introduced errors, **red** is for the residual errors.

- **Case study #1**

  ○ Temporary averaged distribution of the residual error



TN. Empirical distributions of errors, averaged over (a) the whole period, (b) January months, (c) July months
Green color refers to the introduced errors, **red** is for the residual errors



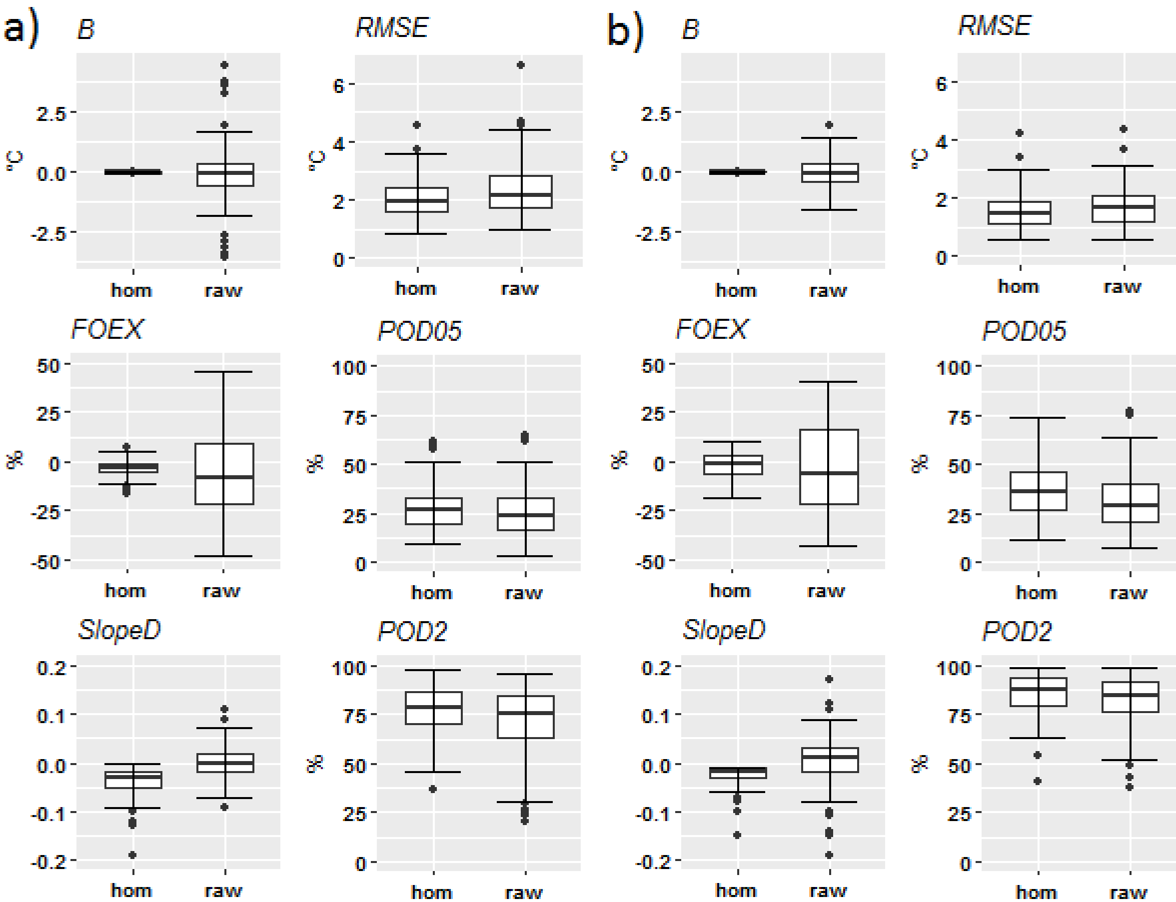
TX. Empirical distributions of errors, averaged over (a) the whole period, (b) January months, (c) July months

Parameters of averaged empirical distributions of errors in TN/TX time series: introduced $\boldsymbol{E}^R$ and residual $\boldsymbol{E}^H$ (all in ^oC)

		Year		January		July	
		$\boldsymbol{E}^R$	$\boldsymbol{E}^H$	$\boldsymbol{E}^R$	$\boldsymbol{E}^H$	$\boldsymbol{E}^R$	$\boldsymbol{E}^H$
TN	Mean	-0.11	-0.03	-0.08	0.40	-0.13	-0.41
	st.d.	2.53	2.15	2.97	2.56	1.85	1.39
	P05	-4.00	-3.20	-4.90	-3.60	-3.20	-2.80
	P95	3.70	3.20	4.60	4.50	2.90	1.70
	P95-P05	7.70	6.40	9.50	8.10	6.10	4.50
TX	Mean	-0.00	-0.02	-0.03	0.28	0.04	-0.22
	st.d.	1.84	1.64	1.78	1.58	1.67	1.48
	P05	-2.70	-2.50	-2.70	-2.00	-2.50	-2.60
	P95	2.60	2.30	2.60	2.60	2.50	1.90
	P95-P05	5.30	4.80	5.30	4.60	5.00	4.50

• Case study #1

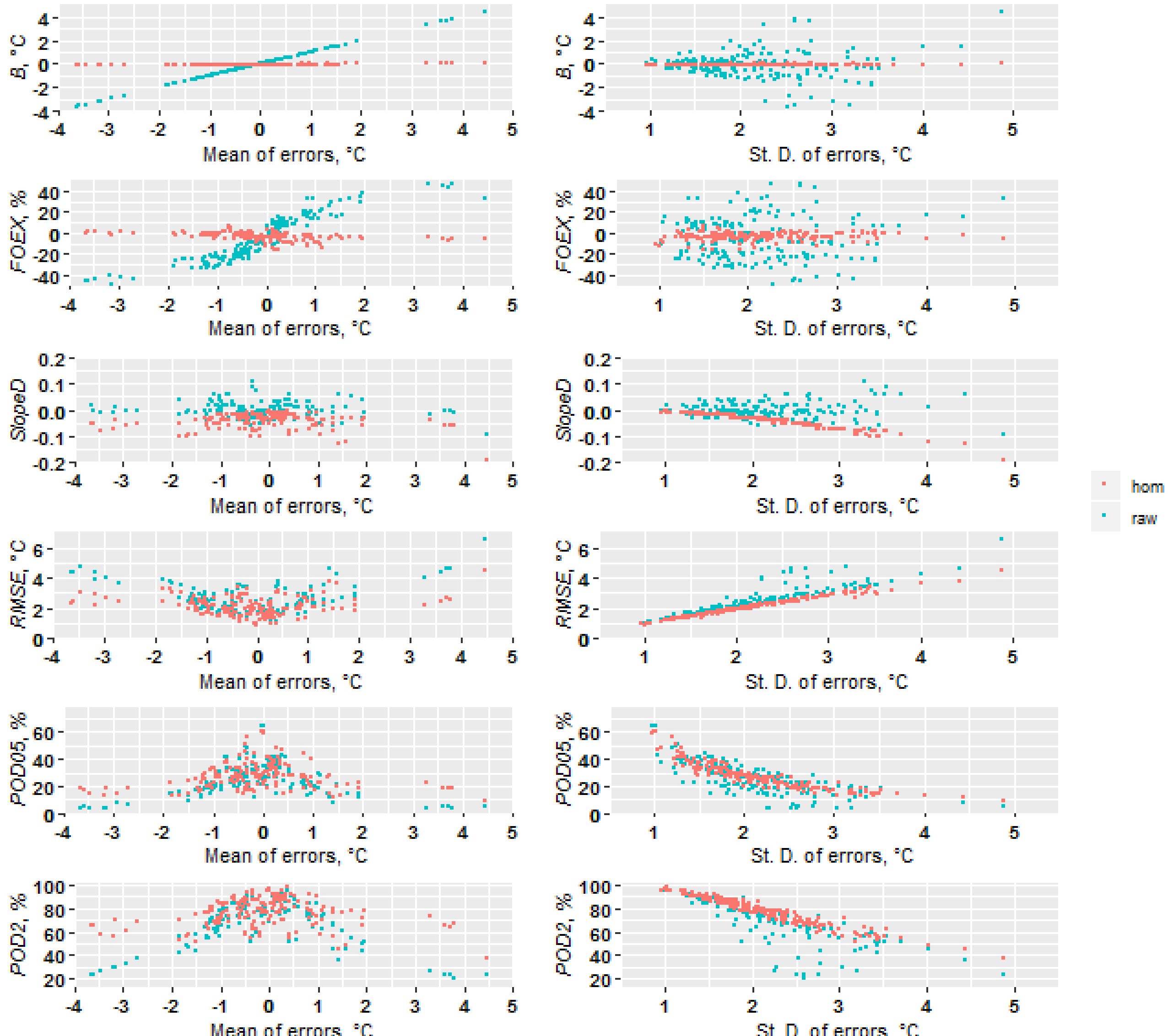
○ Evaluation of homogenization adjustment through the metrics



- Clearly seen 'added value' of Climatol's adjustment
- Mean value of $RMSE$ represents an overall/averaged estimate of the homogenization adjustment uncertainty
- Other metrics give general estimation of Climatol's performance

Boxplots of the metrics, calculated in Case study #1: (a) TN, (b) TX

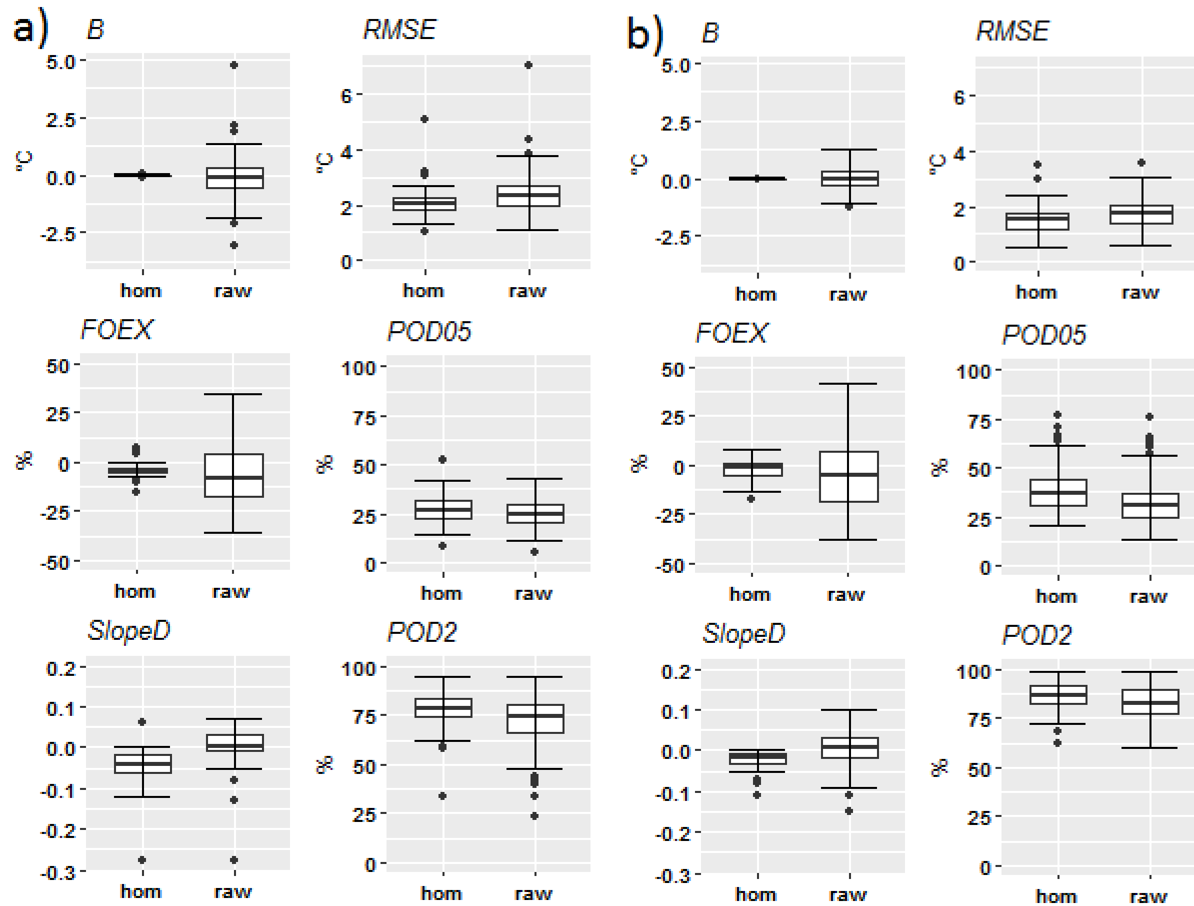
● Case study #1



Relationships between the TN metrics and the main statistical properties of the introduced errors: means/shift amplitudes (left column) and standard deviations (right column). Similar results were obtained for TX

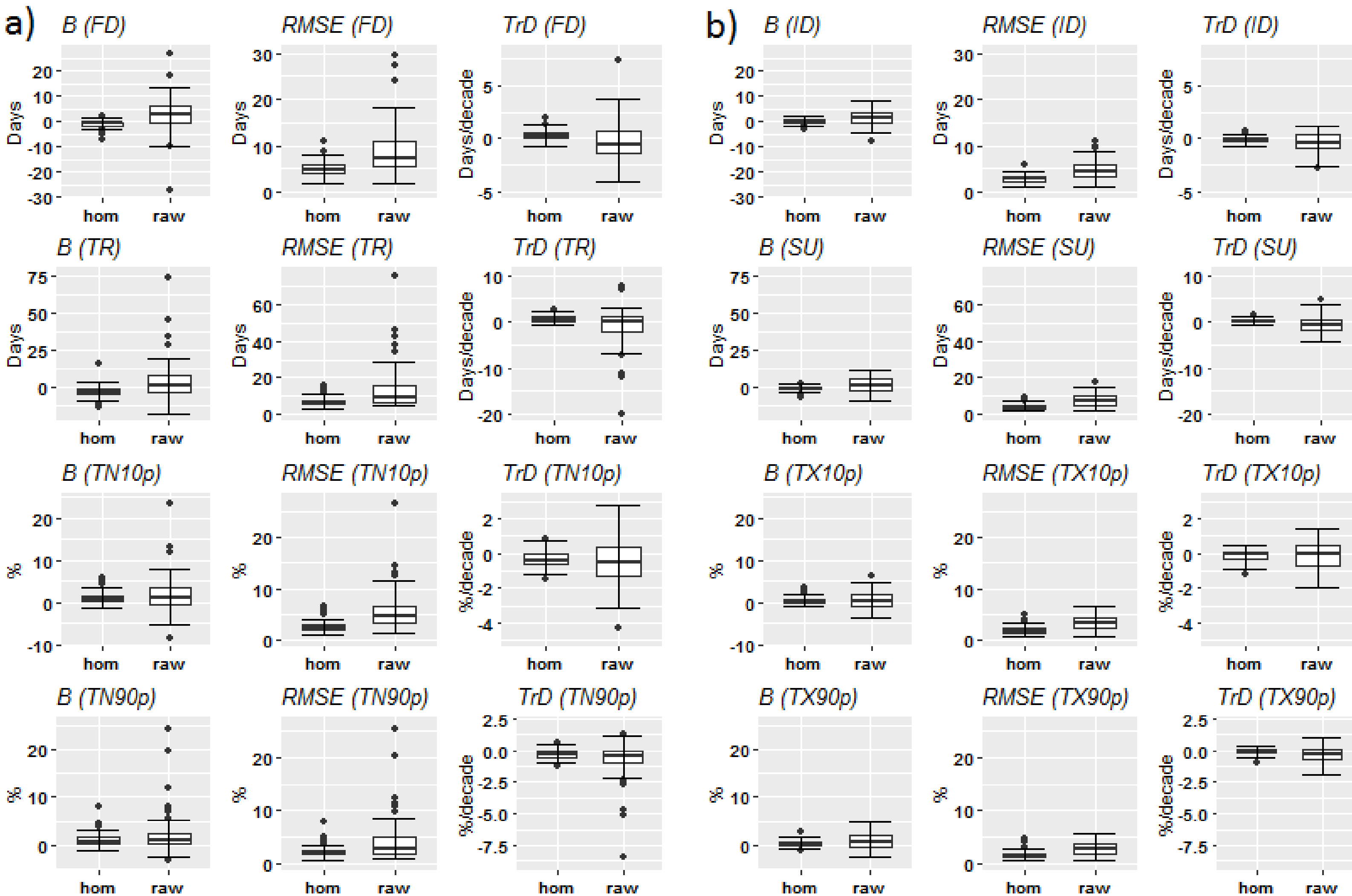
● Case study #2

- The same 10 stations as in Case study #1, period 1950-2005 (56 years)
- Time series of the 9 stations left clean. Time series of the 10-th station is corrupted: multiple break points are allowed with their arbitrary positions
- 94/96 (TN/TX) different variants/realizations of the 10-th corrupted series were created by adding the station signals from $\boldsymbol{E}^{R}$ to the clean data
- Consequently, 94/96 Climatol adjustment exercises were performed creating the corresponding number of homogenized versions of the 10-th series (a sample to evaluate residual error)



Boxplots of the metrics, calculated in Case study #2 (daily scale): (a) TN, (b) TX

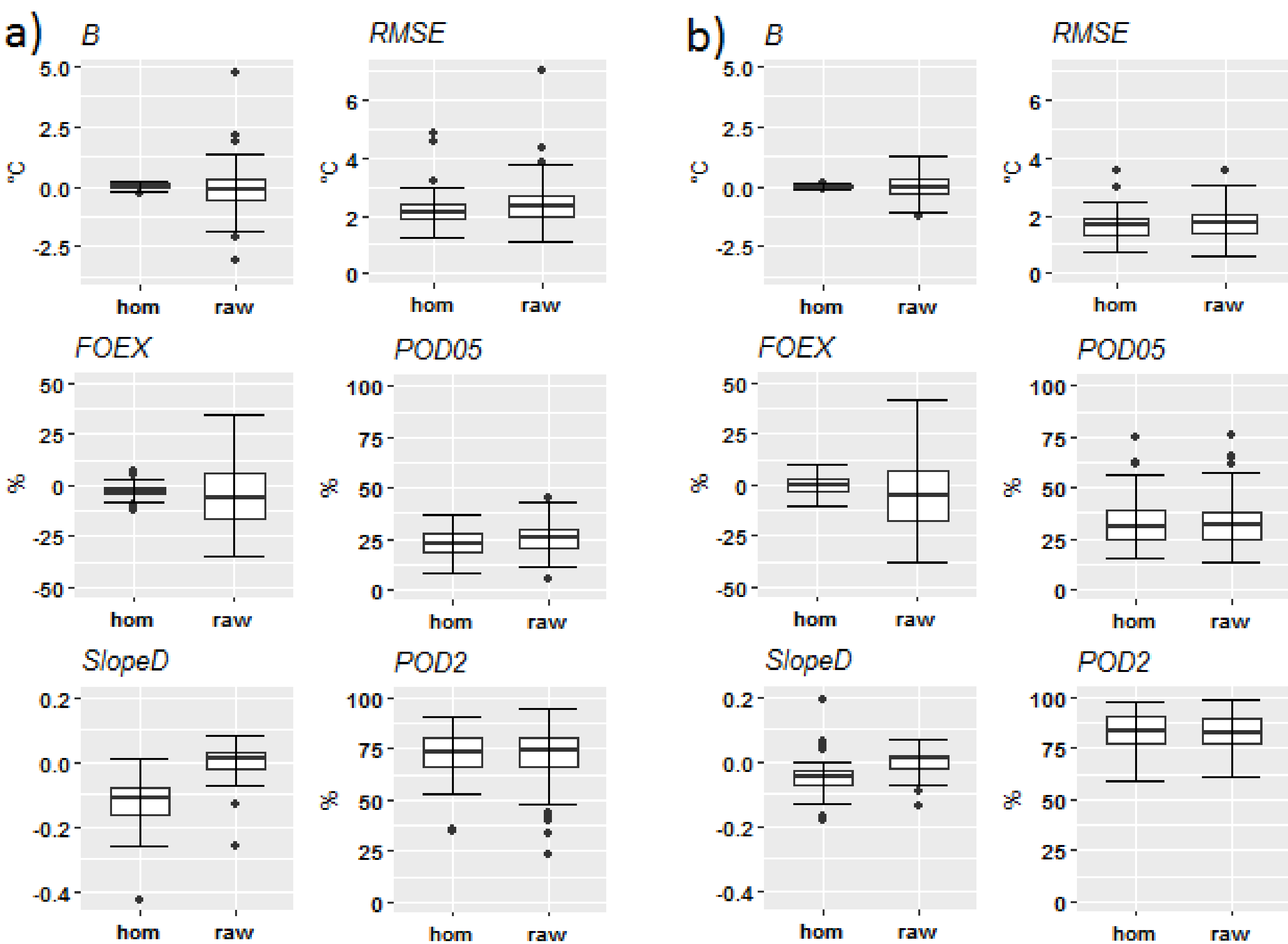
- **Case study #2**



Boxplots of the metrics, calculated based on the yearly series of the climate extremes indices in Case study #2: (a) TN, (b) TX

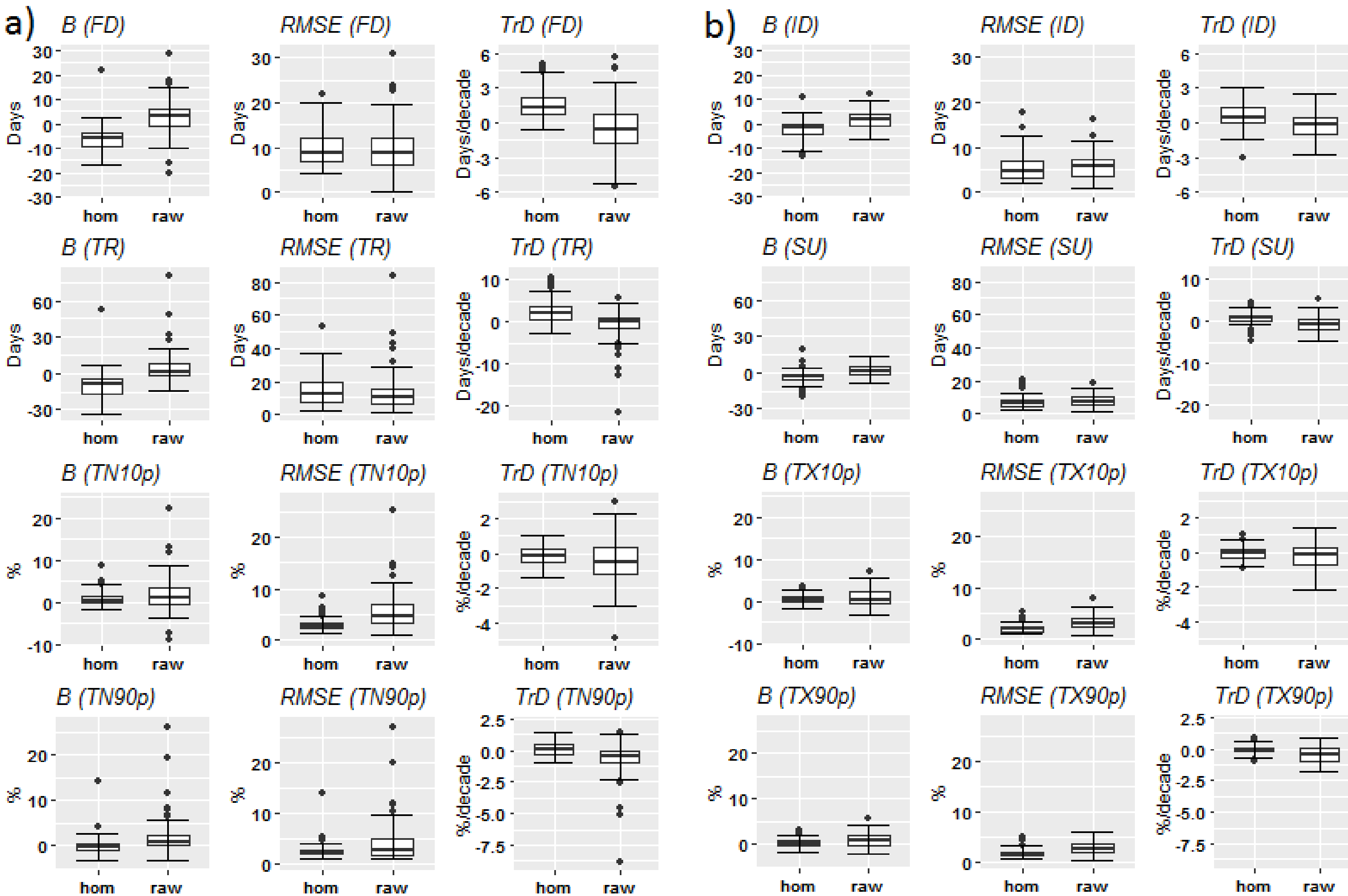
Case study #3

- The original case, initially created in the frame of the INDECIS project: 100 stations, period 1950-2005
- Compared to the previous case, every station signal of $\boldsymbol{E}^R$ will be removed in different local conditions



Boxplots of the metrics, calculated in Case study #3 (daily scale): (a) TN, (b) TX

- **Case study #3**



Boxplots of the metrics calculated based on the yearly series of the climate extremes indices in Case study #3: (a) TN, (b) TX

Case study #3

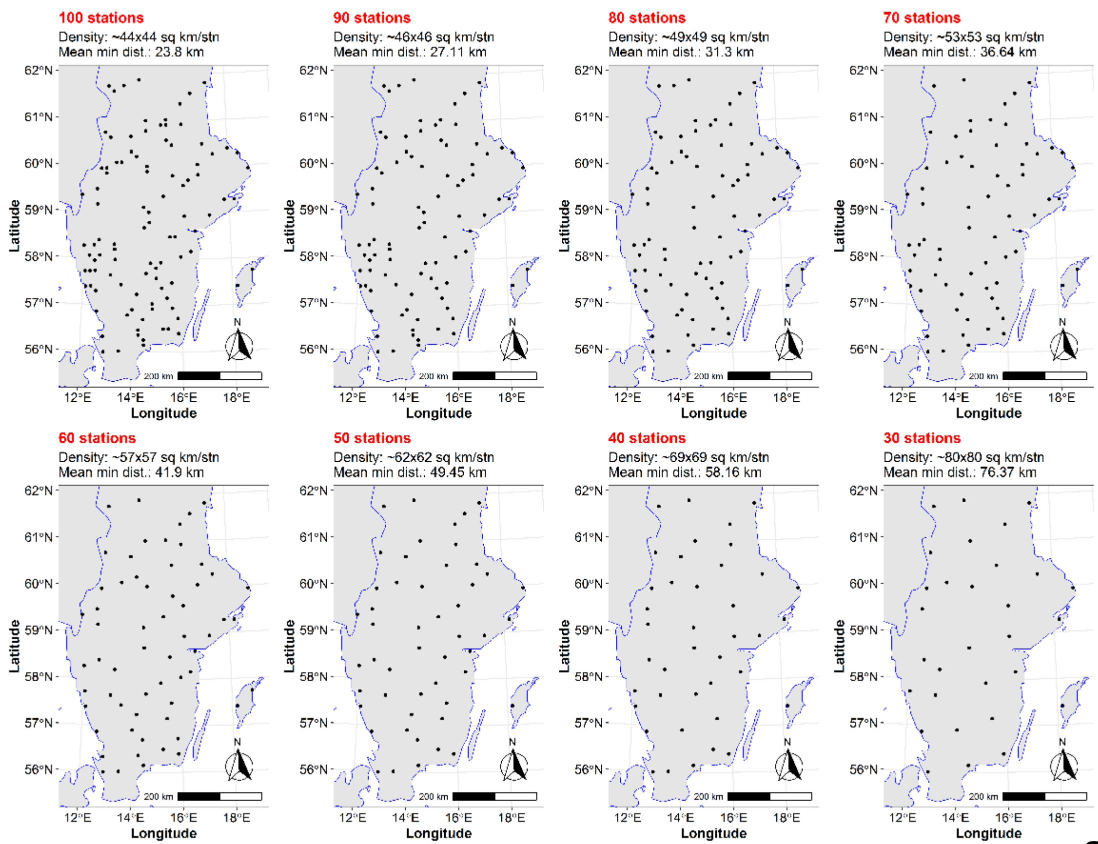
- in order to study how local peculiarities (e.g. subset of neighboring stations chosen to create a composite reference series) influence removing of station signals we performed calculations with 100 random permutations (without repetition) of the introduced error time series in $\boldsymbol{E}^R$

Parameter variability of the metrics empirical distributions due to 100 random permutations of the station signals in $\boldsymbol{E}^R$. Case study #3, daily scale

			B °C		$FOEX$ %		$SlopeD$		$RMSE$ °C		$POD05$ %		$POD2$ %	
			$\boldsymbol{X}^H$	$\boldsymbol{X}^I$	$\boldsymbol{X}^H$	$\boldsymbol{X}^I$	$\boldsymbol{X}^H$	$\boldsymbol{X}^I$	$\boldsymbol{X}^H$	$\boldsymbol{X}^I$	$\boldsymbol{X}^H$	$\boldsymbol{X}^I$	$\boldsymbol{X}^H$	$\boldsymbol{X}^I$
TN	mean	max	0.09	-0.11	-1	-6	-0.09	0.01	2.25	2.44	24	25	75	71
		aver.	0.06	-0.11	-2	-6	-0.10	~0.00	2.18	2.44	23	25	73	71
		min	0.04	-0.11	-3	-6	-0.12	0.00	2.13	2.44	22	25	71	71
	IQR	max	0.13	0.86	4	22	0.11	0.05	0.64	0.72	10	10	16	15
		aver.	0.09	0.86	3	22	0.08	0.04	0.50	0.72	8	10	13	15
		min	0.06	0.86	2	22	0.05	0.03	0.39	0.72	7	10	10	15
TX	mean	max	0.05	-0.02	2	-5	-0.04	~0.00	1.65	1.75	35	33	86	83
		aver.	0.03	-0.02	1	-5	-0.05	0.00	1.61	1.75	33	33	85	83
		min	0.01	-0.02	0	-5	-0.06	~-0.0	1.59	1.75	32	33	84	83
	IQR	max	0.10	0.64	7	25	0.07	0.06	0.61	0.65	15	13	13	12
		aver.	0.07	0.64	5	25	0.05	0.04	0.54	0.65	12	13	10	12
		min	0.05	0.64	4	25	0.03	0.03	0.49	0.65	9	13	9	12

● Case study #4

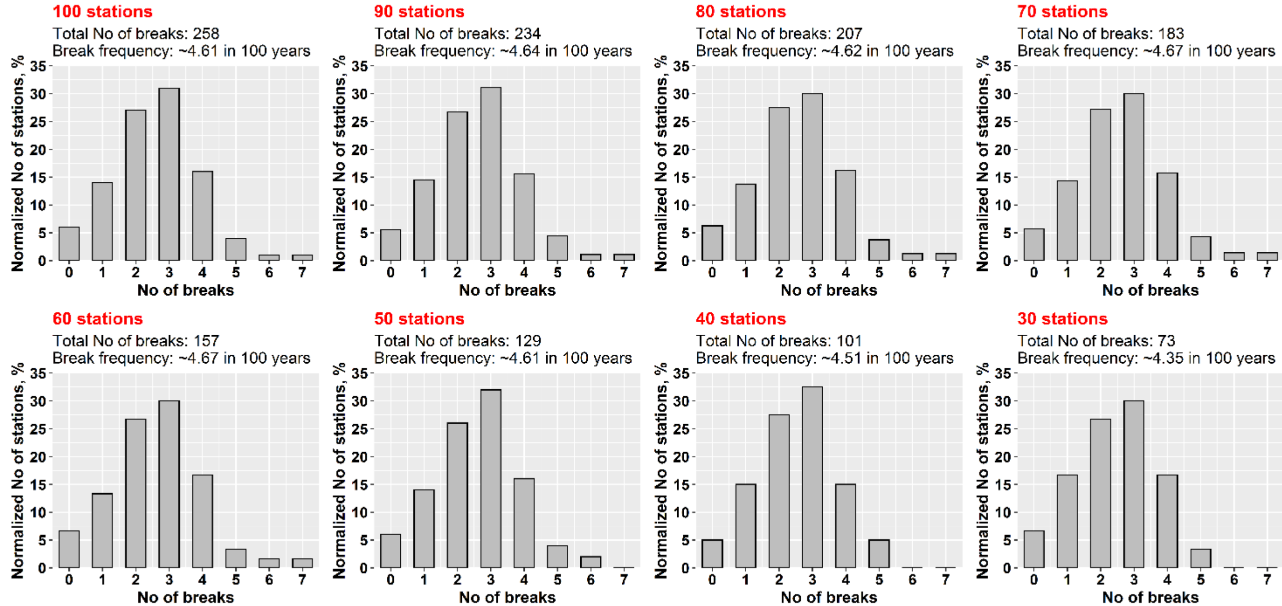
- The purpose of Case study #4 was to investigate other sources of the residual errors, i.e. length of time series and station density
- It consolidates 56 different cases: combinations of 7 periods (1950-2005, 1954-2005, …, 1974-2005) and 8 cases of station densities (see figure below)
- The station selection for each of the eight cases shown in the figure was not random: while excluding 10 stations on every step, we tried to preserve the structure of break point system (their frequency, the distributions of shift amplitudes/factors etc.; see the next slide)
- Calculations with 100 random permutations in the station signals were performed in each of 56 cases (5600 homogenization exercises were performed was for each of the two climate variables).



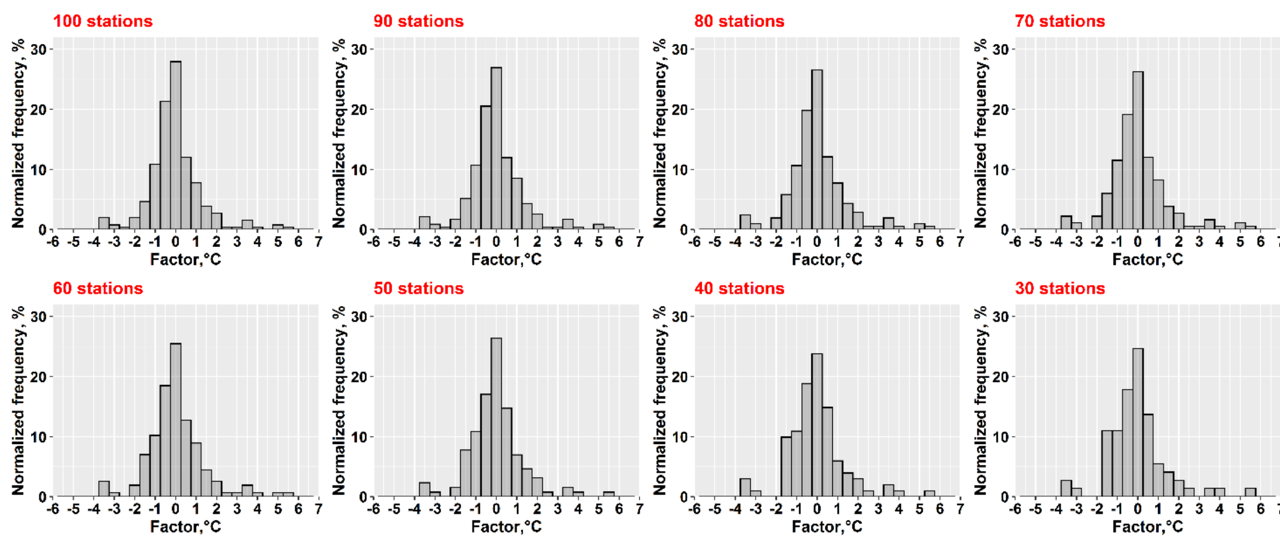
Spatial distribution of the stations in Case study #4

• Case study #4

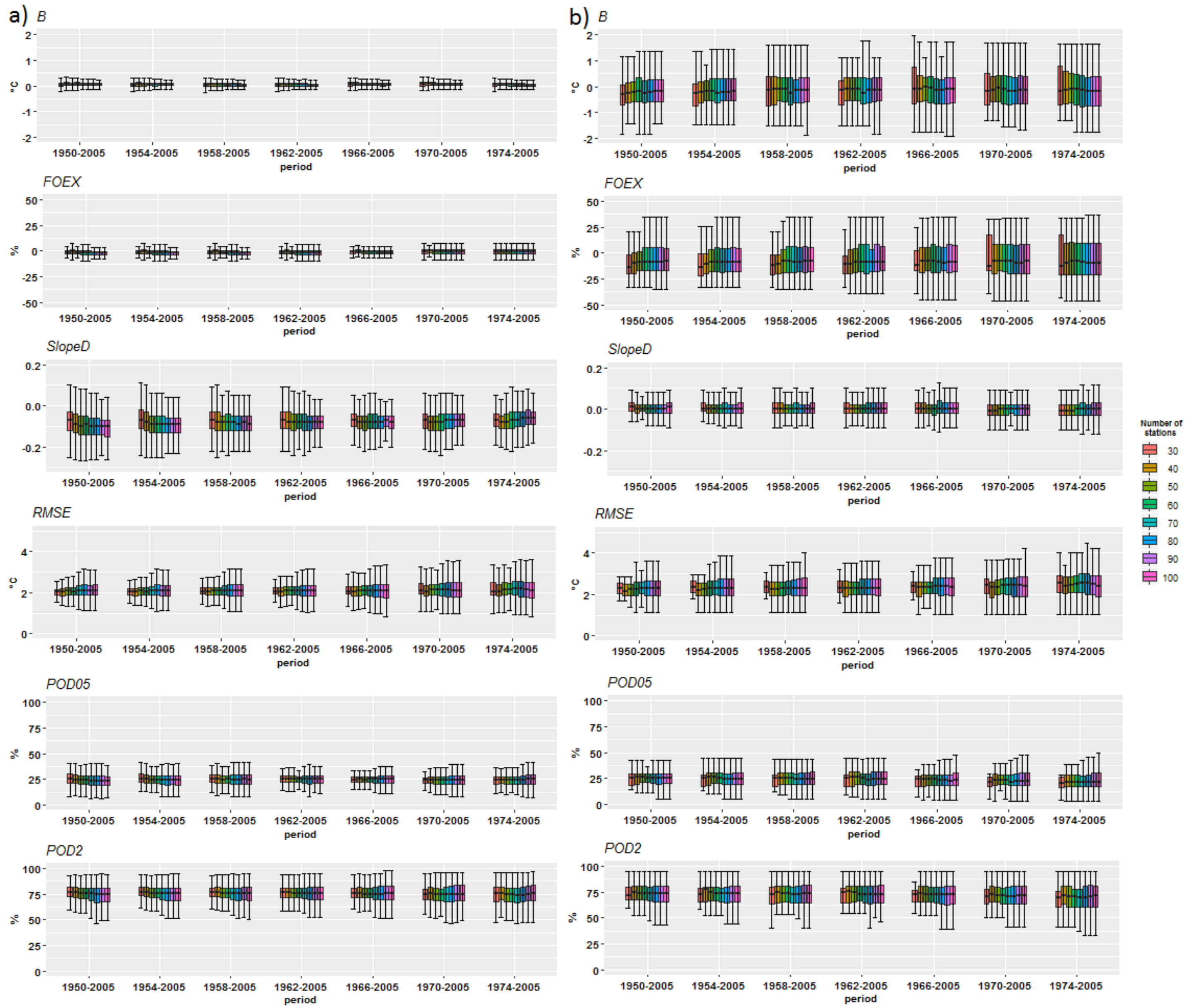
The structure of the break system in Case study #4 (the period of 1950-2005): the distribution of the number of stations/time series with respect to the number of breaks in one time series (TN data)



The structure of the break system in Case study #4 (the period of 1950-2005): histograms of shift factors (TN data)



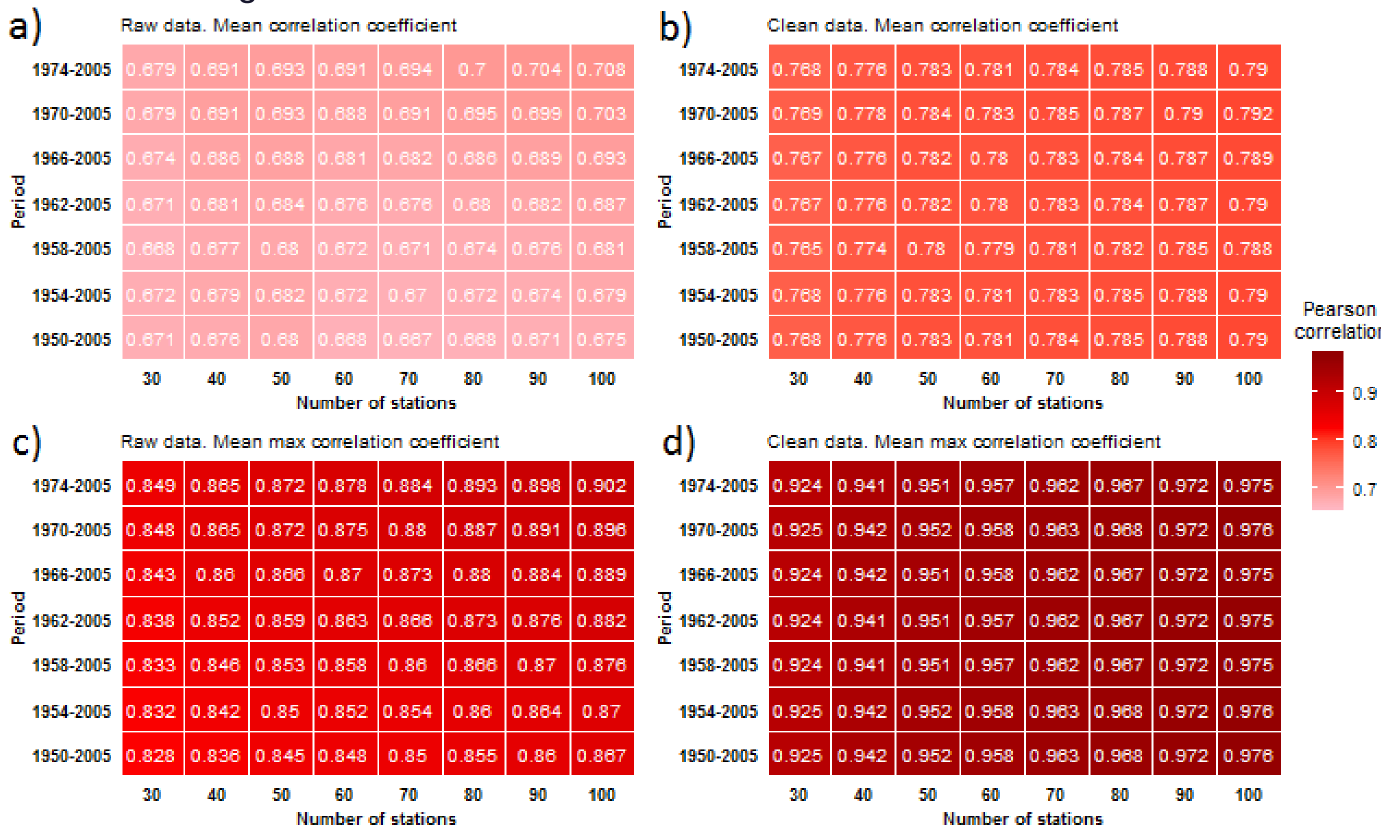
- **Case study #4**



Boxplots of the metric, calculated for TN adjusted (a) and raw (b) time series in Case study #4. Similar results were obtained for TX

● Case study #4

- Probable reason for the absence of any noticeable relation between time series length/station density and the metric values is good correlation between time series in $\boldsymbol{X}^I$ and $\boldsymbol{X}^C$



Estimated average correlation coefficients for each of the 56 considered cases for TN: (a, c) the raw data; (b, d) clean data. Top panel (a, b) represents mean values of corresponding correlation matrices, bottom panel (c, d) contains mean max values (averaged over a set of maximal correlation coefficients obtained for each time series). Year cycles were removed before calculations similarly to (Vincent et al., 2018, doi:10.1002/joc.5203)

Conclusion remarks

• Climatol's adjustment uncertainty on day-to-day scale depends on seasons. In summer months, the residual errors of daily TN and TX temperature values belong to the intervals, $(P05, P95)$, $(-2.8, 1.7)$ and $(-2.6, 1.9)$ $({}^{o}C)$ respectively. In winter months, the ranges of the errors for TN/TX are larger $(-3.6, 4.5)/(-2.0, 2.6)$ $({}^{o}C)$.

• Overall adjustment uncertainty (averaged over all seasons) can be evaluated as the error ranges, $(P05, P95)$, of $(-3.2, 3.2)/(-2.5, 2.3)$ $({}^{o}C)$. In terms of standard deviations of the residual error distributions, the overall uncertainty can be estimated as $2.15/1.64$ $({}^{o}C)$.

• The latter estimates agree well with mean values of $RMSE$, calculated based on adjusted and clean data.

• Non-stationarity of means of the residual errors can also be reported. It is related to underestimation of an amplitude of the seasonal cycle. Metric $SlopeD$ provides additional evidence for such conclusion.

• The underestimation of the temperature seasonal cycle might be reflected in yearly time series of climate extremes indices with the absolute thresholds, like ID and TR for TN or FD and SU for TX. However, its influence on linear trends in these yearly time series is not so noticeable.

• Climatol removes very well systematic errors related to jumps in the mean. The ability of Climatol to remove systematic bias is valid for shifts of any magnitude and does not depend on the number of break points in the raw time series.

Acknowledgement

This work was performed in a frame of the INDECIS project. INDECIS is a part of ERA4CS, an ERA-NET initiated by JPI Climate, and funded by FORMAS (SE), DLR (DE), BMWFW (AT), IFD (DK), MINECO (ES), ANR (FR) with co-funding by the European Union (Grant 690462).

The work has been partially supported by the Ministry of Education and Science of Kazakhstan (Grant BR05236454) and Nazarbayev University (Grant 090118FD5345).