



Norwegian  
Meteorological  
Institute



# Long-term homogenised precipitation data sets for Norway

Elin Lundstad, Department of climate services, Norwegian  
Meteorological institute

Petr Stepanek & Pavel Zahradníček, Global Change Research  
Institute, Czech Academy of Sciences

19.04.2017

# Objectives

- Establish a quality assurance tools to identify and adjust for homogeneity breaks
- Develop methodology to generate "homogenized" daily values of precipitation and temperature for given locations
- Produce homogenized monthly and daily values of temperature and precipitation for a number of long climate series
- Faciliate analysis by providing homogenized data to external users

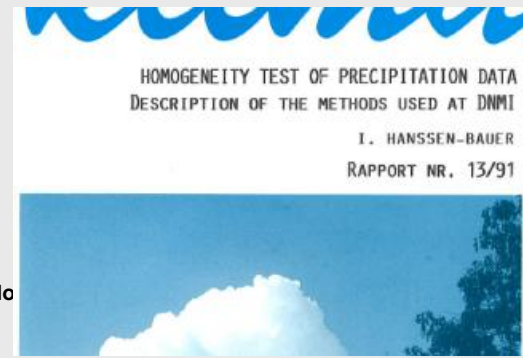
# Background for this work

- Cooperation on the field of homogenization of the Norwegian meteorological institute and Czech hydrometeorological institute (past, around 2004 - 2010, air temperature) and Global Change Research Institute, Czech Academy of Sciences (now, 2016 - X, precipitation)
- Air temperature records – finished in the last years (Home.R, MASH), now the task is precipitation
- Homogenization results applying SNHT, Home.R and MASH
- New results (for precipitation):
  - November 2016, the first version
  - March 2017, cleaned version from main problems (detected in the first version, time shift etc.)

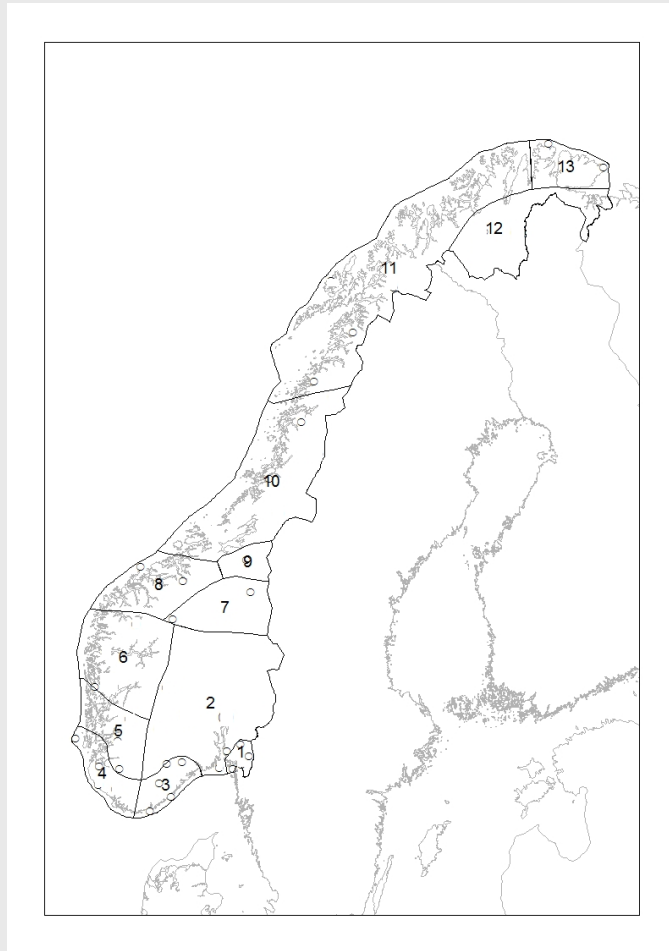
# Precipitation homogenization, past

- About 20 years ago (1991-96) the SNHT method was used on some Norwegian annual precipitation series and monthly temperature series.
- The first results were presented in the DNMI report KLIMA publications (Hanssen-Bauer et al. 1991) and (Førland and Hanssen-Bauer, 1992, in Norwegian).
- The last one was published in English in 1994 (Hanssen-Bauer and Førland).
- In the first one 151 annual precipitation series of 75 years length were tested; 52 stations were classified as homogeneous, 99 with at least one break point

Homogenizing Long Norwegian Precipitation Series.  
Source: I. Bauer-Hanssen and E.J. Førland, MET NO  
AMS 1994 (7) No. 6



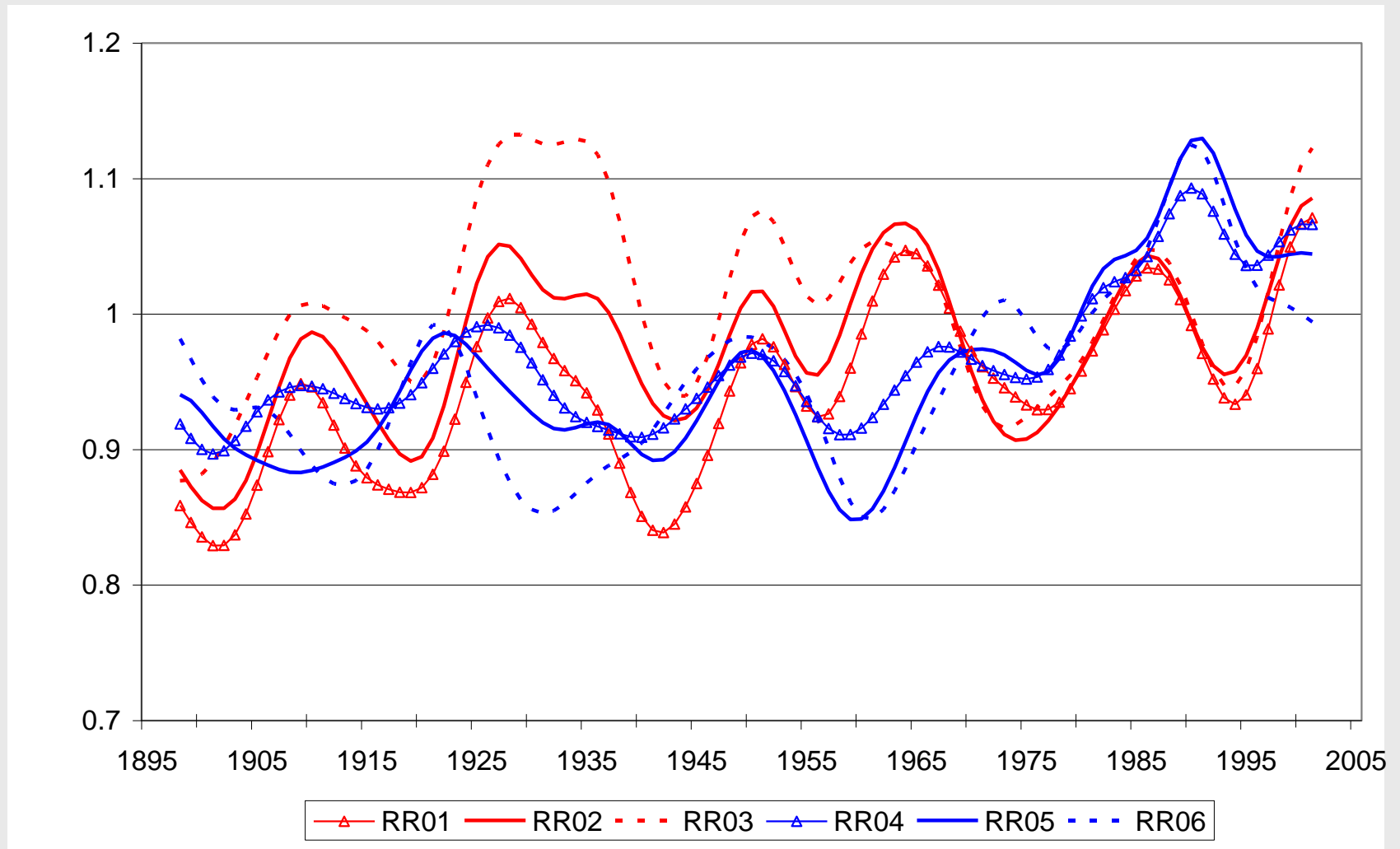
# Homogenization of precipitation data in Norway (past)



**The norwegian  
precipitation  
regions**

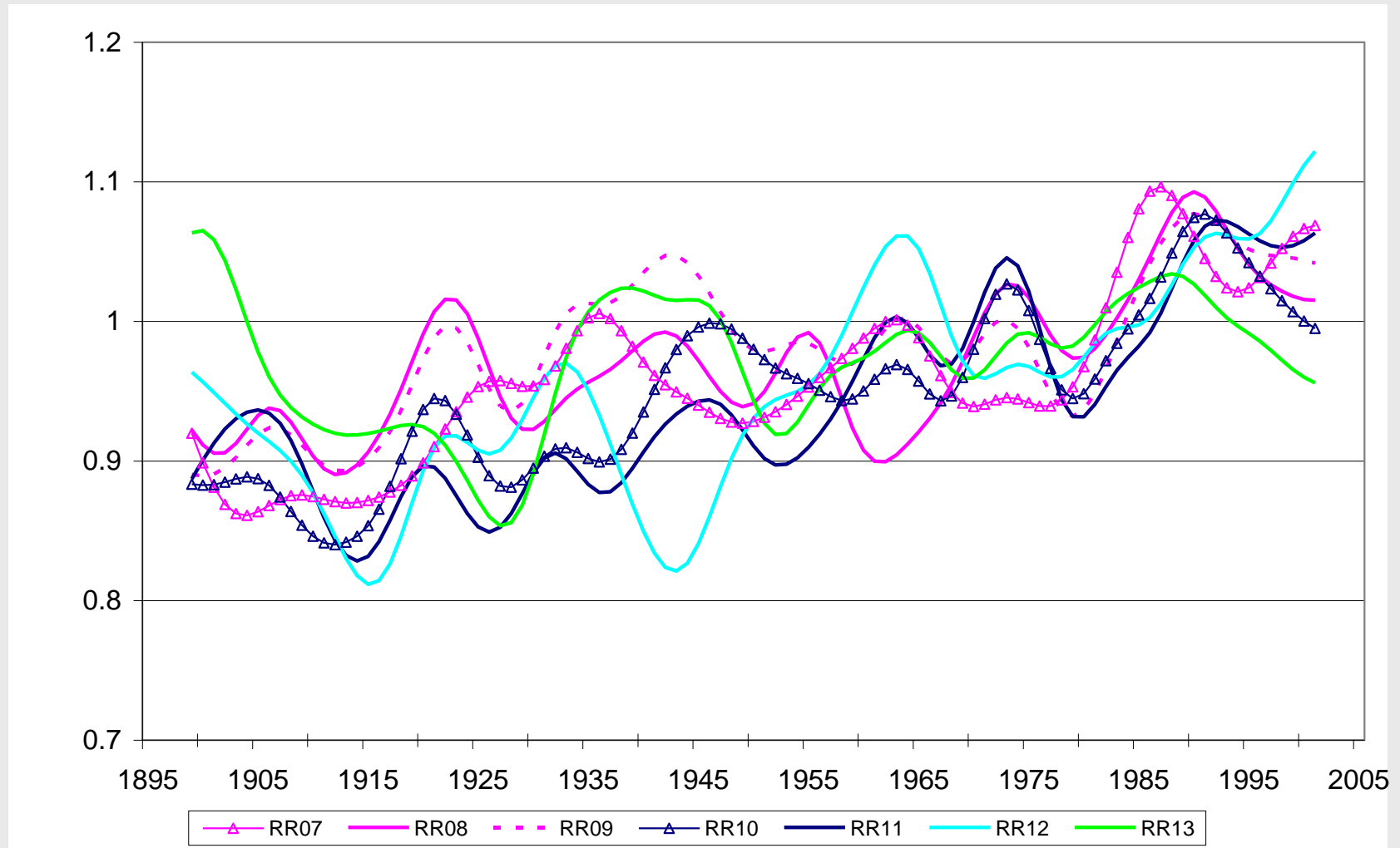
# Regional precipitation the last 100 years

## Region 1-6



# Regional precipitation the last 100 years

## Region 7-13

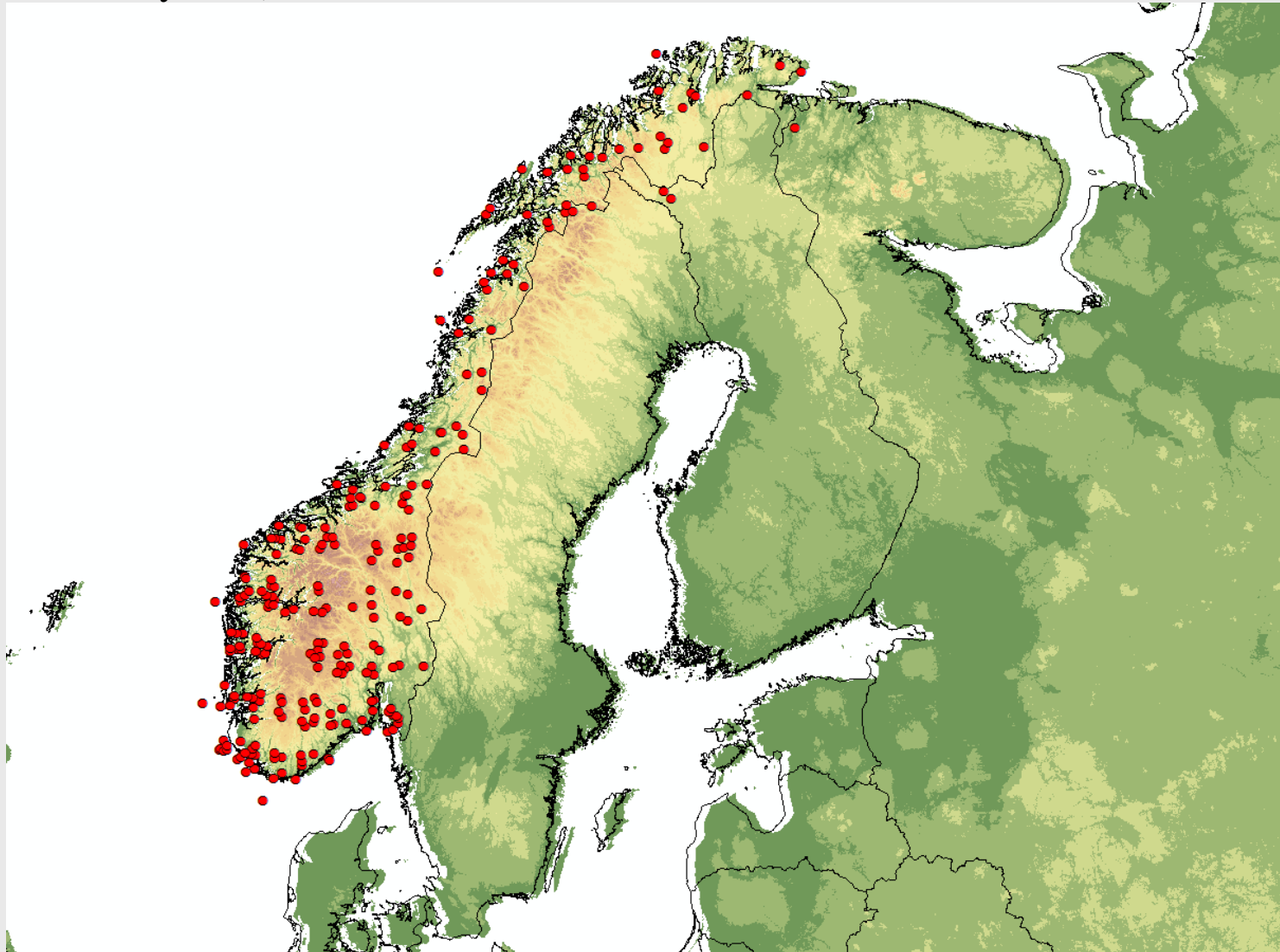


# Precipitation homogenization, update



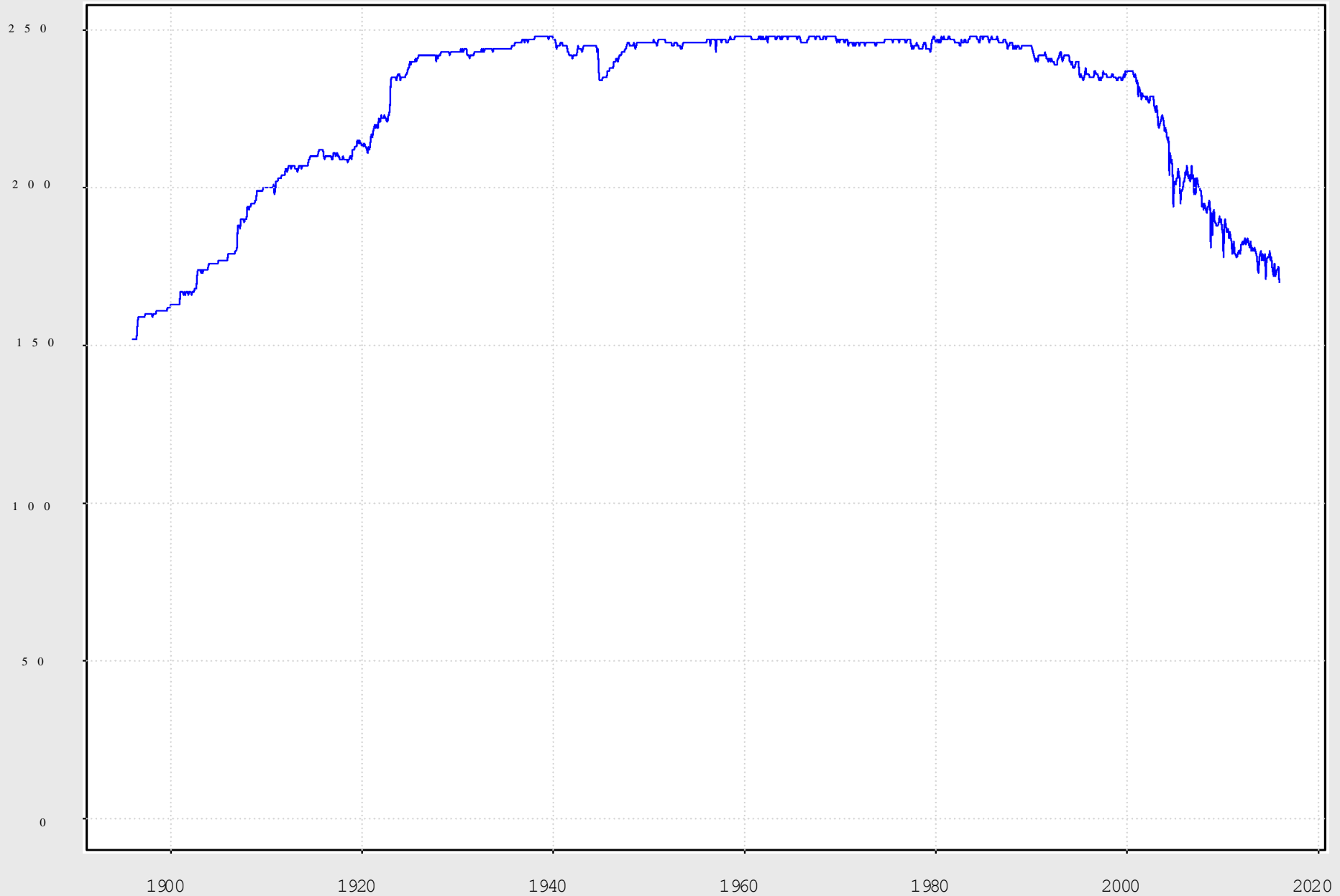
# Precipitation measurements

- 248 long precipitation series (1896-2015, various length), monthly data (with outlook to process daily data)

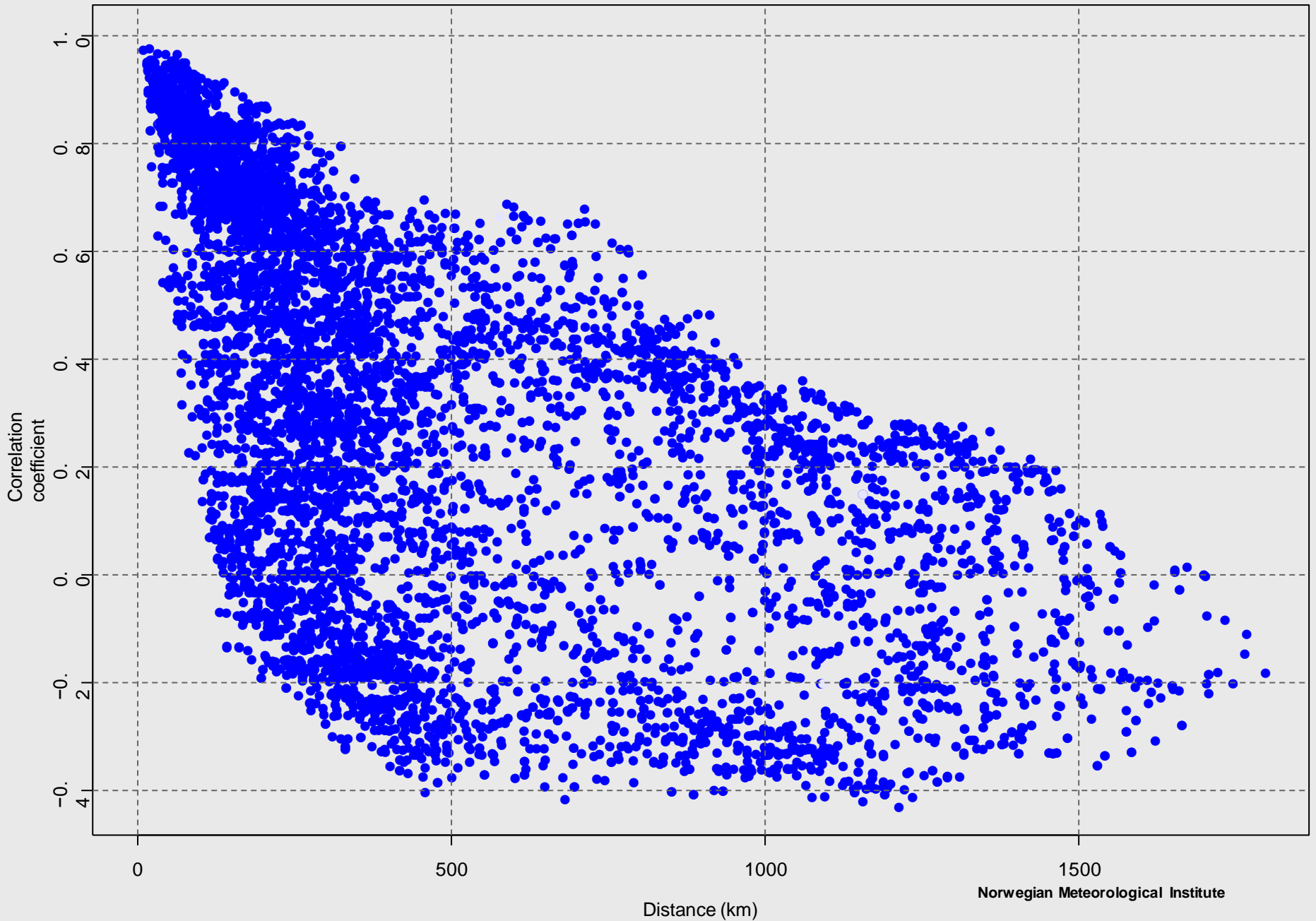


# Precipitation Stations available through the time

Nr of RR data in network 00124



Correlogram of first difference 100 sampled series

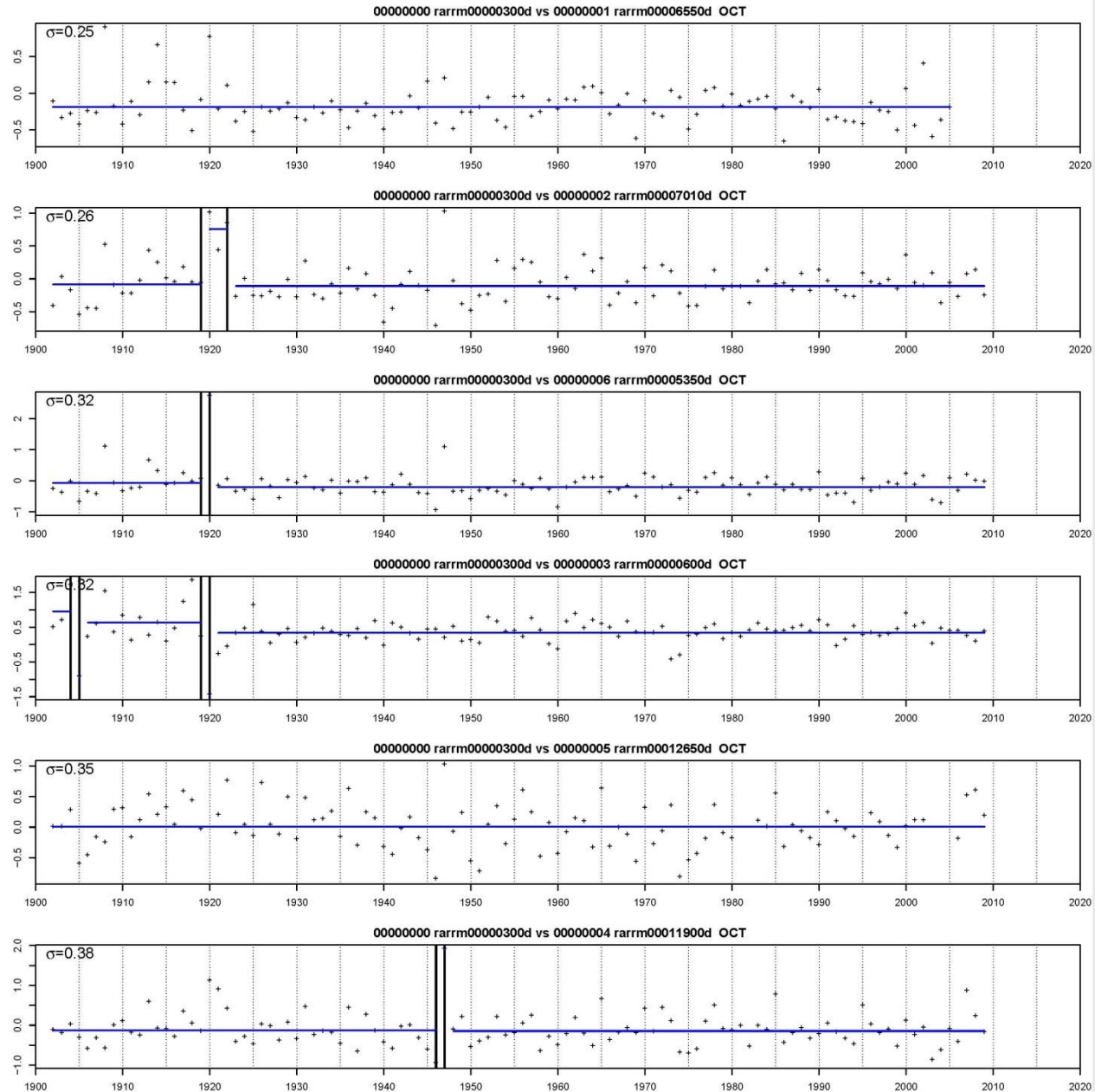


# Precipitation homogenization, update

- Home.R and MASH applied in the first step
- But Home.R (applying former definition of the regions): even if it worked excellently for air temperature, it was not the case for precipitation (finding outliers, finding breaks, ... )
- Only MASH could be used so far
- Need to compare it with another independent source (verification): AnClim + ProClim DB solution

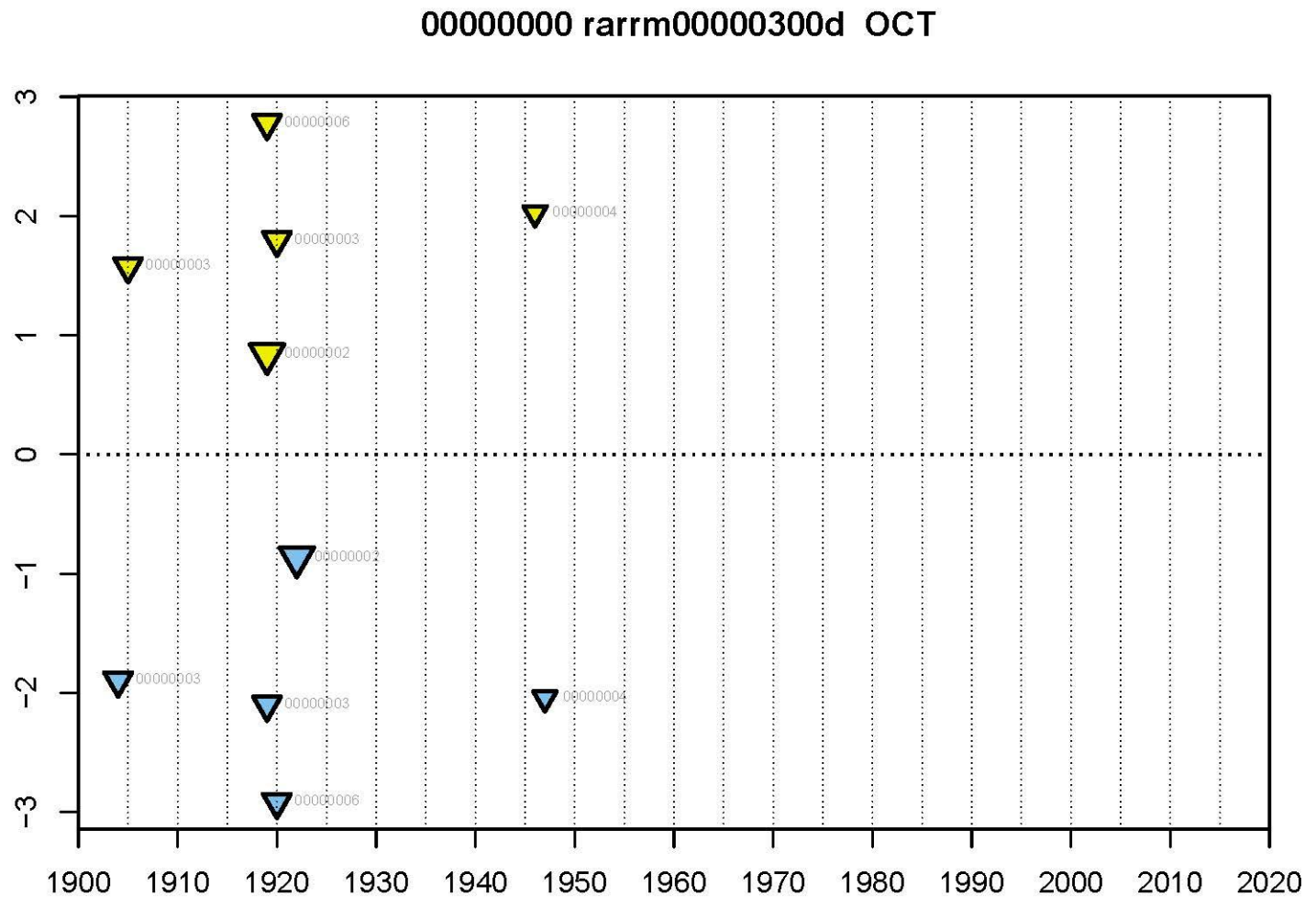
# Other Home.R precipitation issues

... ?



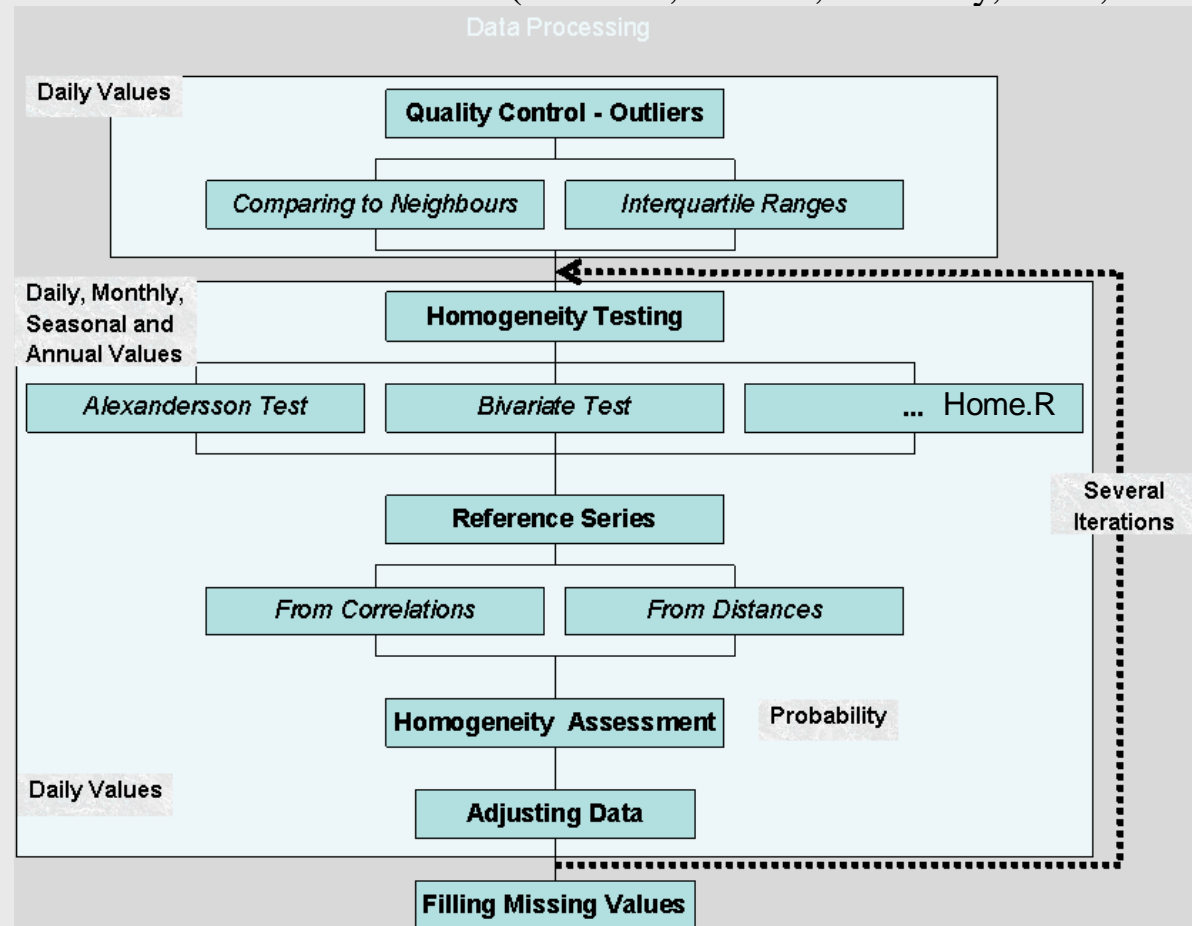
# Other Home.R precipitation issues

... ?



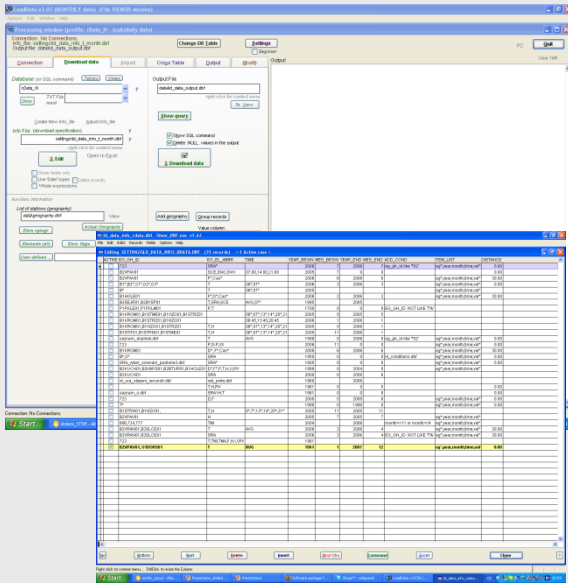
# Homogenization using AnClim and ProClimDB

- Being developed since 1996
- Combination of several methods (break detection & correction, spatial interpolation)
- Experience with data from other countries (Slovakia, Austria, Germany, USA, Croatia, Bhutan and others)



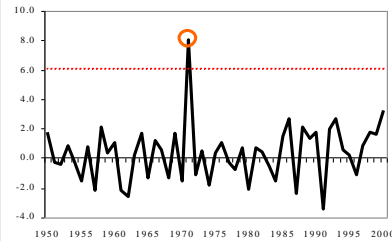
# Download data from database (e.g. Oracle)

(LoadData)



# Quality control

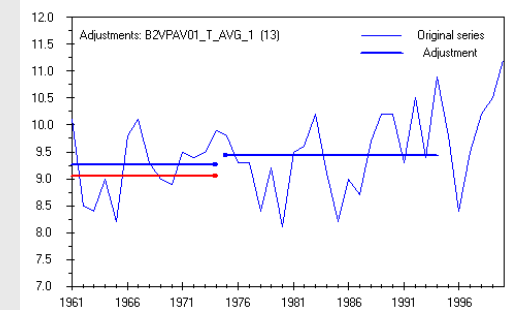
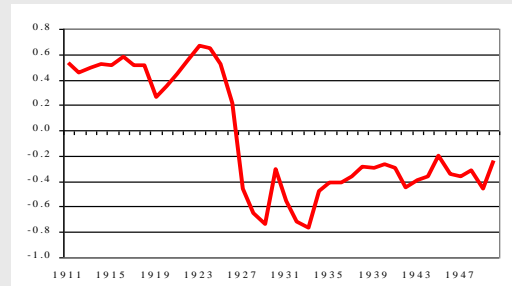
(ProClimDB)



STATION_ID	STATION_NAME	STATION_CODE	STATION_TYPE	STATION_CATEGORY	STATION_STATUS	STATION_HEIGHT	STATION_LATITUDE	STATION_LONGITUDE	STATION_ALTITUDE	STATION_INSTRUMENT	STATION_OPERATOR	STATION_START_DATE	STATION_END_DATE	STATION_DATA_SOURCE
T_03_30	BURTURHI_T_03_30					241,00			225,00	670,00	203,00	210,00	749,00	I
T_03_30	BURCARHI_T_03_30								11,28		36,85			
T_03_30	BIFPROTHI_T_03_30								st_2_6		36,85	59,12		
T_03_30	COPPERHI_T_03_30								st_3_6				62,85	91,95
T_03_30	COLOLOHI_T_03_30								st_4_6					
T_03_30	COCEBETHI_T_03_30								st_5_6					
T_03_30	COCEBETHI_T_03_30	2866	6	25	27,38	17,28			17,30	16,18	16,50	16,88	16,18	-
T_03_45	BURTURHI_T_03_45					241,00			225,00	670,00	203,00	210,00	749,00	I
T_03_45	BURCARHI_T_03_45								11,28		36,85			
T_03_45	BIFPROTHI_T_03_45								st_1_6		36,85	59,12		
T_03_45	COPPERHI_T_03_45								st_2_6				62,85	91,95
T_03_45	COLOLOHI_T_03_45								st_3_6					
T_03_45	COCEBETHI_T_03_45								st_4_6					
T_03_45	COCEBETHI_T_03_45	2866	6	25	26,28	17,28			17,30	16,18	16,50	16,88	16,28	-
T_04_00	BURTURHI_T_04_00					241,00			225,00	670,00	203,00	210,00	749,00	I
T_04_00	BURCARHI_T_04_00								11,28		36,85			
T_04_00	BIFPROTHI_T_04_00								st_1_6		36,85	59,12		
T_04_00	COPPERHI_T_04_00								st_2_6				62,85	91,95
T_04_00	COLOLOHI_T_04_00								st_3_6					
T_04_00	COCEBETHI_T_04_00								st_4_6					
T_04_00	COCEBETHI_T_04_00	2866	6	25	24,78	17,42			17,30	17,28	17,30	16,30	17,28	-
T_05_00	BURTURHI_T_05_00					241,00			225,00	670,00	203,00	210,00	749,00	I
T_05_00	BURCARHI_T_05_00								11,28		36,85			
T_05_00	BIFPROTHI_T_05_00								st_1_6		36,85	59,12		
T_05_00	COPPERHI_T_05_00								st_2_6				62,85	91,95
T_05_00	COLOLOHI_T_05_00								st_3_6					
T_05_00	COCEBETHI_T_05_00								st_4_6					
T_05_00	COCEBETHI_T_05_00	2866	6	25	24,78	17,42			17,30	17,28	17,30	16,30	17,28	-
T_06_00	BURTURHI_T_06_00					241,00			225,00	670,00	203,00	210,00	749,00	I
T_06_00	BURCARHI_T_06_00								11,28		36,85			
T_06_00	BIFPROTHI_T_06_00								st_1_6		36,85	59,12		
T_06_00	COPPERHI_T_06_00								st_2_6				62,85	91,95
T_06_00	COLOLOHI_T_06_00								st_3_6					
T_06_00	COCEBETHI_T_06_00								st_4_6					

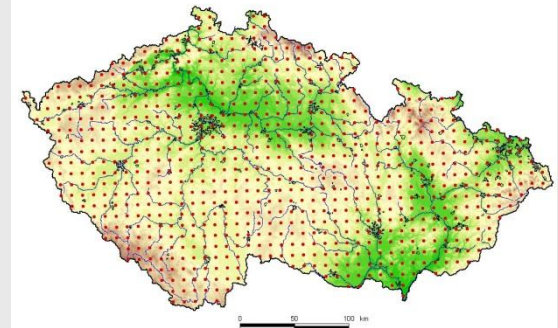
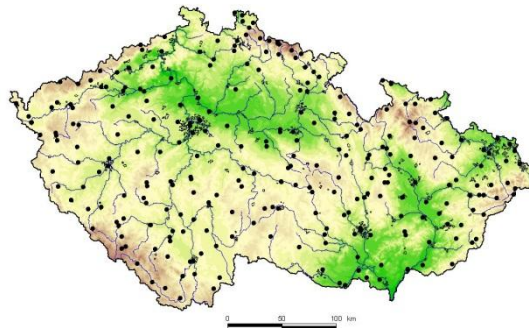
# Homogenization

(ProClimDB/AnClim)



# „Technical“ series and grid points calculation

(ProClimDB)





# ProcData software, only one Data file, accompanied by Info\_file

database processing

Processing window (profile: slovensko)

Menu: **Reference**

Calculates reference series for each station given

Item: **From Correlations**

Selects given Number of stations with average correlation

Source files:

right click for context menu

Data file

:\e\t\_hurv\_mes\_new\_reconstr2

(Data Info file)

data\data\_info.dbf

ID	EG_EL_ABBR	YEAR	DAY	TIME	N1	N2	N3	N4	N5	N6	N7	N8
B1BYSH01_SCE_07:00	SCE	2006	24	07:00	30.00	10.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	25	07:00	28.00	10.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	26	07:00	28.00	12.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	27	07:00	28.00	9.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	28	07:00	28.00	9.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	29	07:00	28.00	-999.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	30	07:00	28.00	-999.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	31	07:00	27.00	-999.00	0.00	-999.00	0.00	-999.00	0.00	-999.00
B1BYSH01_SNO_07:00	SNO	1961	1	07:00	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	2	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	3	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	4	07:00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	5	07:00	0.00	5.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	6	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	7	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	8	07:00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00

	NAME	ID	B	E	L	IDXXX	III	REGION	LATITUDE	LONGITUDE	ALTITUDE	BEGIN	END	LENGTH	MISS_CN
<input checked="" type="checkbox"/>	Bystřice pod Hostýnem	B1BYSH01_SCE_07:00				B1BYSH01		SCE	17.67	49.40	315	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Bystřice pod Hostýnem	B1BYSH01_SNO_07:00				B1BYSH01		SNO	17.67	49.40	315	1.1.1961	31.1.2006	46	0.00
	Bystřice pod Hostýnem	B1BYSH01_SRA_07:00				B1BYSH01		SRA	17.67	49.40	315	1.1.1872	31.1.2006	135	0.00
<input checked="" type="checkbox"/>	Bystřice pod Hostýnem	B1BYSH01_SVH_07:00				B1BYSH01		SVH	17.67	49.40	315	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Holešov	B1HOLE01_SCE_07:00				B1HOLE01		SCE	17.57	49.32	224	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Holešov	B1HOLE01_SNO_07:00				B1HOLE01		SNO	17.57	49.32	224	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Holešov	B1HOLE01_SRA_07:00				B1HOLE01		SRA	17.57	49.32	224	1.1.1953	31.1.2006	54	0.00
<input checked="" type="checkbox"/>	Holešov	B1HOLE01_SVH_07:00				B1HOLE01		SVH	17.57	49.32	224	1.1.1979	31.1.2006	28	0.00
<input checked="" type="checkbox"/>	Napajedla	B1NAPA01_SCE_07:00				B1NAPA01		SCE	17.52	49.18	185	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Napajedla	B1NAPA01_SNO_07:00				B1NAPA01		SNO	17.52	49.18	185	1.1.1961	31.1.2006	46	0.00
	Napajedla	B1NAPA01_SRA_07:00				B1NAPA01		SRA	17.52	49.18	185	1.1.1889	31.1.2006	118	0.00
<input checked="" type="checkbox"/>	Napajedla	B1NAPA01_SVH_07:00				B1NAPA01		SVH	17.52	49.18	185	1.1.1977	31.1.2006	30	0.00
	Brno	B2BKVE01_SCE_07:00				B2BKVE01		SCE	16.57	49.19	223	2.1.1922	31.1.1970	49	0.00
	Brno	B2BKVE01_SNO_07:00				B2BKVE01		SNO	16.57	49.19	223	3.1.1931	31.1.1970	40	0.00
	Brno	B2BKVE01_SRA_07:00				B2BKVE01		SRA	16.57	49.19	223	1.1.1922	31.1.1970	49	0.00
	Brno	B2BPIS01_SCE_07:00				B2BPIS01		SCE	16.57	49.20	203	1.1.1919	31.1.1979	61	0.00
	Brno	B2BPIS01_SNO_07:00				B2BPIS01		SNO	16.57	49.20	203	4.1.1931	31.1.1979	49	0.00
	Brno	B2BPIS01_SRA_07:00				B2BPIS01		SRA	16.57	49.20	203	1.1.1916	31.1.1979	64	0.00
<input checked="" type="checkbox"/>	Brno	B2BPIS01_SVH_07:00				B2BPIS01		SVH	16.57	49.20	203	1.1.1961	31.1.1979	19	0.00
<input checked="" type="checkbox"/>	Brno	B2BTUR01_SCE_07:00				B2BTUR01		SCE	16.70	49.16	241	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Brno	B2BTUR01_SNO_07:00				B2BTUR01		SNO	16.70	49.16	241	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Brno	B2BTUR01_SRA_07:00				B2BTUR01		SRA	16.70	49.16	241	1.1.1961	31.1.2006	46	0.00
<input checked="" type="checkbox"/>	Brno	B2BTUR01_SVH_07:00				B2BTUR01		SVH	16.70	49.16	241	1.1.1969	31.1.2006	38	0.00
<input checked="" type="checkbox"/>	Jihlava	B2JIHL01_SCE_07:00				B2JIHL01		SCE	15.54	49.39	560	1.1.1961	31.1.1969	9	0.00
<input checked="" type="checkbox"/>	Jihlava	B2JIHL01_SNO_07:00				B2JIHL01		SNO	15.54	49.39	560	1.1.1961	31.1.1969	9	0.00

Correlations column

K13

Run

Last Output

Quit

# ProClimDB software

**ProClimDB v7.61 (MONTHLY data)**

Options Edit Get info Tools Transf Calculate Calc2 Neighbors Anomalies Reference Homog Adjust Fill Miss Window Help

**Processing window (profile: slovensko)**

**Menu : Reference** 8 **Settings**  
*Calculates reference series for each station given in Info File*

**Item : From Correlations** 2 **Change PROFILE**  
*Selects given Number of stations with average correlation higher than a Limit and creates reference series*

**Source files:** *right click for context menu*

Data file	:_et_hurv_mes_new_reconstr2.dbf
(Data Info file)	data\data_info.dbf
Correlations	data\correl.dbf

**Destination files:** *right click for context menu*

Refer. Series	data\ref_series.dbf
Ref Info file	data\ref_ser_info.dbf

**Settings**

Create Info File only

Number of Stations: 5

Limit - correlation: 0.2;100

Maximum altitude diff.: -100

Weighted average

Years per one part: [ ]

Overlap - years: [ ]

Allow lenght +/- overlay

Correlations column: K13

**Process info:**

Number of stations: 5  
Difference in measuring periods (base and selected stations) is not taken into account!  
Neighbours selected according to: correlations (limit value: 0.200), based on K13 column  
- additional condition: limit distance: maximum: 100 km

Neighbours can differ in altitude at least: 100 m  
Base station has to have a lenght at least: 20 years.  
Neighbours have to have a lenght at least: 20 years.  
Minimum length of period in common: 10 years (selecting 5 stations out of 5).  
Selected stations from the same region only! (Column 'Region' in the Info\_file).

Stations processed:  
1:B1BRBY01\_TMA\_21  
2:BRBY01\_TMA\_21

**Run** **Last Output** **Quit**

Context menu over 'Ref Info file':  
Open File  
Save as ... (Copy)  
Save as DBF IV  
View / Edit Table ...  
Open in Excel ...  
Load Template  
Undo  
Copy Name to Clipboard  
Paste Name from Clipboard

Ready for action

NUM

Institute

# ProClimDB software

7.61 (MONTHLY data)

**Processing window (profiles slovensko)**

**Menu: Reference**  
Calculates reference series for each station given in Info File

**Item: From Correlations**  
Selects given Number of stations with average correlation higher than a Limit and creates reference

**Source files:**  
Data file:   
(Data Info file)   
Correlations

**Destination files:**  
Refer. Series   
Ref info file

**Settings**

Create Info File only  
Number of Stations   
Limit - correlation   
Maximum altitude diff.   
 Weighted average  
Years per one part   
Overlap - years   
 Allow length +/- error  
Correlations column

**Process info:**  
Number of stations: 5  
Difference in measuring periods (base and selected) taken into account  
Neighbours selected according to: correlator based on k13 column - additional condition: limit distance: maximum  
Neighbours can differ in altitude at least: 100 m  
Base station has to have a length at least: 20 years  
Neighbours have to have a length at least: 20 years  
Minimum length of period in common: 10 years (select if 0)  
Selected stations from the same region only (Column 5 Info file)  
Stations processed:  
1: B1BRBY01\_TMA\_21

ref info t.dbf - Show\_DBF.exe v1.2.4

File Edit Edit2 Records Fields Options Help

Editing D:\dokumenty\progr\proc data\DATA\zprac\_CR\Vse\_od61\ref info t.dbf (12306 records, 20 marked for deleting)

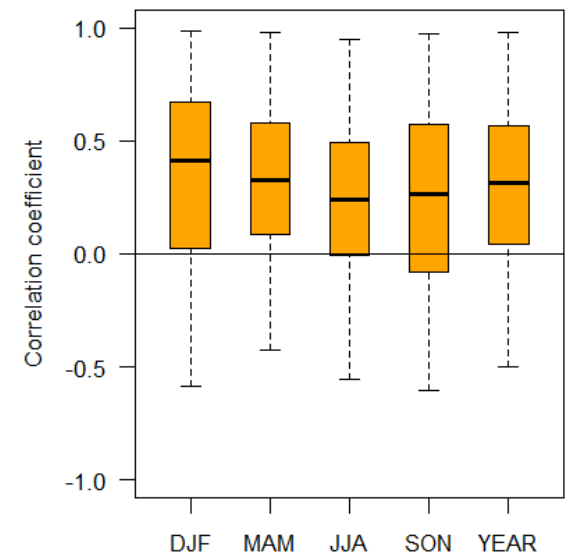
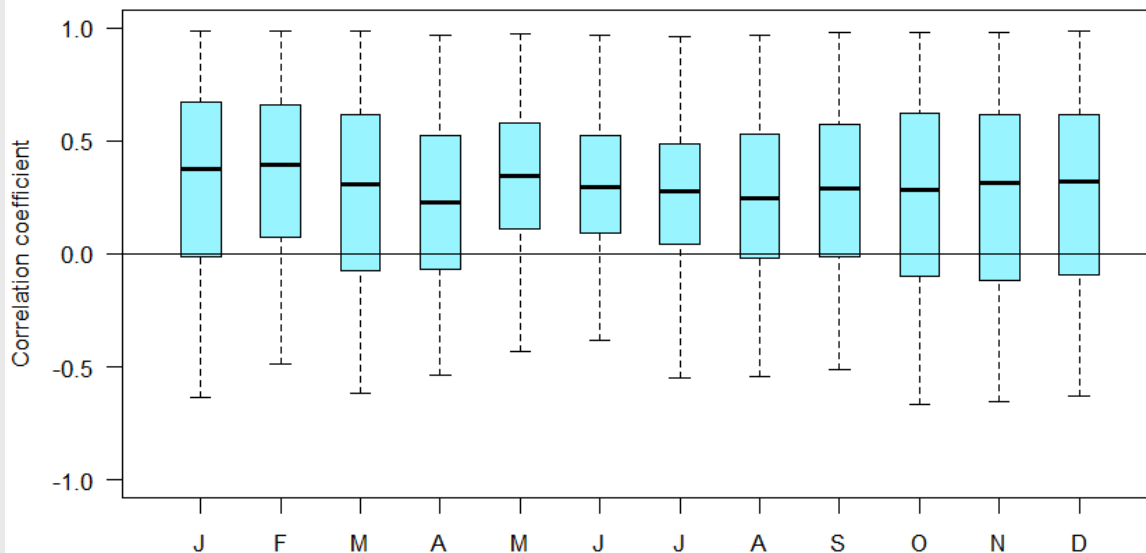
ID_1	ID_2	REGION	BEGIN	END	LENGTH	REMARK	CORREL	DISTANCE	AZIMUTH	AL
B1BRBY01_T_07:00	B1BRBY01_T_07:00_1_d	T_07:00	1.1.1960	31.12.1989	10958	0st.		0.00	0.0	50
B1BRBY01_T_07:00	B1BRBY01_T_07:00_2_d	T_07:00	31.12.1964	31.12.1994	10957	5st. (l:29.3		92.80	122.8	50
	B1LUHA01_T_07:00	T_07:00	31.12.1960	31.12.2007		10957 y. com		18.25	176.4	50
	B1VIZO01_T_07:00	T_07:00	31.12.1960	31.12.2007		10957 y. com		18.71	134.4	50
	O3HUSL01_T_07:00	T_07:00	31.12.1960	31.12.2007		10957 y. com		23.66	70.3	50
	O3VSET01_T_07:00	T_07:00	31.12.1960	31.12.2007		10957 y. com		26.76	93.1	50
	B1ZLIN01_T_07:00	T_07:00	31.12.1960	31.12.1996		10957 y. com		29.30	150.3	50
B1BRBY01_T_14:00	B1BRBY01_T_14:00_1_d	T_14:00	1.1.1960	31.12.1989	10958	0st.		0.00	0.0	50
B1BRBY01_T_14:00	B1BRBY01_T_14:00_2_d	T_14:00	31.12.1964	31.12.1994	10957	5st. (l:29.3		92.80	122.8	50
	B1LUHA01_T_14:00	T_14:00	31.12.1960	31.12.2007		10957 y. com		18.25	176.4	50
	B1VIZO01_T_14:00	T_14:00	31.12.1960	31.12.2007		10957 y. com		18.71	134.4	50
	O3HUSL01_T_14:00	T_14:00	31.12.1960	31.12.2007		10957 y. com		23.66	70.3	50
	O3VSET01_T_14:00	T_14:00	31.12.1960	31.12.2007		10957 y. com		26.76	93.1	50
	B1ZLIN01_T_14:00	T_14:00	31.12.1960	31.12.1996		10957 y. com		29.30	150.3	50
B1BRBY01_T_21:00	B1BRBY01_T_21:00_1_d	T_21:00	1.1.1960	31.12.1989	10958	0st.		0.00	0.0	50
B1BRBY01_T_21:00	B1BRBY01_T_21:00_2_d	T_21:00	31.12.1964	31.12.1994	10957	5st. (l:29.3		92.80	122.8	50
	B1LUHA01_T_21:00	T_21:00	31.12.1960	31.12.2007		10957 y. com		18.25	176.4	50
	B1VIZO01_T_21:00	T_21:00	31.12.1960	31.12.2007		10957 y. com		18.71	134.4	50
	O3HUSL01_T_21:00	T_21:00	31.12.1960	31.12.2007		10957 y. com		23.66	70.3	50
	O3VSET01_T_21:00	T_21:00	31.12.1960	31.12.2007		10957 y. com		26.76	93.1	50
	B1ZLIN01_T_21:00	T_21:00	31.12.1960	31.12.1996		10957 y. com		29.30	150.3	50
B1BRBY01_T_AVG	B1BRBY01_T_AVG_1_d	T_AVG						0.00	0.0	50
B1BRBY01_T_AVG	B1BRBY01_T_AVG_2_d	T_AVG						92.80	122.8	50
	B1LUHA01_T_AVG	T_AVG						18.25	176.4	50
	B1VIZO01_T_AVG	T_AVG						18.71	134.4	50
	O3HUSL01_T_AVG	T_AVG						23.66	70.3	50
	O3VSET01_T_AVG	T_AVG	31.12.1960	31.12.2007		10957 y. com		26.76	93.1	50
	B1ZLIN01_T_AVG	T_AVG	31.12.1960	31.12.1996		10957 y. com		29.30	150.3	50

Right click for context menu ...

- Sort data according to this column
- Sort data according to All columns CTRL+O
- Find a string CTRL+F
- Find next F3
- Replace strings CTRL+L
- List cases of the column CTRL+T
- Filter (show rows of a particular case)**
- Filter out into new Application
- Blank the cell CTRL+B
- Insert row CTRL+I
- Mark/Unmark record for deleting CTRL+D
- Delete rest (mark) CTRL+A
- Recall rest (unmark) CTRL+R
- Copy row(s) to Clipboard CTRL+W
- Paste row(s) from Clipboard CTRL+E
- Display DBF file
- Quit viewer CTRL+Q

# Spatial dependence

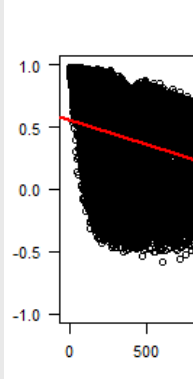
- A strong spatial dependence is crucial for successful homogenization
- Precipitation may occur very locally
- With distance and higher altitude difference, correlations drops quickly down
- In case of Norway we find the strongest dependence in winter and lowest in summer or autumn. From individual months, the weakest correlation is in April



# Spatial dependence

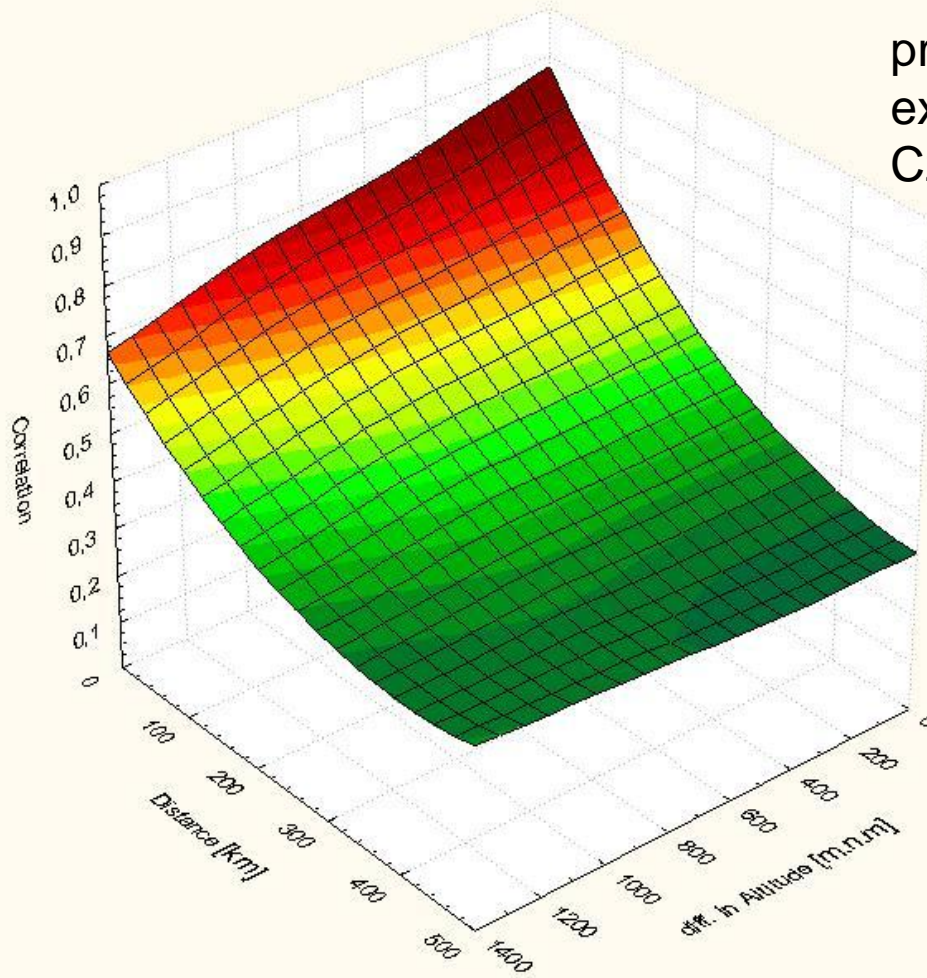
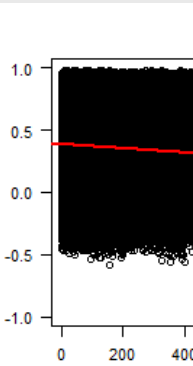
DISTANCE

DJF

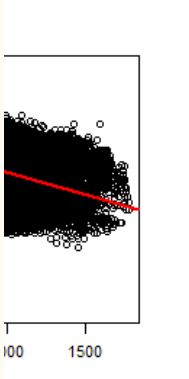


ALTITUDE

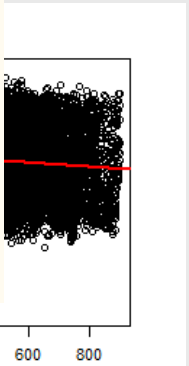
DJF



precipitation:  
example from the  
Czech Republic

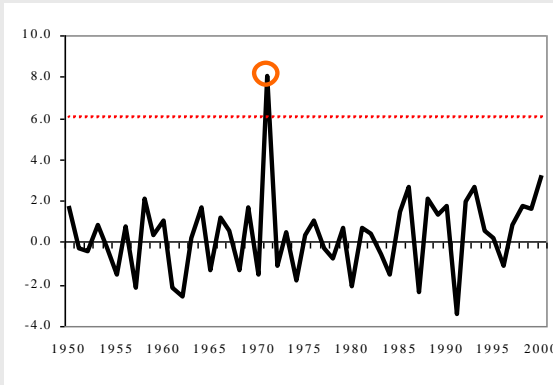


- > 0,8
- < 0,8
- < 0,7
- < 0,6
- < 0,5
- < 0,4

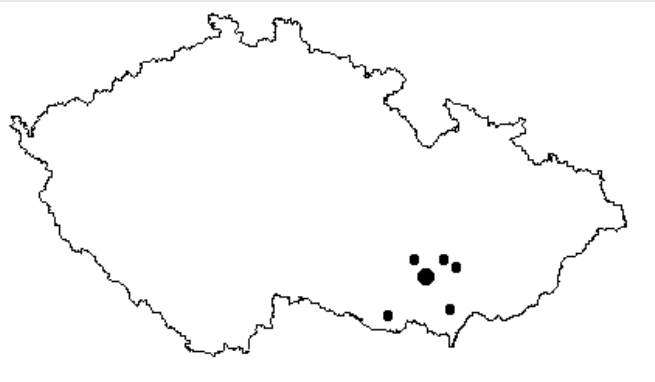


# Data Quality Control

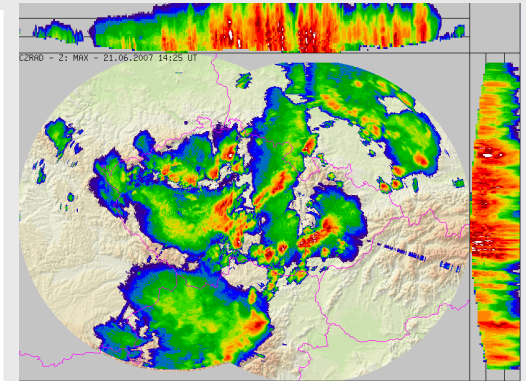
- Own approach, combination of several methods



Interquartile ranges



Comparing with neighbours  
Comparing with expected values



Comparing with radar information (not possible for Norway)

A	B	C	D	E	F	G	H	I	J	K	L	M	N
REGIC	ID	YE	MON1	DA	ST_BASE	EXPECT	REMAR	ST_1	ST_2	ST_3	ST_4	ST_5	DIF1_S
T_03:30	B2BTUR01_T_03:30				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	1
T_03:30	B2BZAB01_T_03:30						st_1, di	11,58					
T_03:30	B1PROT01_T_03:30						st_2, di		36,85				
T_03:30	O3PRER01_T_03:30						st_3, di			59,12			
T_03:30	O2OLOM01_T_03:30						st_4, di				62,88		
T_03:30	O1CERV01_T_03:30						st_5, di					91,95	
<b>T_03:30</b>	<b>B2BTUR01_T_03:30</b>	<b>2006</b>	<b>6</b>	<b>25</b>	<b>27,30</b>	<b>17,28</b>		<b>17,30</b>	<b>16,10</b>	<b>15,50</b>	<b>15,80</b>	<b>16,10</b>	<b>-7</b>

# Data quality control

Year/  
Month



Test  
values



Expect  
values



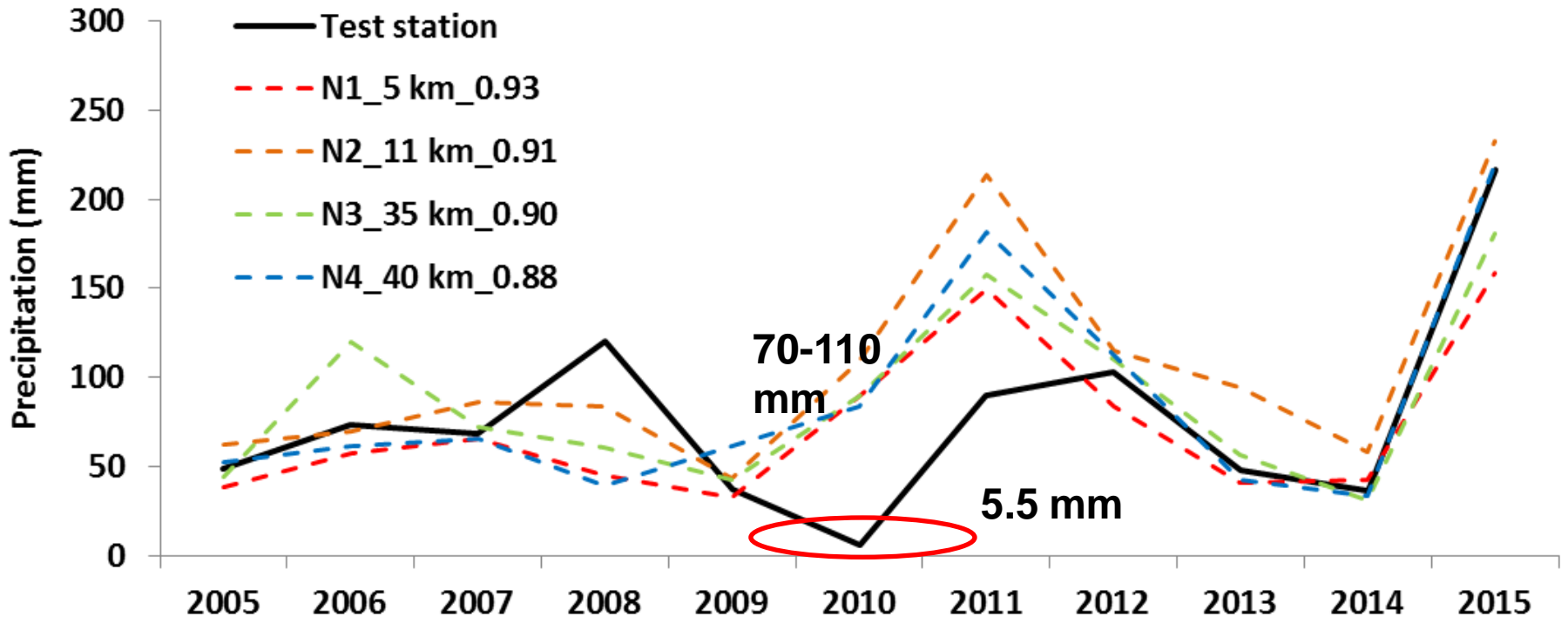
Neighbours stations

REGION	ID	YEAR	MONTH	ST_BASE	EXPECT_TRF	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5	ST_6
	rarrm00018450d			173.0		Altitude	94.0	360.0	92.0	144.0	164.0	372.0
	rarrm00018700d					st_1, Co	0.9			<b>Distance</b>		
	rarrm00018500d					st_2, Co		0.9				
	rarrm00017850d					st_3, Co	<b>Correlation</b>		0.9			
	rarrm00003780d					st_4, Co				0.9		
	rarrm00004050d					st_5, Co					0.9	
	rarrm00020520d					st_6, Co						0.9
	<b>rarrm00018450d</b>	<b>2010</b>	<b>9</b>	<b>5.5</b>	<b>85.407</b>		<b>89.1</b>	<b>109.5</b>	<b>89.5</b>	<b>83.5</b>	<b>71.3</b>	<b>79.7</b>
	rarrm00029600d			870.0		Altitude	720.0	828.0	166.0	474.0	628.0	890.0
	rarrm00025300d					st_1, Co	0.9					
	rarrm00029800d					st_2, Co		0.9				
	rarrm00024890d					st_3, Co			0.9			
	rarrm00022730d					st_4, Co				0.8		
	rarrm00022840d					st_5, Co					0.8	
	rarrm00023400d					st_6, Co						0.8
	<b>rarrm00029600d</b>	<b>1977</b>	<b>6</b>	<b>173.8</b>	<b>82.517</b>		<b>74.2</b>	<b>70.0</b>	<b>59.6</b>	<b>137.5</b>	<b>68.3</b>	<b>92.1</b>
	rarrm00039750d			207.0		Altitude	278.0	504.0	220.0	295.0	151.0	245.0
	rarrm00041480d					st_1, Co	0.9					
	rarrm00041550d					st_2, Co		0.9				
	rarrm00038800d					st_3, Co			0.9			
	rarrm00037740d					st_4, Co				0.9		
	rarrm00039220d					st_5, Co					0.9	
	rarrm00038600d					st_6, Co						0.9
	<b>rarrm00039750d</b>	<b>2011</b>	<b>9</b>	<b>49.2</b>	<b>201.865</b>		<b>307.0</b>	<b>266.5</b>	<b>223.8</b>	<b>179.0</b>	<b>289.1</b>	<b>183.5</b>

# Data quality control

- 22 suspicious values detected in first version
- 19 suspicious values detected in second clean-up version

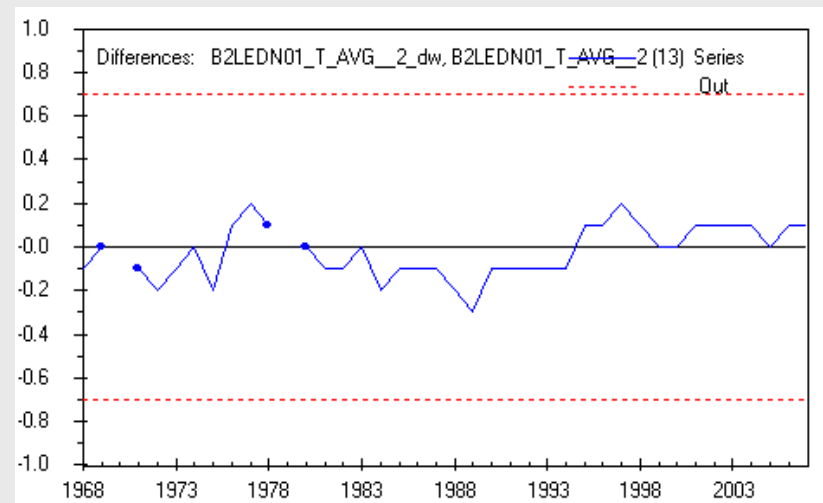
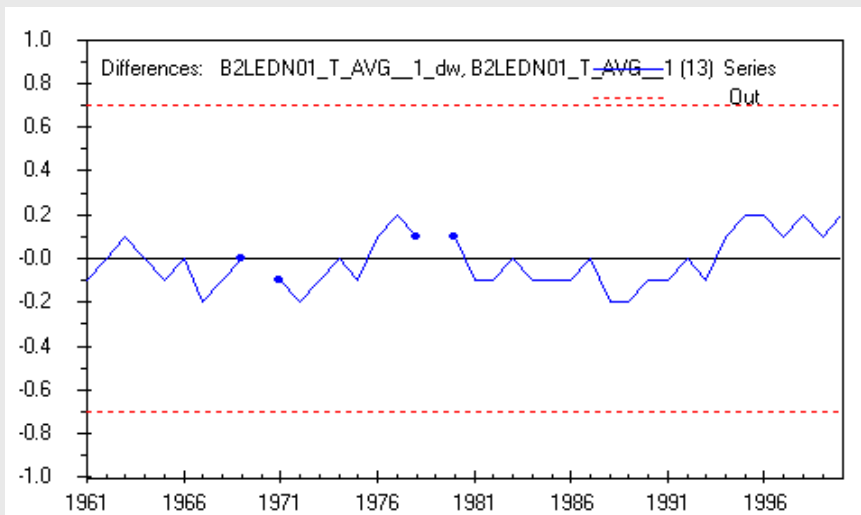
REGION	ID	YEAR	MONTH	ST_BASE	EXPECT_TRF	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5	ST_6
	rarrm00018450d			173.0		Altitude	94.0	360.0	92.0	144.0	164.0	372.0
	rarrm00018700d					st_1, Co	0.9					
	rarrm00018500d					st_2, Co		0.9				
	rarrm00017850d					st_3, Co			0.9			
	rarrm00003780d					st_4, Co				0.9		





# Homogenization

- **Detection – monthly data**
- **Two types of reference series**
  - one reference series calculated from the nearest or the best correlated neighbours stations
  - **Pair-wise detection** – comparison with each neighbours station individually
- **SNHT, Bivariate and t-test, and Home.R**

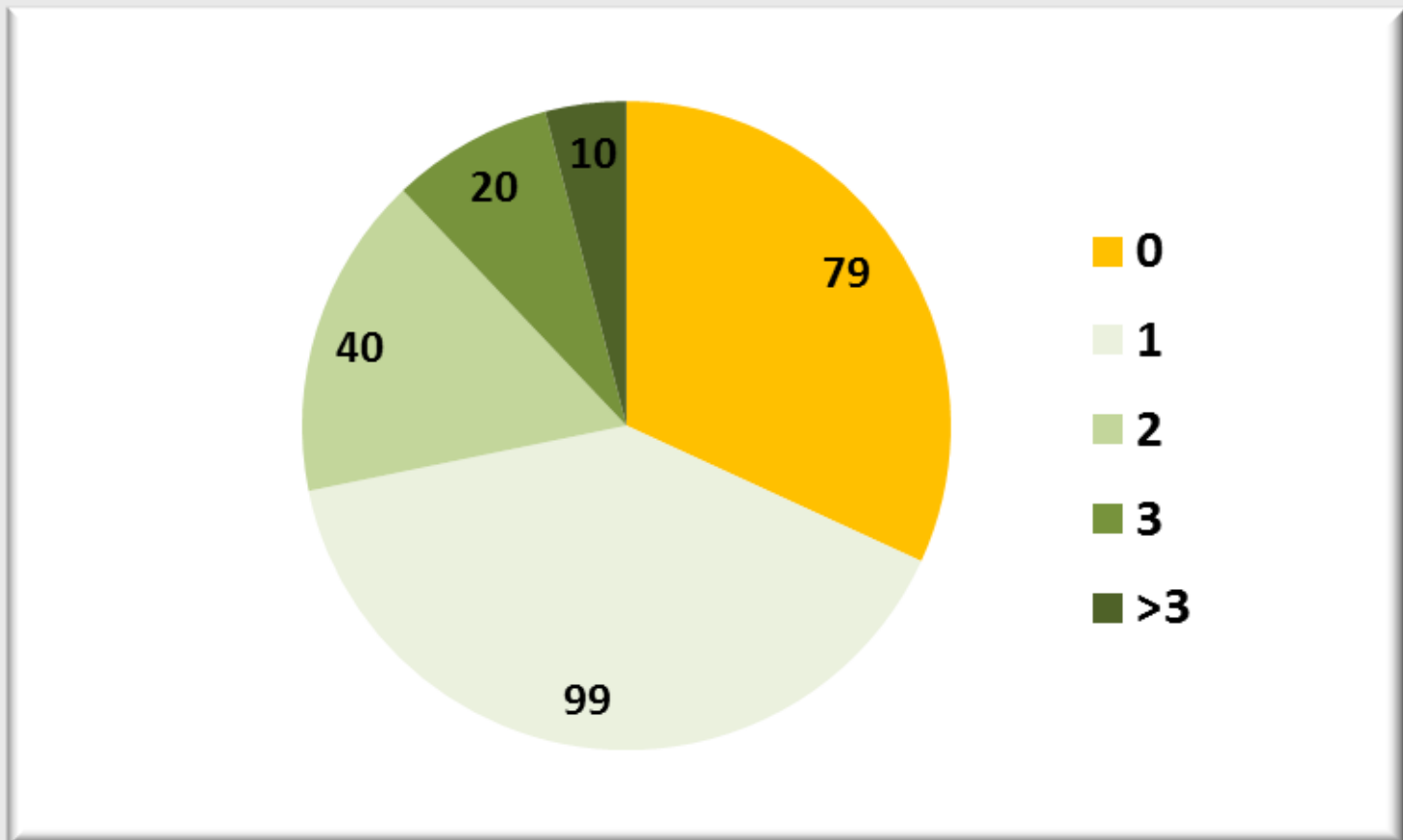




# Homogenization - results

- 1st version: 245 stations and 307 detected breaks
- 2nd version (cleanup): 248 stations and 222 detected breaks

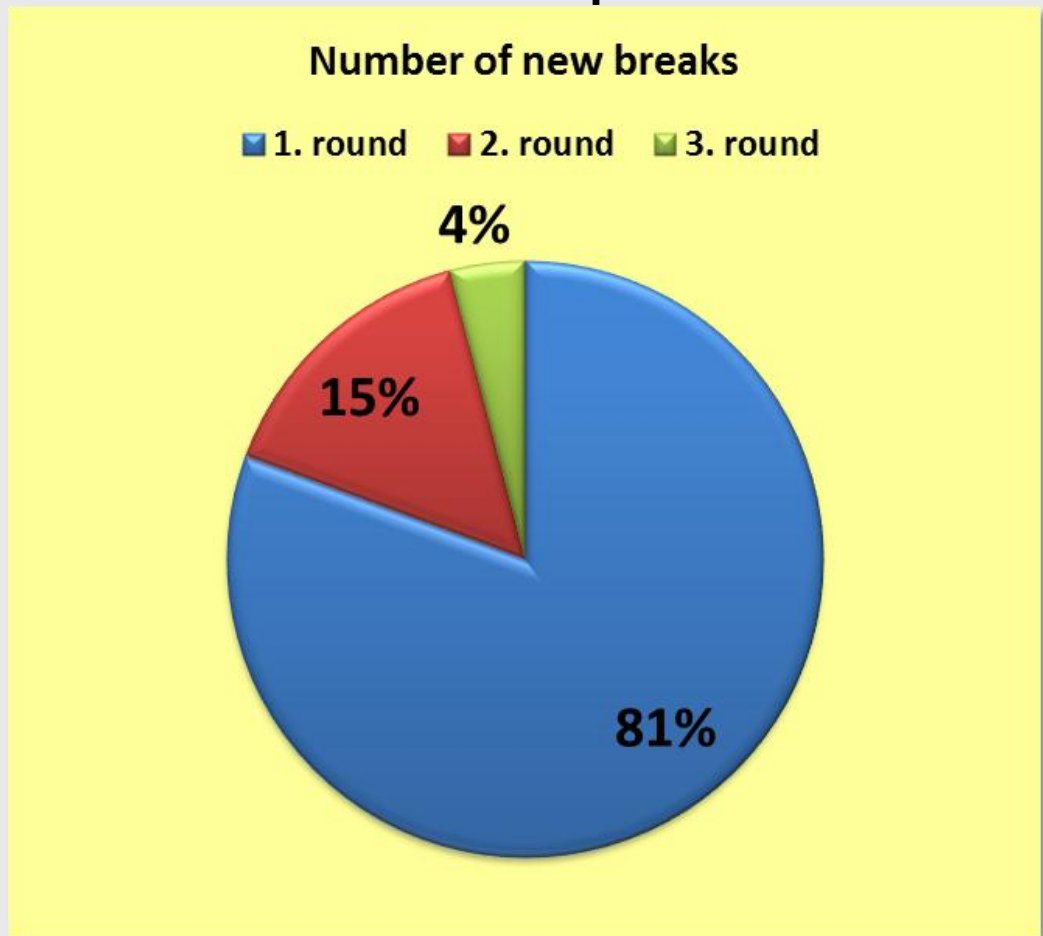
**Number of stations with given number of breaks**



# Homogenization - results

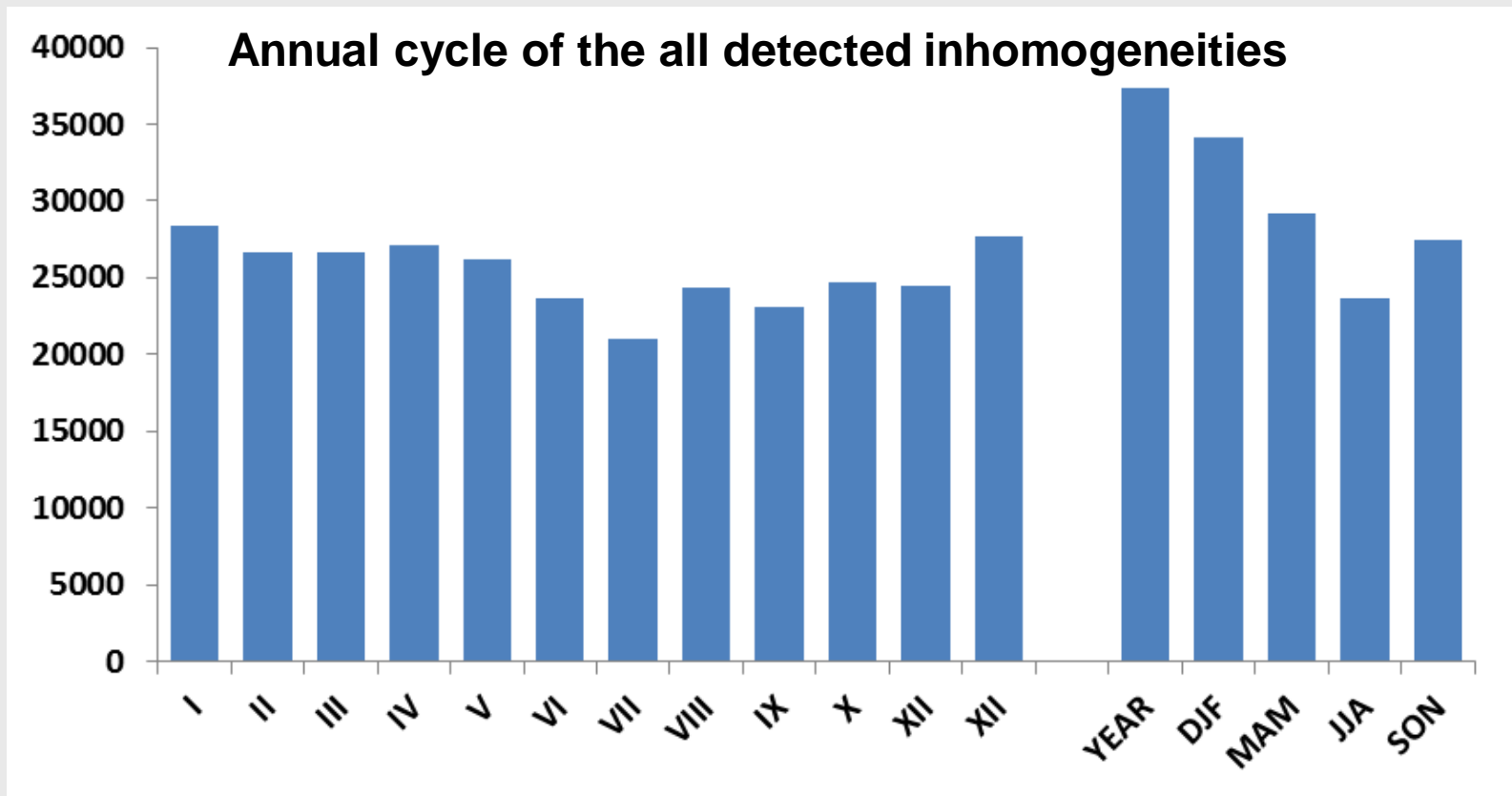
- 1. version: 245 stations and 307 detected breaks
- 2. version (cleanup): 248 stations and 222 detected breaks

## Number of breaks found per iteration



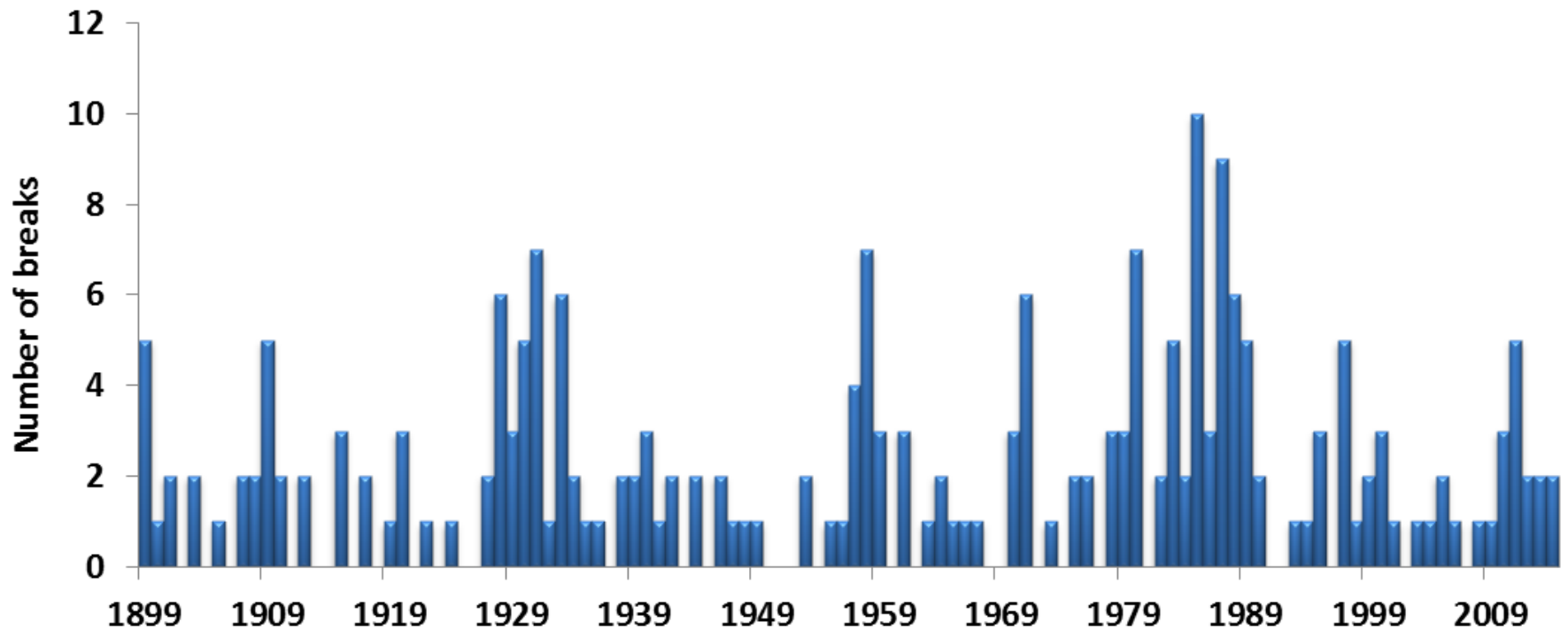
# Inhomogeneities detection

- Large number of detections (tens of thousands)
- It is necessary to establish threshold – significance (empirically or through testing)
- Most inhomogeneities are detected in winter months and in annual values



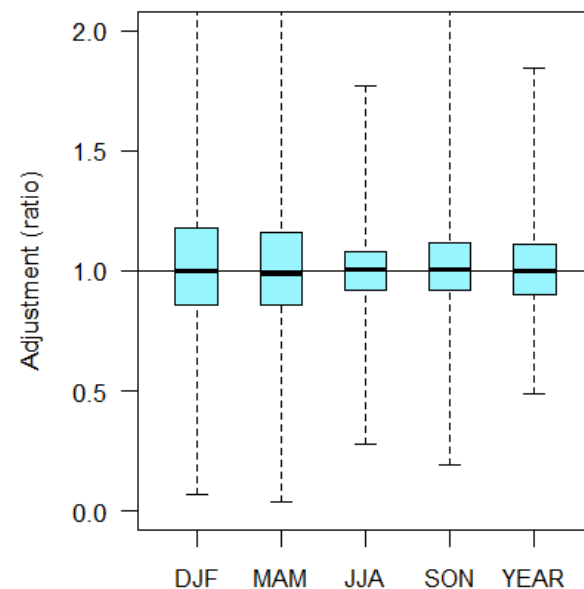
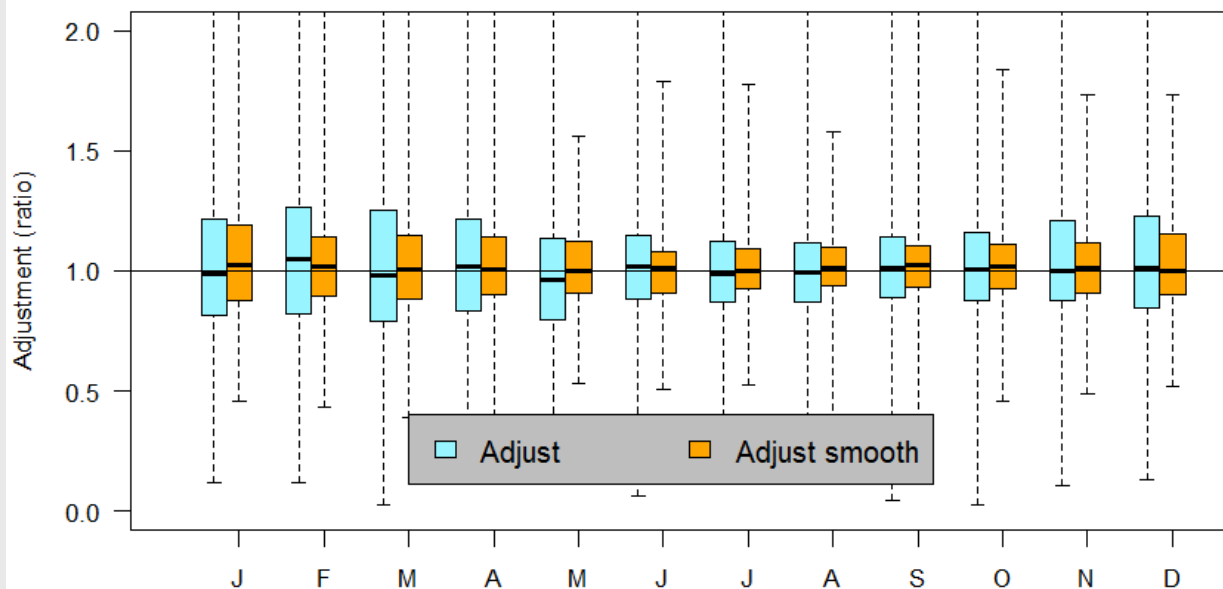
# Inhomogeneities detection

Number of breaks in individual years



# Inhomogeneities adjustment

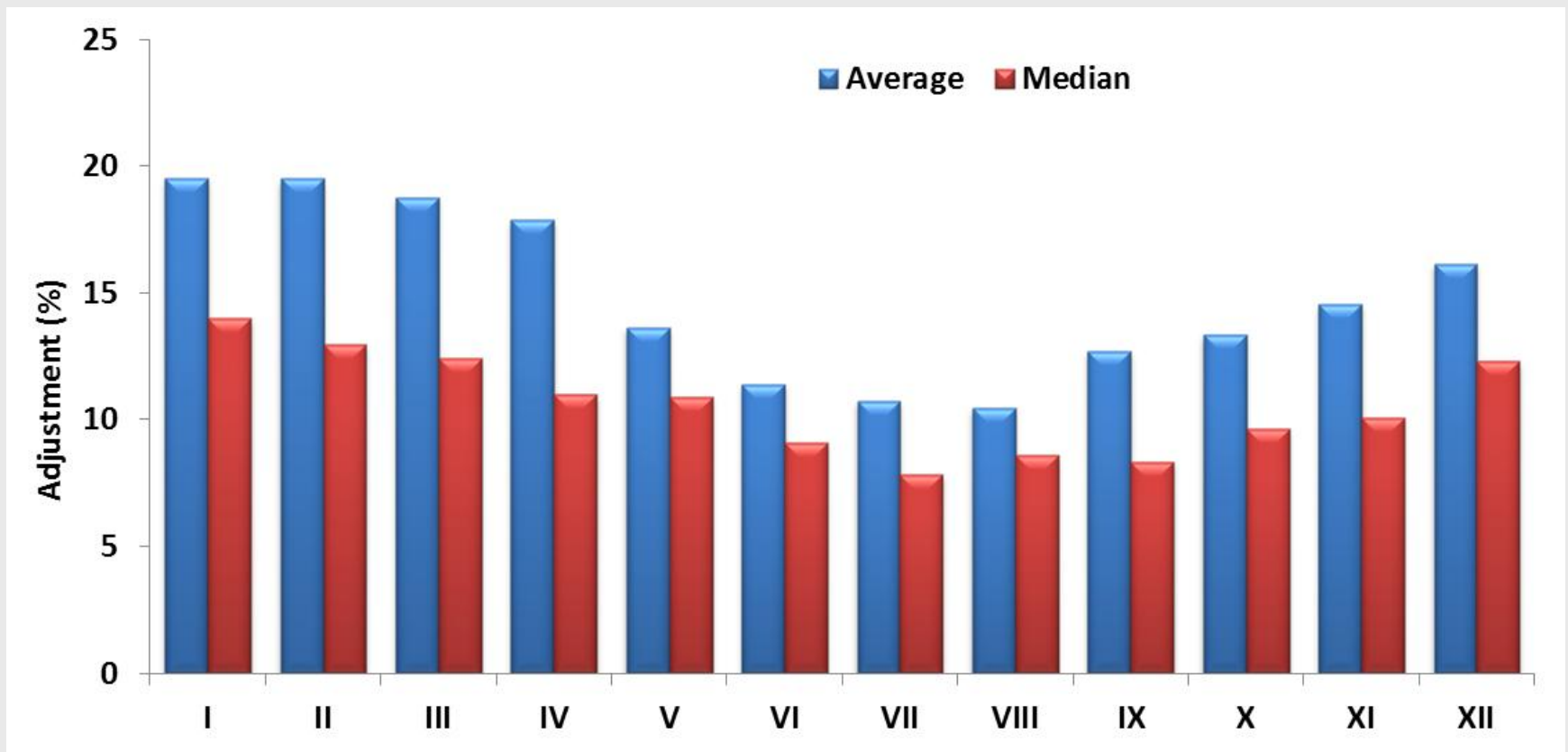
- Reference series calculated from 5 neighbours best correlated or nearest stations
- Correction factor is calculated from differences 20 years before and after breaks
- Final correction factor is smooth by Gauss low pass filter for individual month
- Normally we use own method for correcting daily data, in this phase we used only monthly data



# Inhomogeneities adjustment

- Average absolute size is much higher than median – mainly in winter months
- Average correction factor is 14.9 % and median is 10.6 %
- In winter months correction is two times higher than in summer

**Absolute size of the adjustment (ratio)**

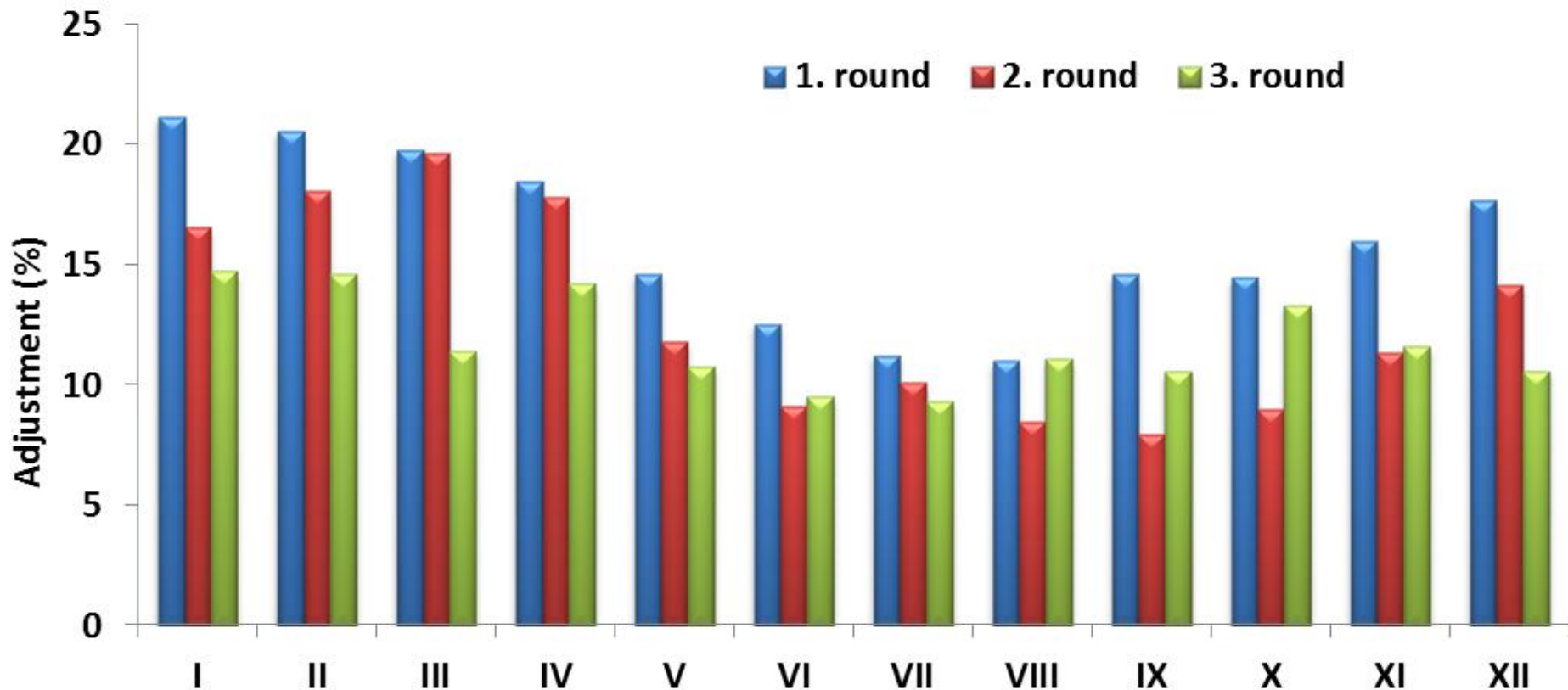




# Inhomogeneities adjustment

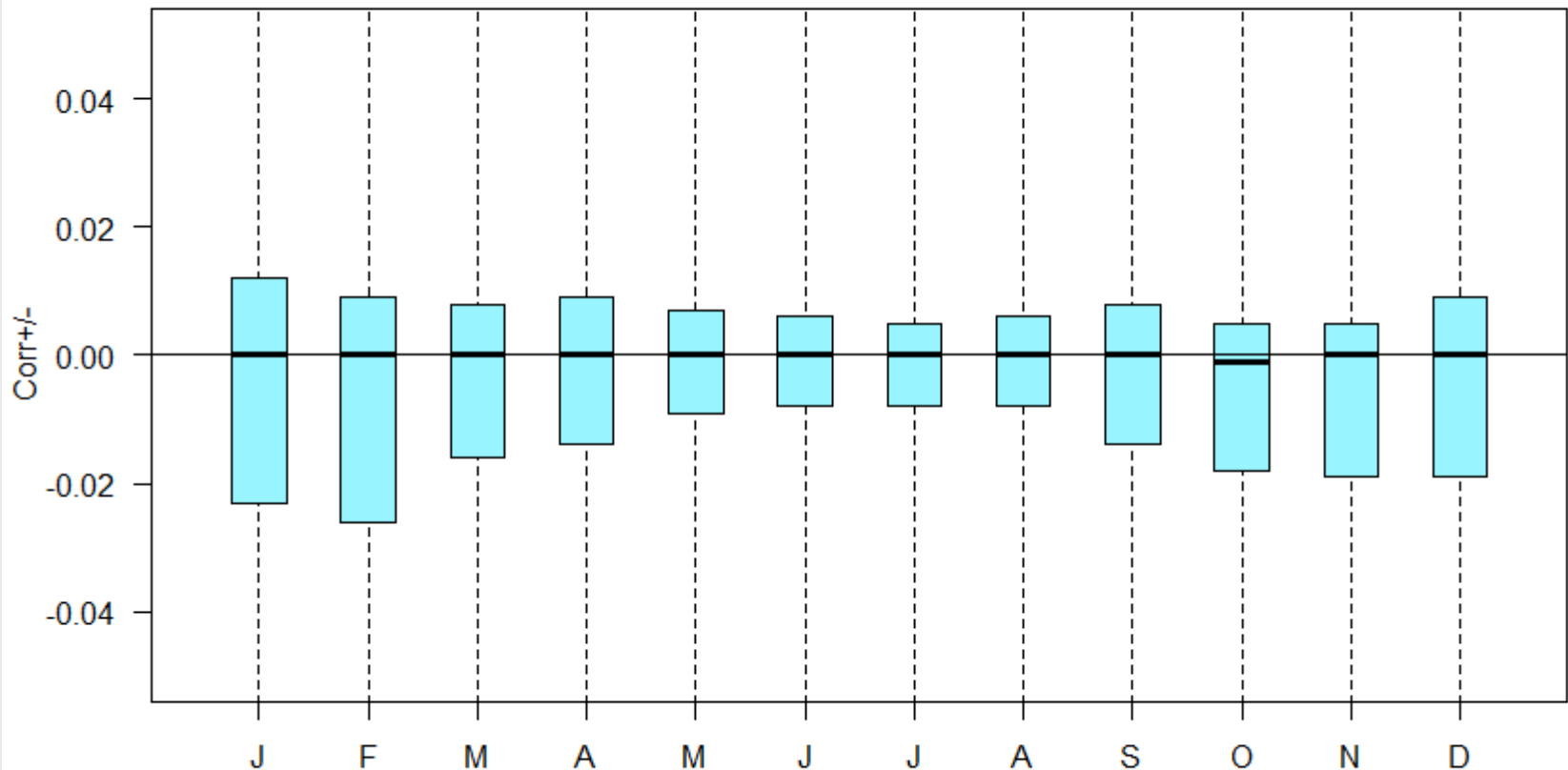
- Correction factor was higher in the 1st iteration of detection/correction (largest breaks were detected in the 1st iteration)
- 1st iteration (16.8 %), 2nd iteration (12.8 %), 3rd iteration (11.8 %)

**Absolute size of the adjustment (ratio)**

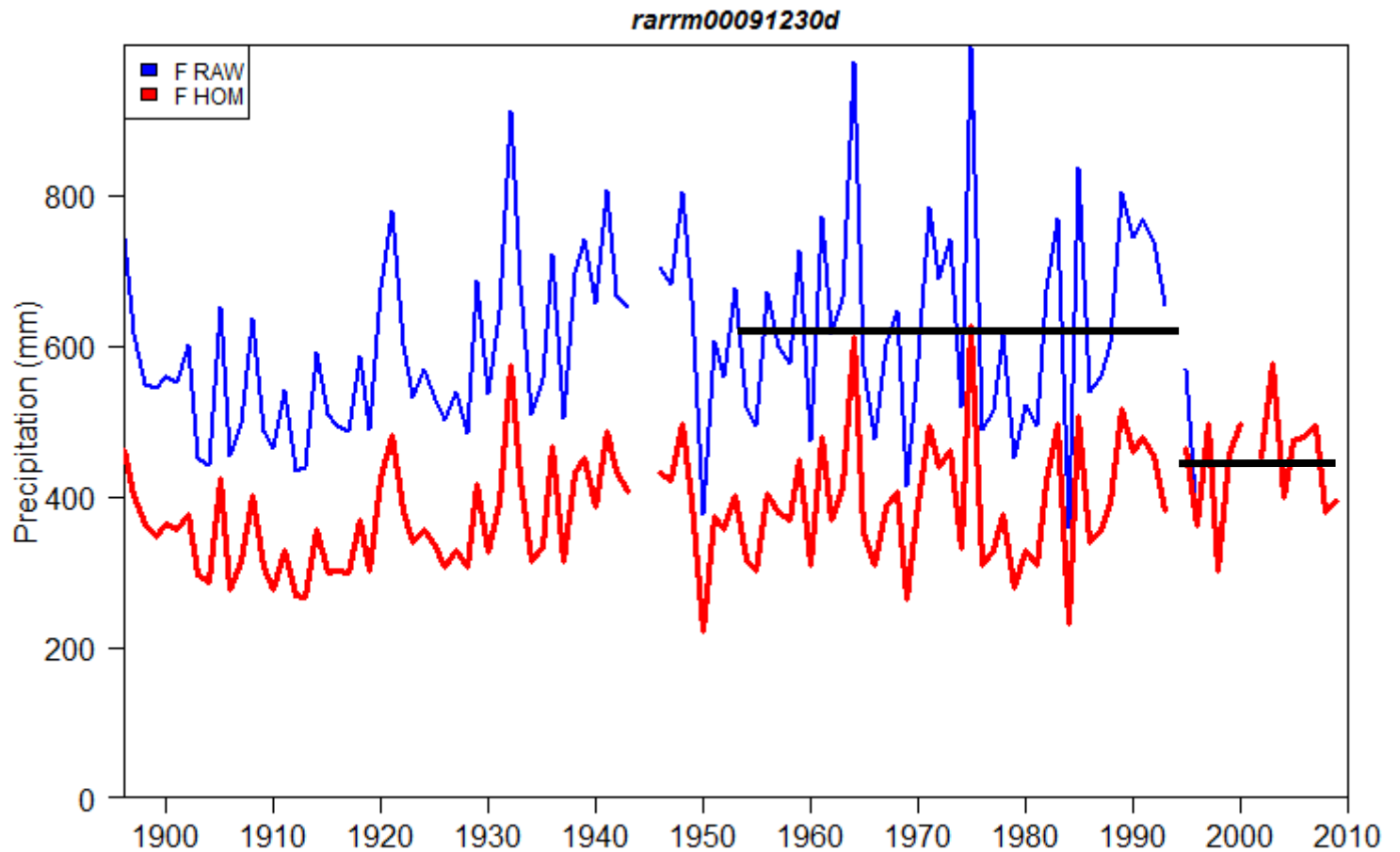


# Inhomogeneities adjustment

Change of the correlations after adjustment (in case of negative correlations, the series were not adjusted)



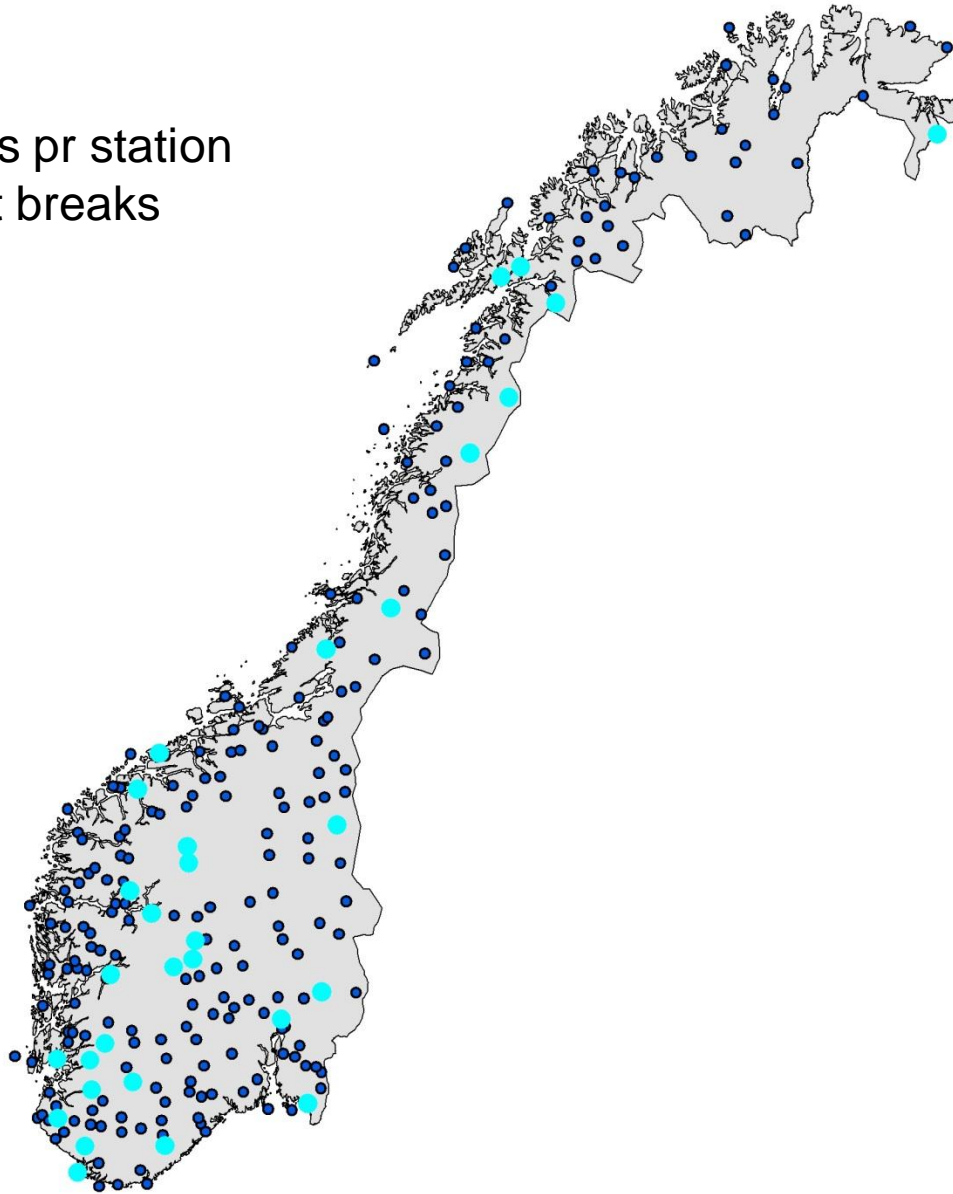
# Homogenization – new series



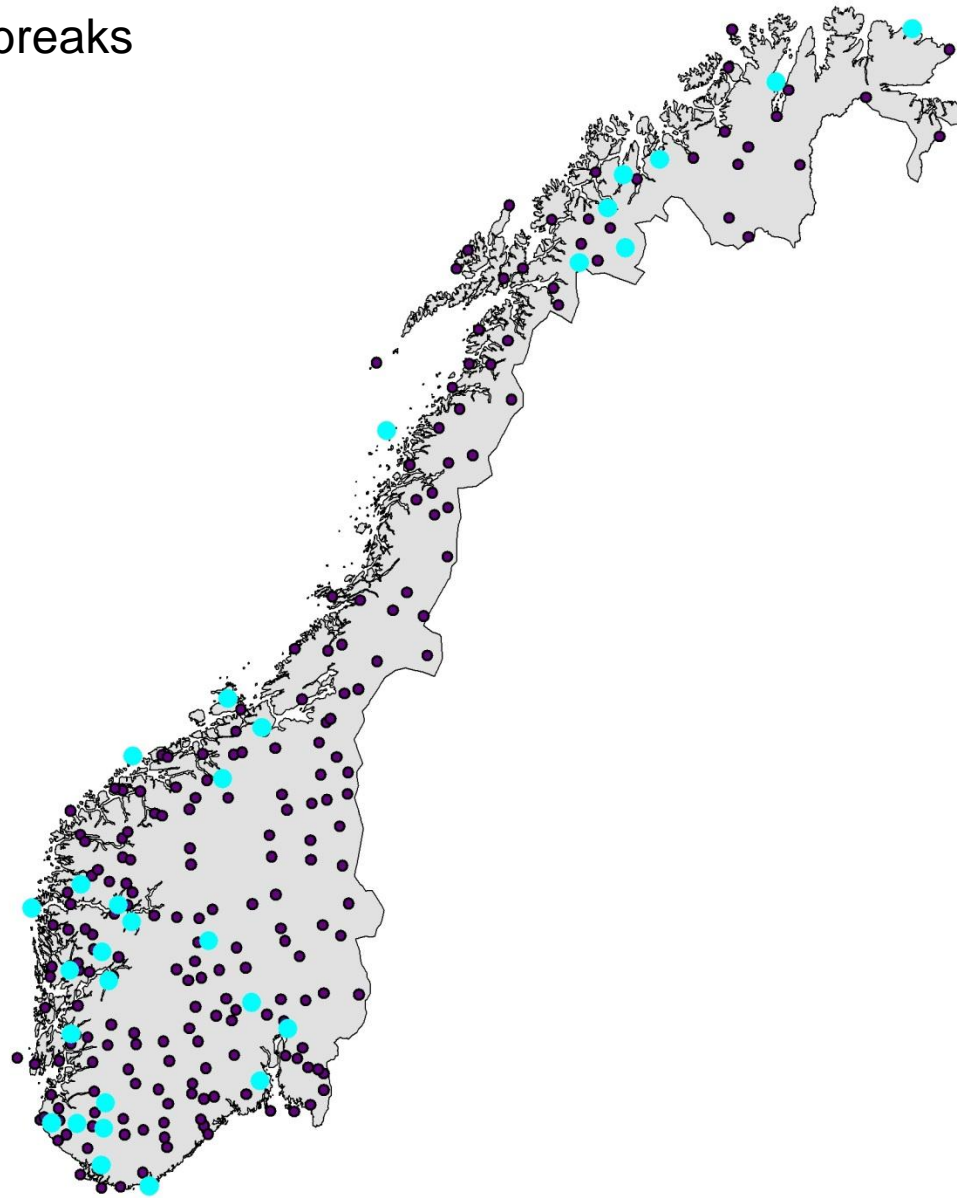
# MASH homogenization results

## 1314 Breaks

- Average 5,2 breaks pr station
- 30 stations without breaks



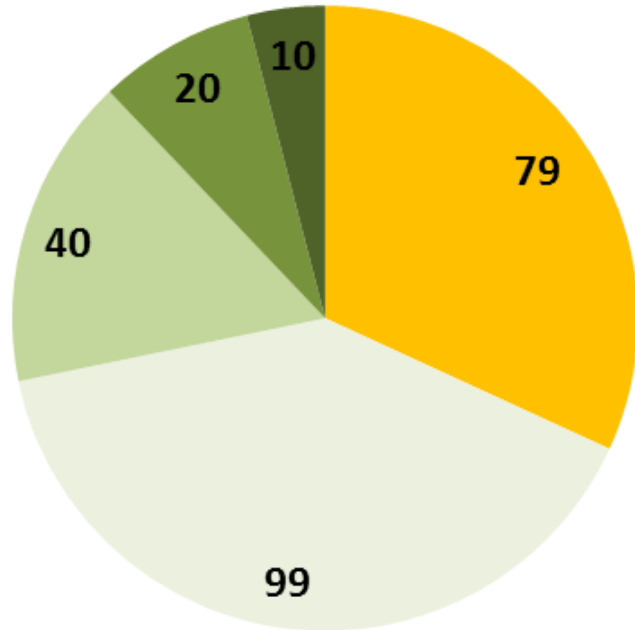
- 30 stations: > 11 breaks



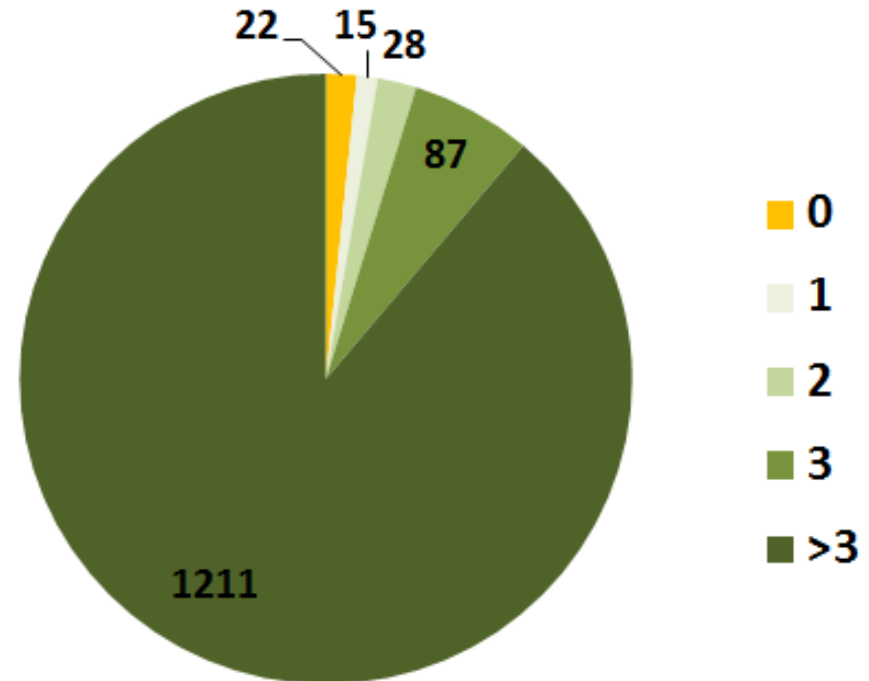
# Comparison with MASH results

Number of stations with given number of breaks

ProClimDB



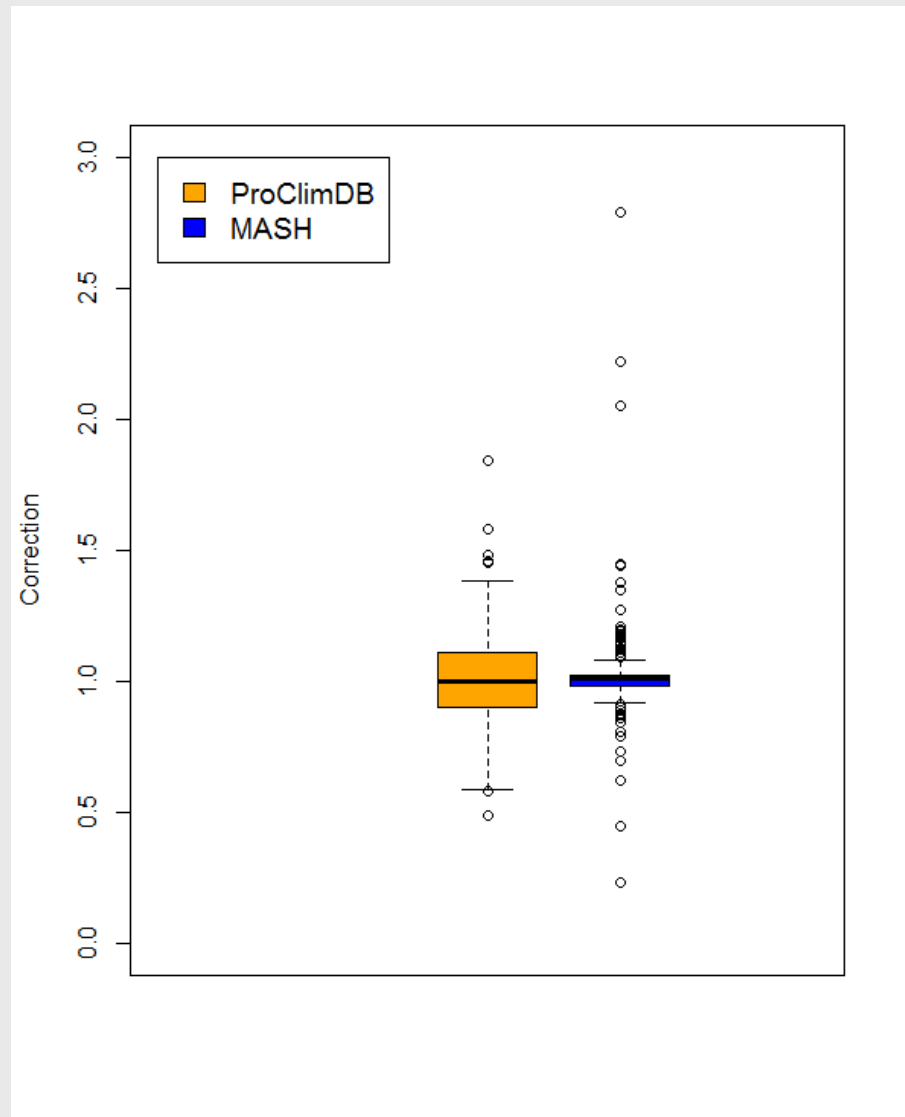
MASH



# Comparison with MASH results

## Inhomogeneities adjustment (correction factor)

- ProClimDB:
  - Average is 14.9 %
  - median is 10.6 %
- MASH (six times more breaks):
  - Average is 3,9 %
  - median is 2.0%





# Lesson Learned

Won knowledge and experience with state-of-the-art algorithms for homogeneity testing (Home.R, MASH)

The algorithms are essentially suitable, but must be interpreted and compared carefully with metadata series.

Important to analyze annual, seasonal and monthly values.

Benefit to apply various algorithms, provide more robust detection of violations.

Homogenization of daily values are important (tails of the frequency distribution and extreme values) – it will be the next step



Norwegian  
Meteorological  
Institute



# Thanks for the attention!

Telephone: +47 22 96 31 69

E-mail: [elinl@met.no](mailto:elinl@met.no)

Twitter: @nile\_2011