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

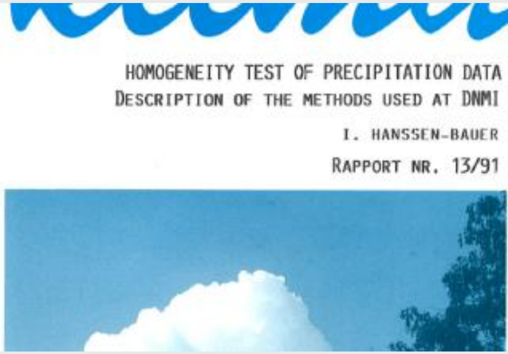
- Homogenization results applying SNHT, Home.R and MASH

- New results (for precipitation):
  - November 2016, the first version
  - March 2017, cleaned version from main problems (detected in the first version, time shift etc.)

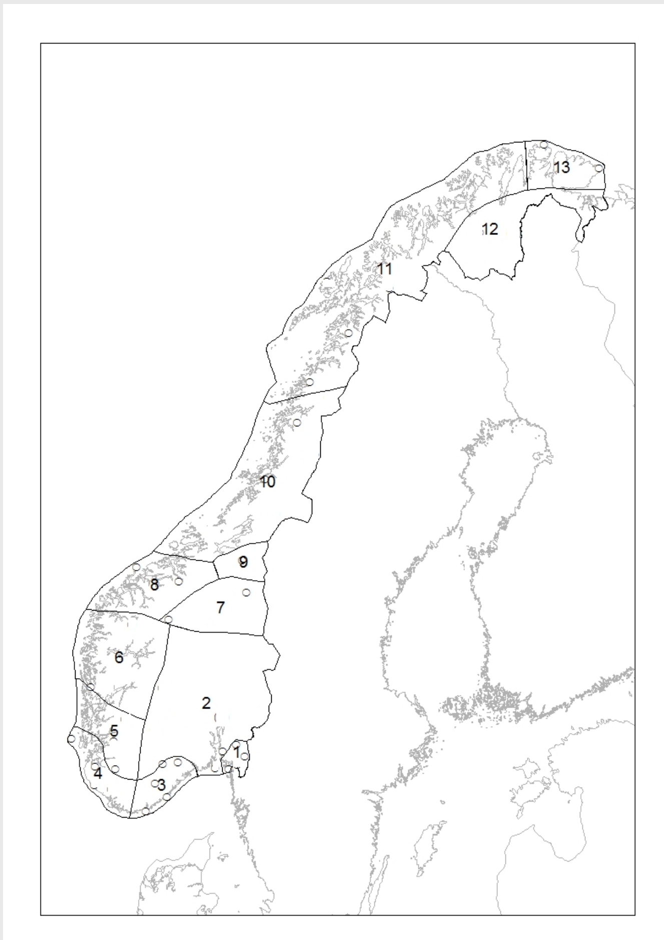
Precipitation homogenization, past

- 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

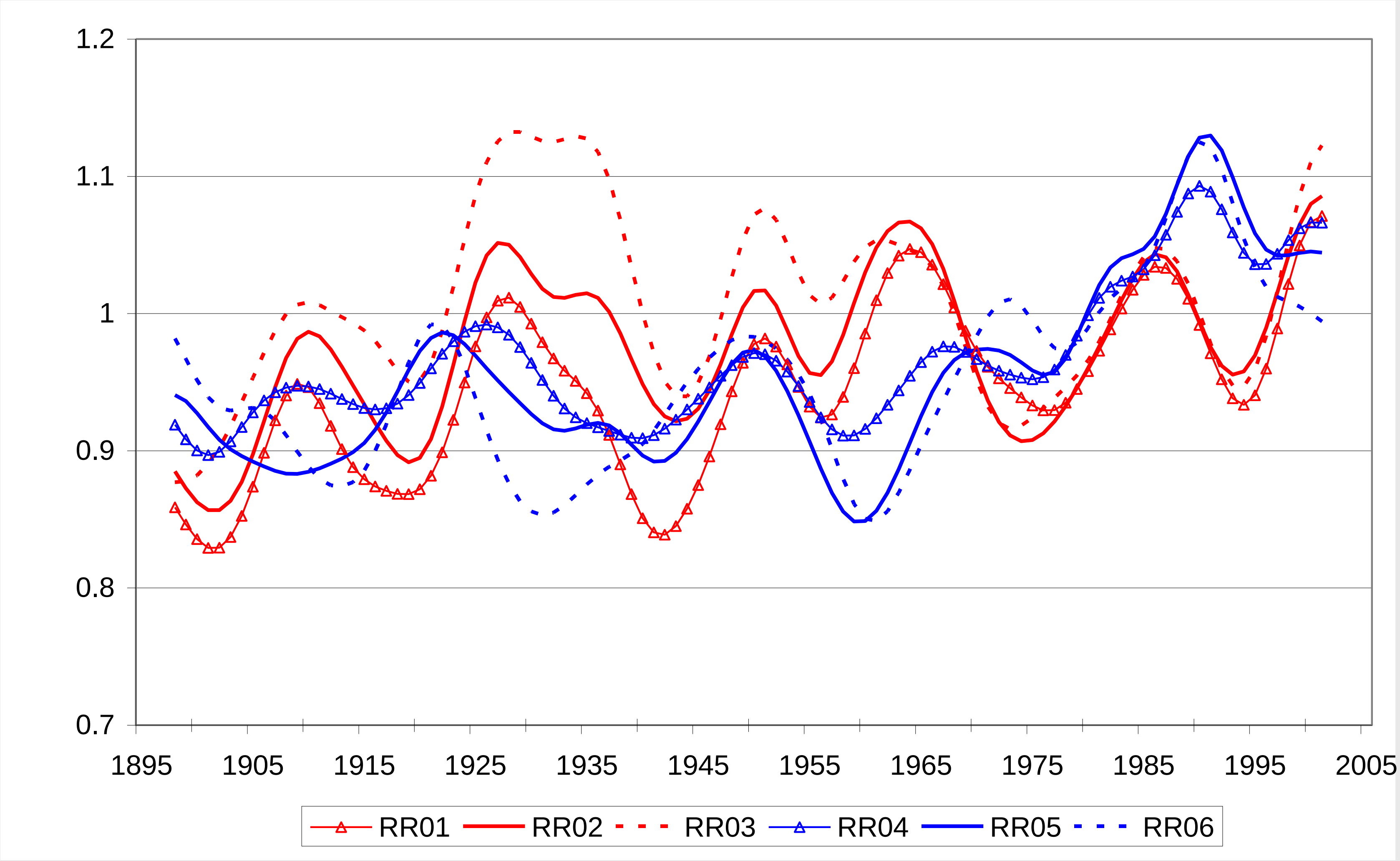


Homogenization of precipitation data in Norway (past)

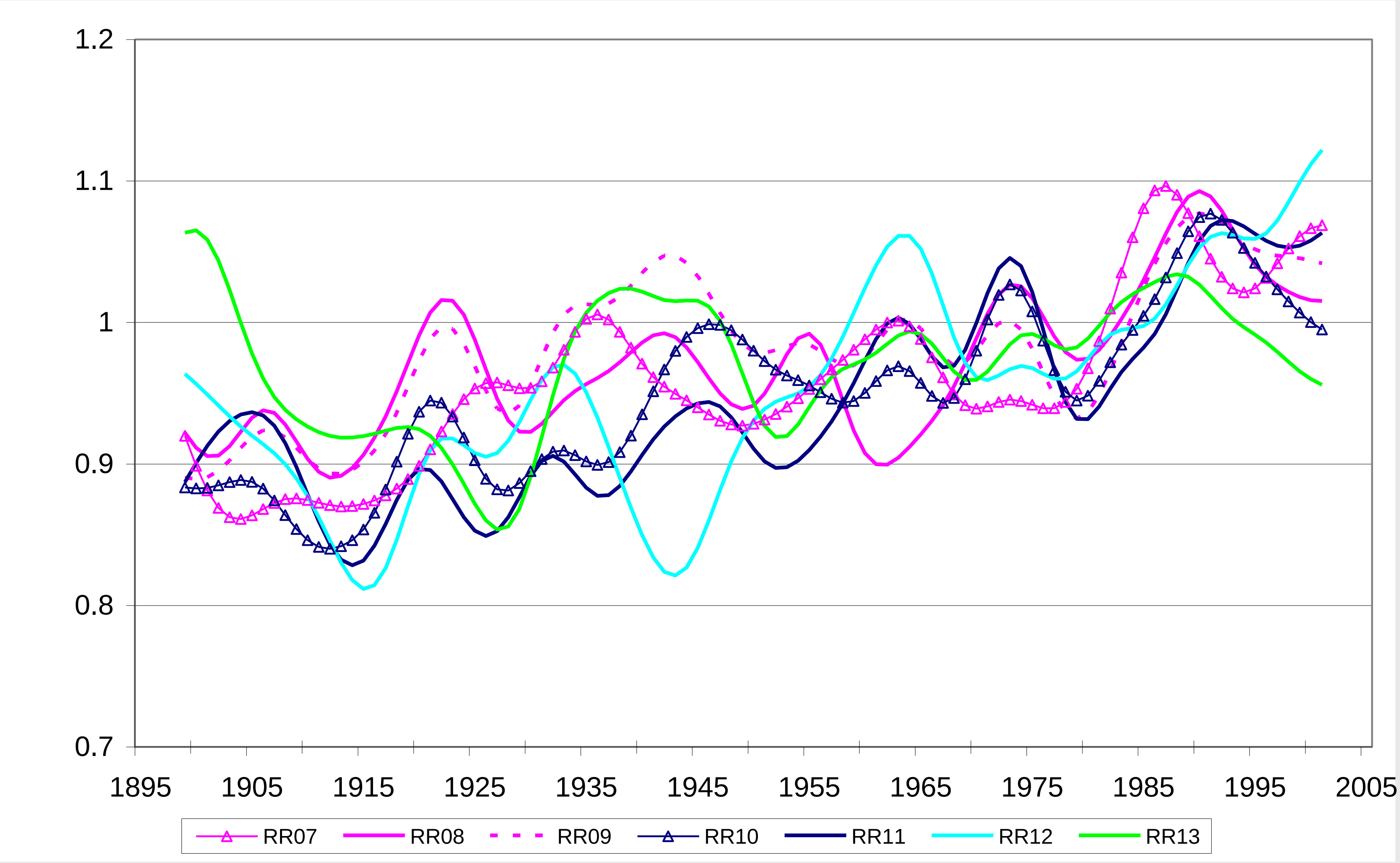


The norwegian precipitation regions

Regional precipitation the last 100 years
Region 1-6



Regional precipitation the last 100 years
Region 7-13

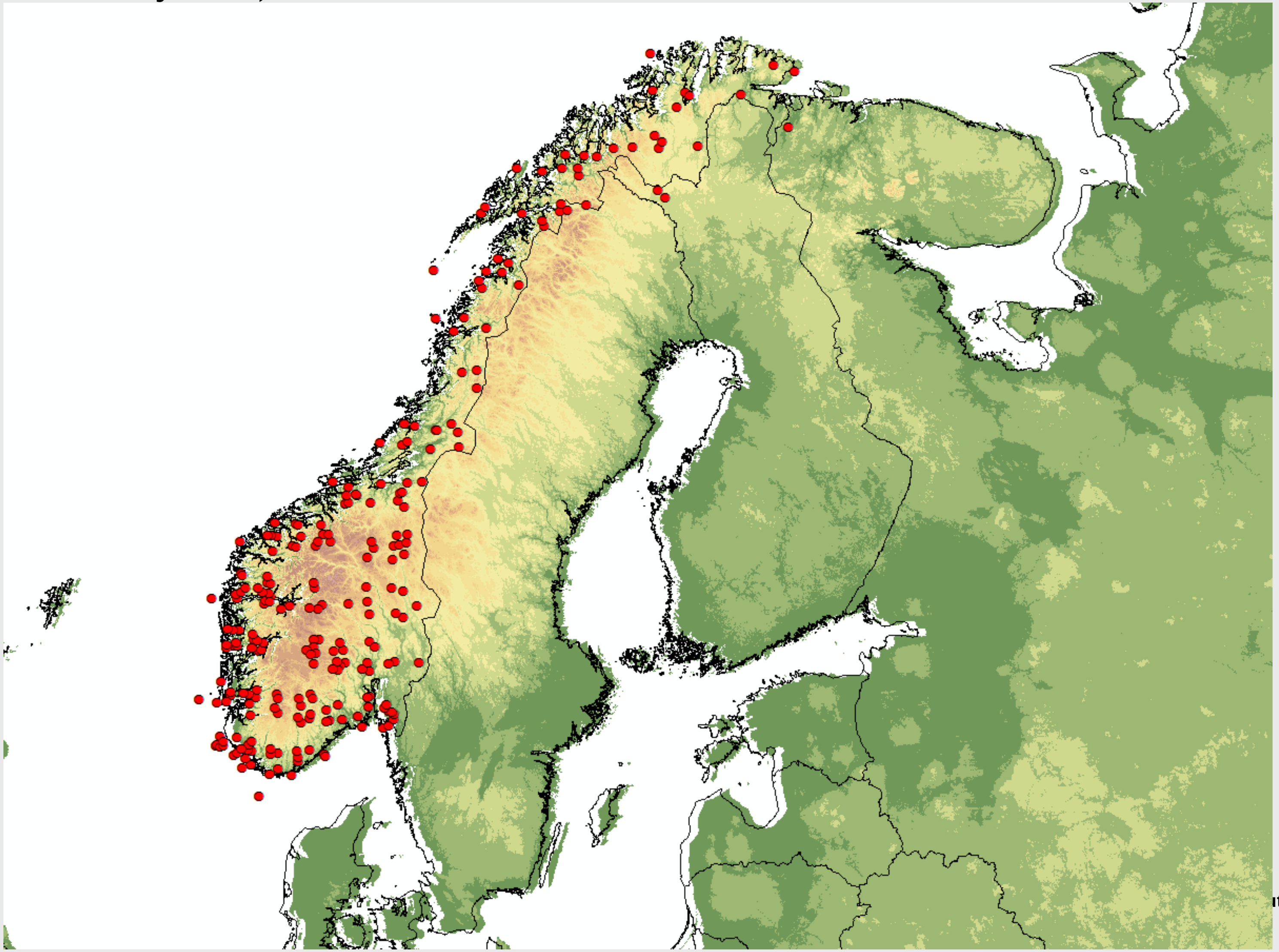


Precipitation homogenization, update

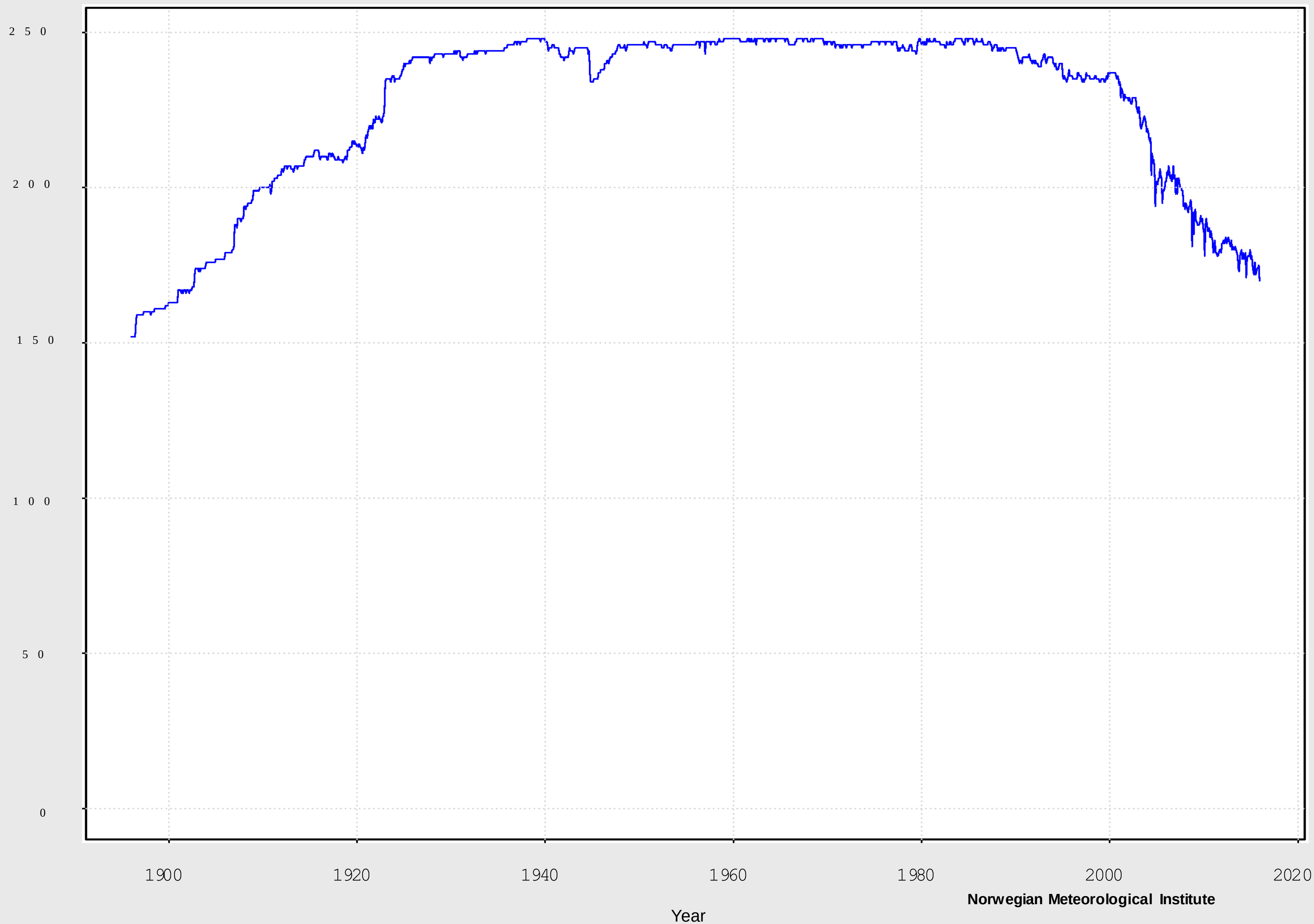
Precipitation measurements

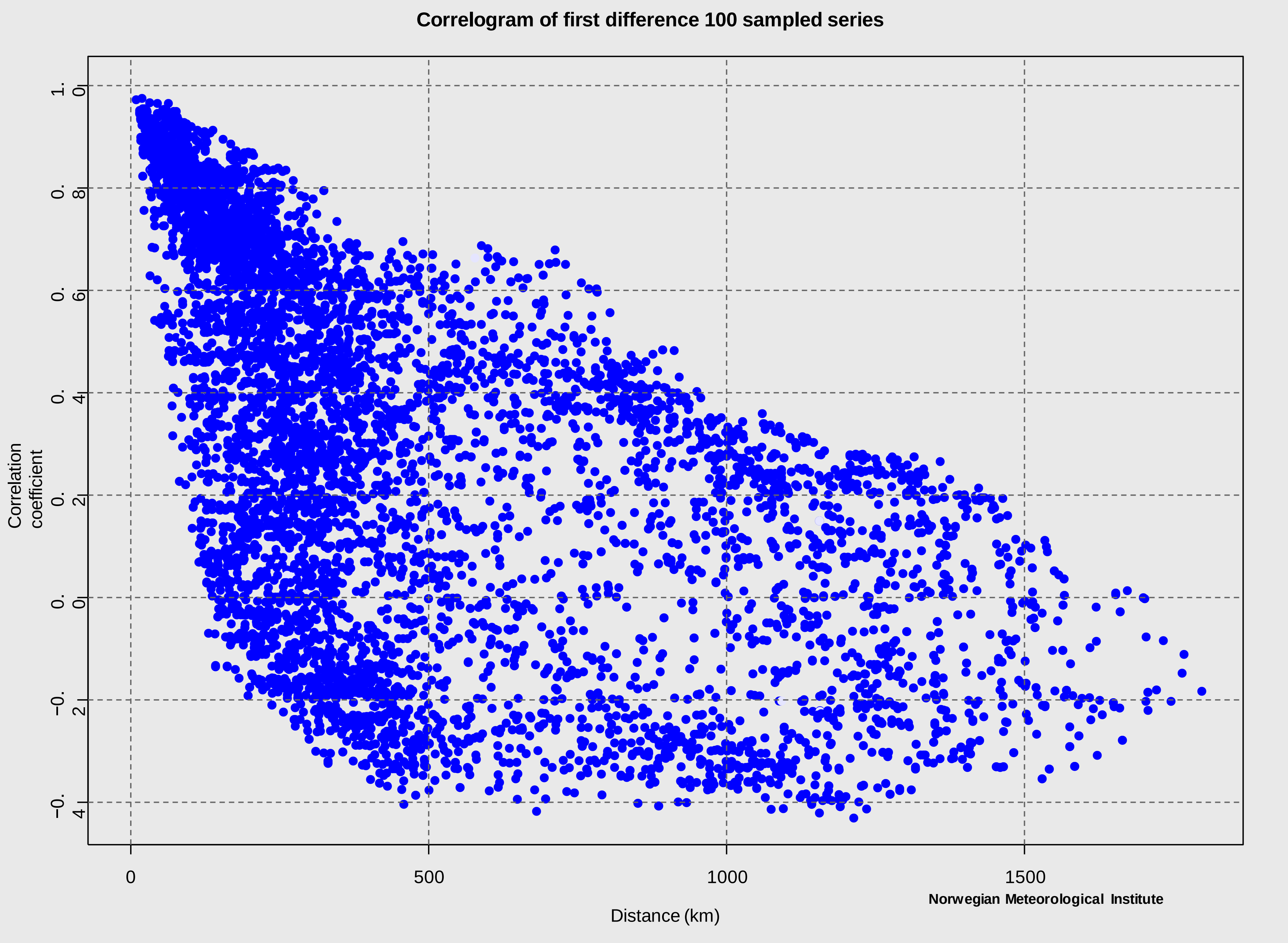
Meteorologisk institutt

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



Precipitation Stations available throught the time

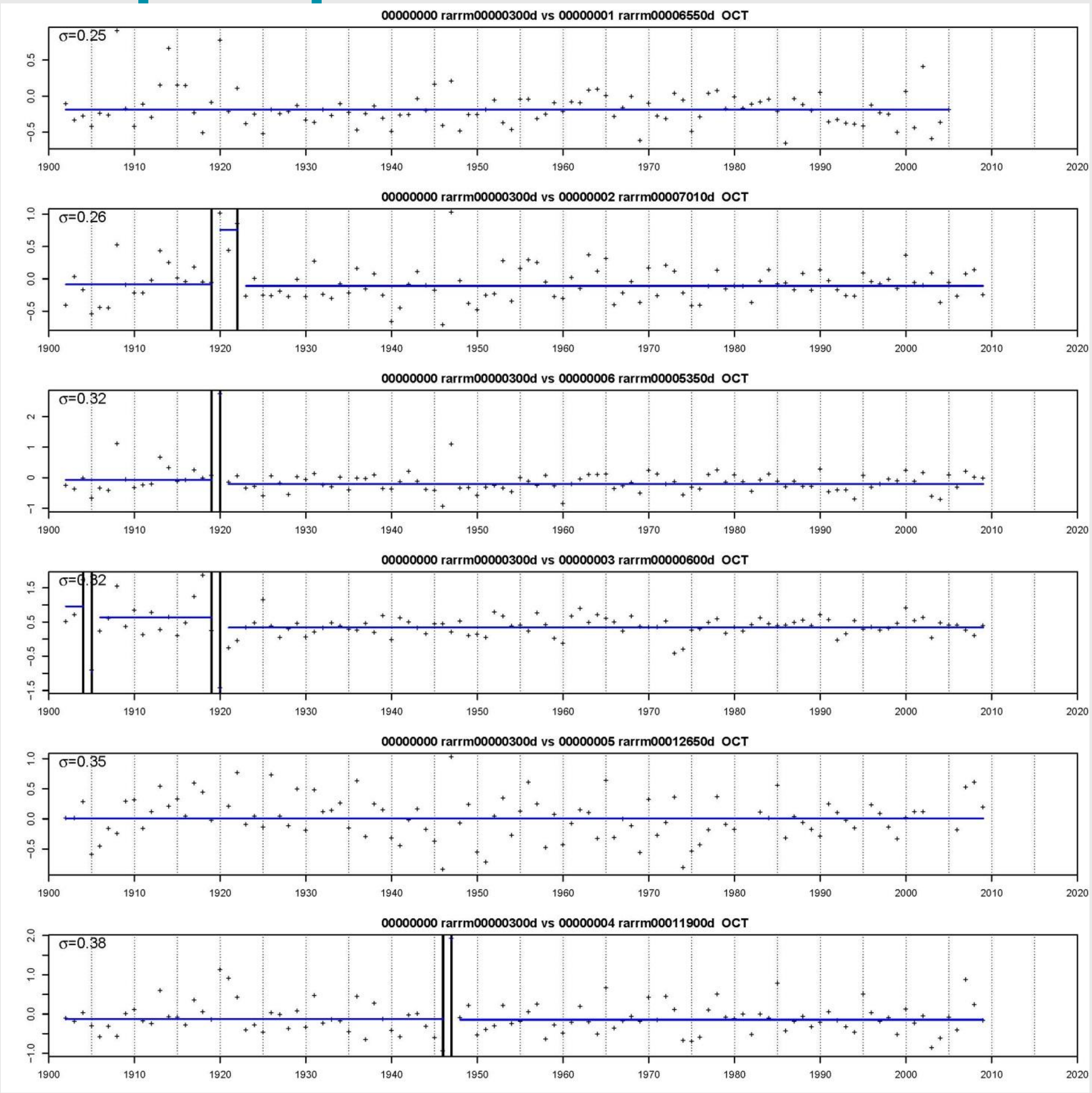



Correlogram of first difference 100 sampled series
Correlation coefficient
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
0
500
1000
1500
Distance (km)
Norwegian Meteorological Institute

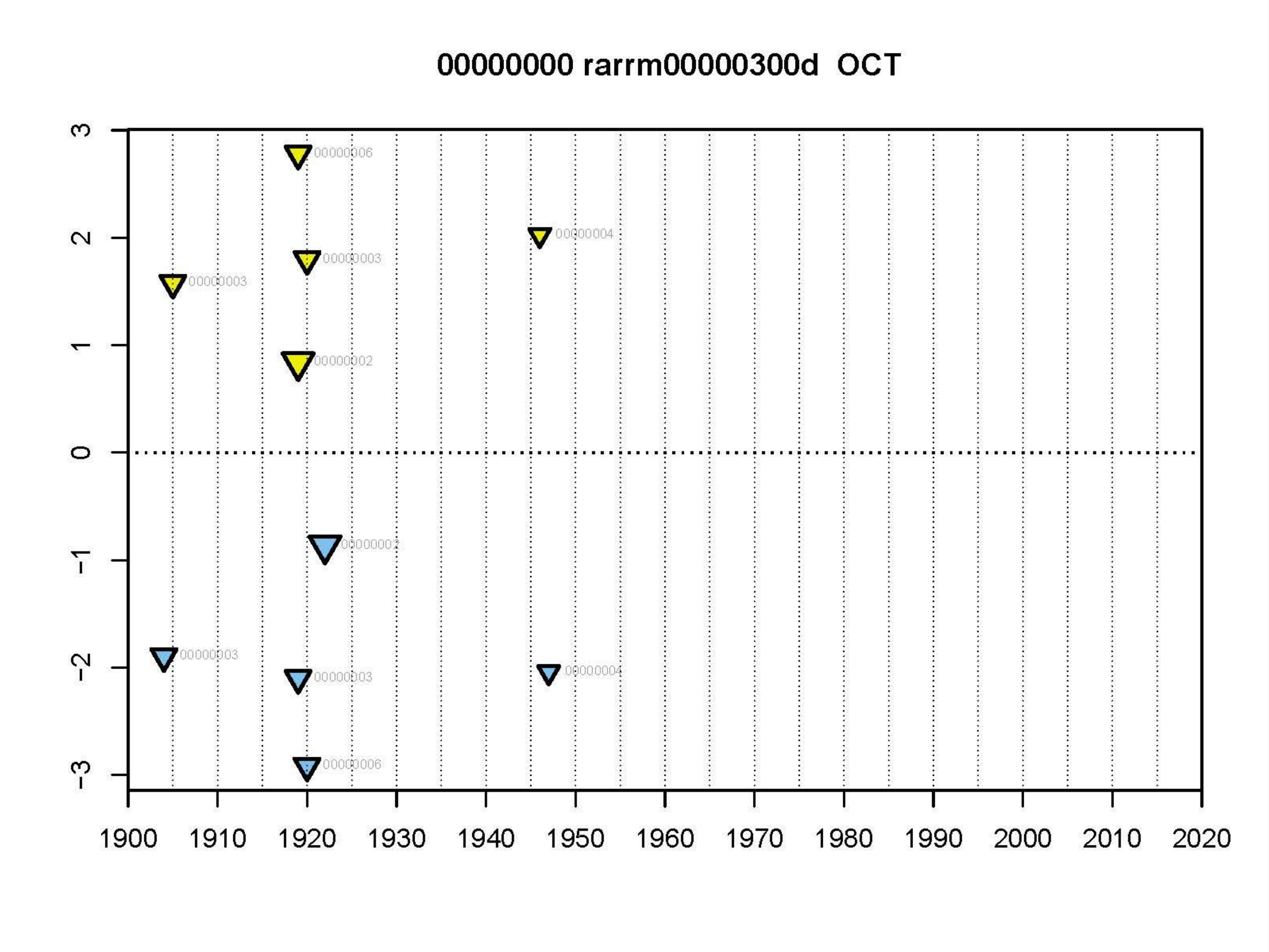
Precipitation homogenization, update

- 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 … ?

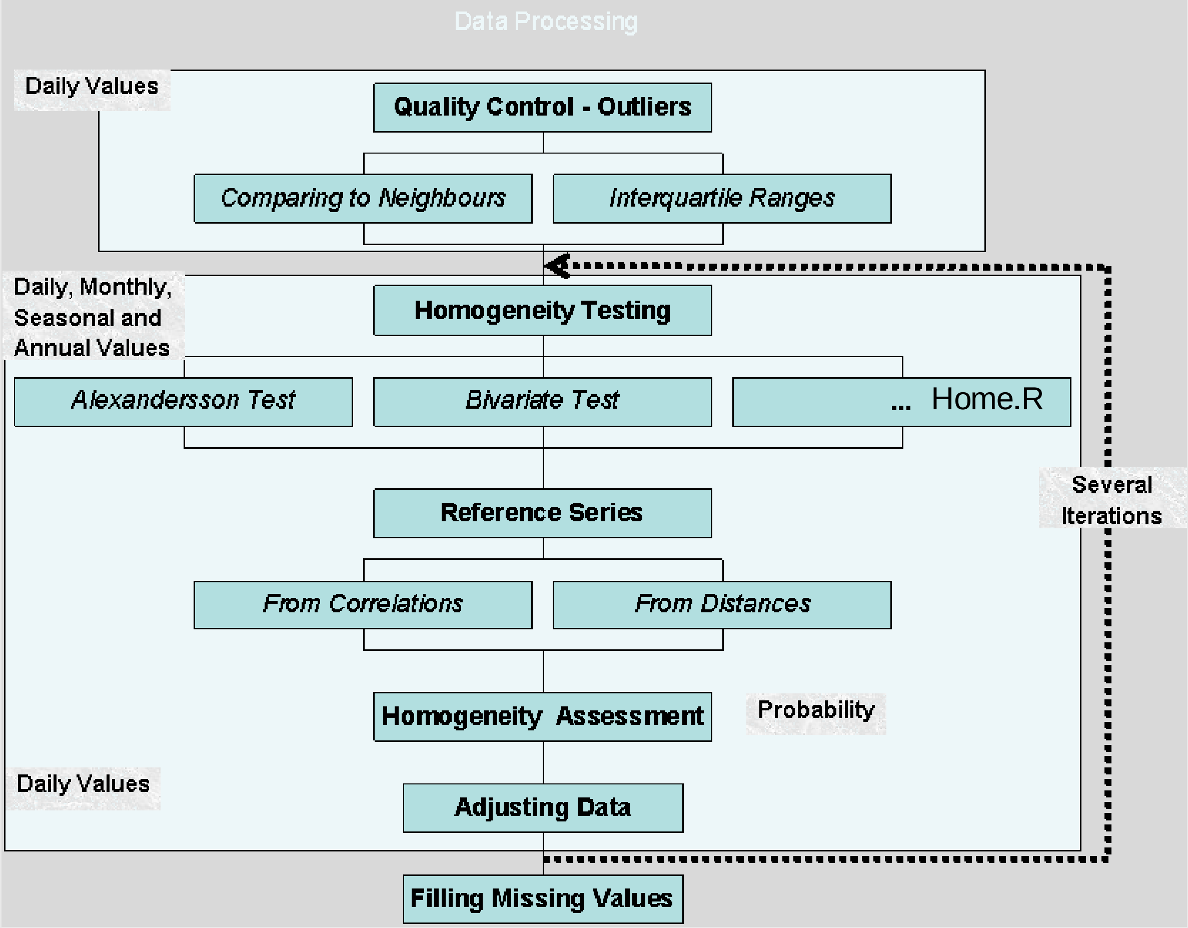


Other Home.R precipitation issues … ?



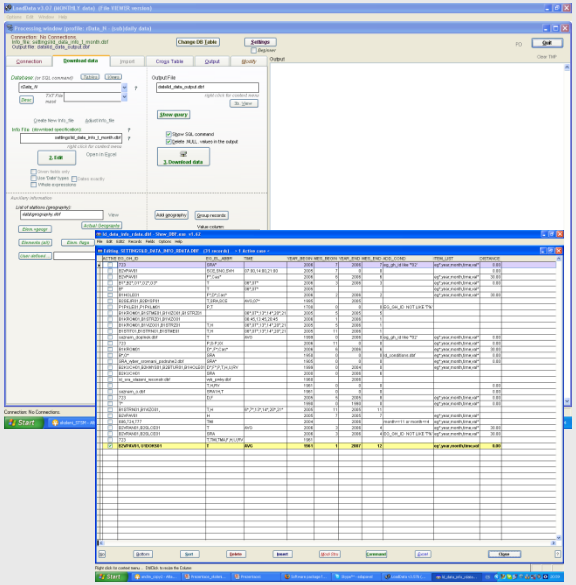
Homogenization using AnClim and ProClimDB

- 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)



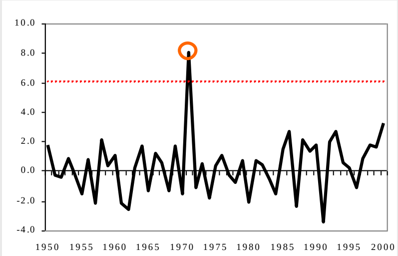
Download data from database (e.g. Oracle)

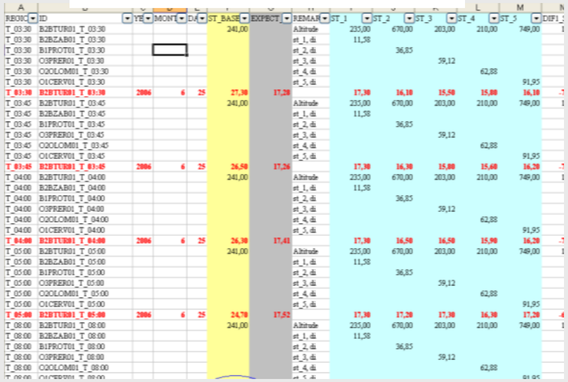
(LoadData)



Quality control

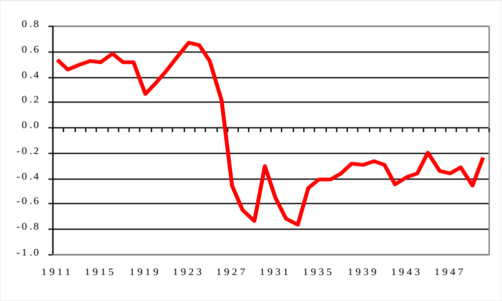
(ProClimDB)

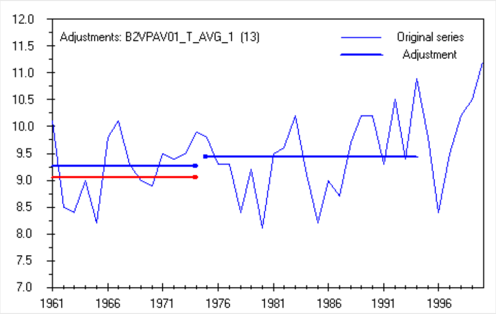




Homogenization

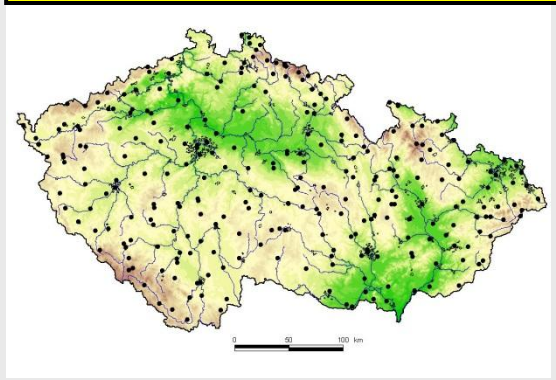
(ProClimDB/AnClim)

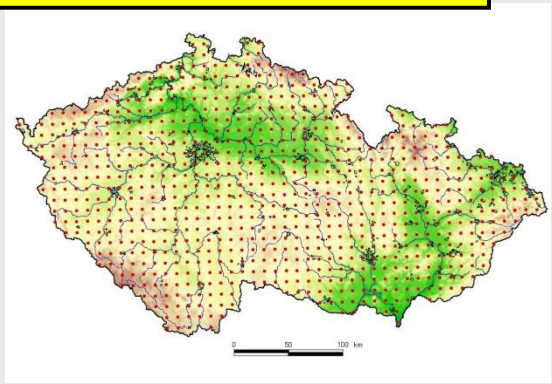




„Technical“ series and grid points calculation

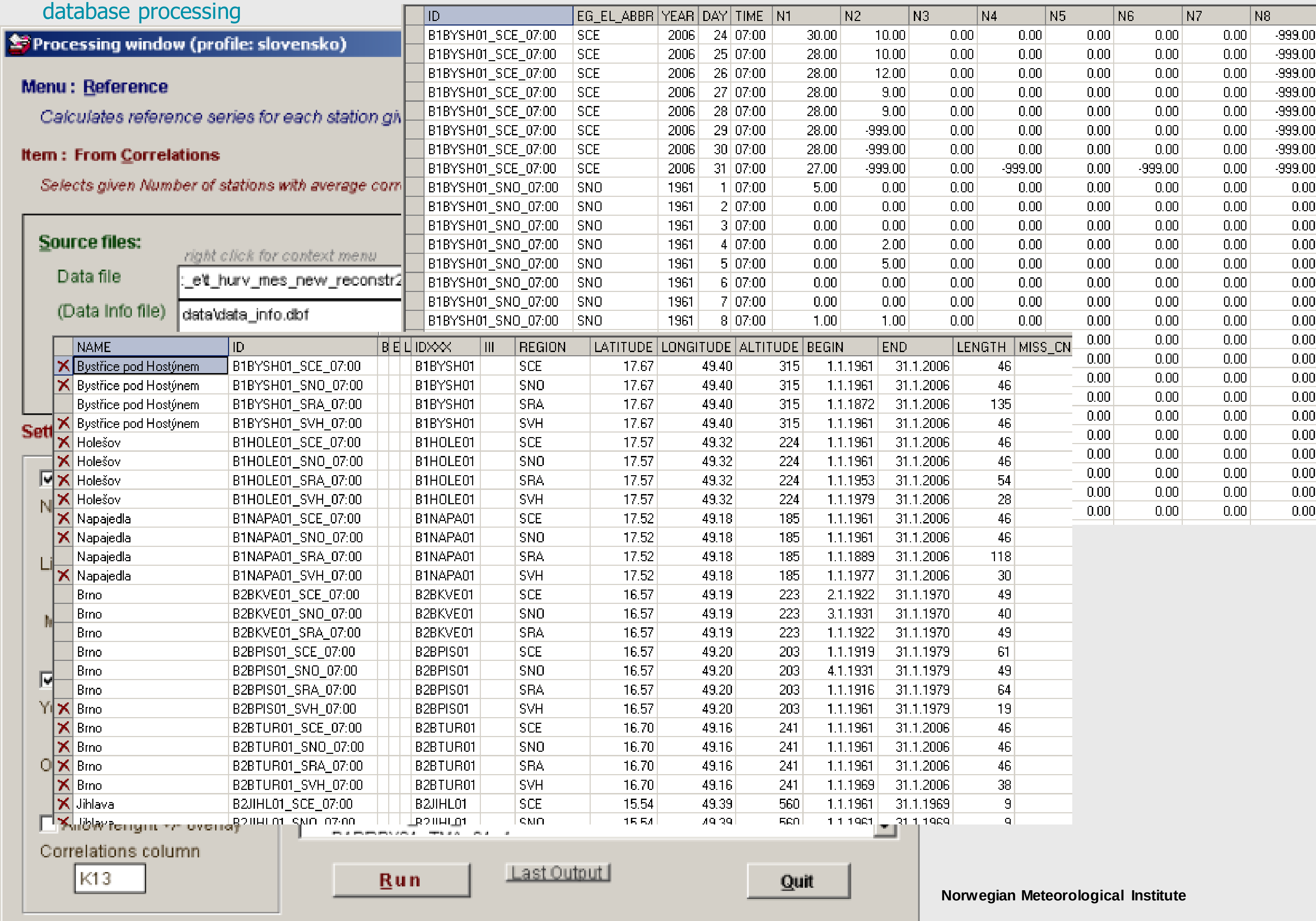
(ProClimDB)





ProcData software, only one Data file, accompanied by Info_file

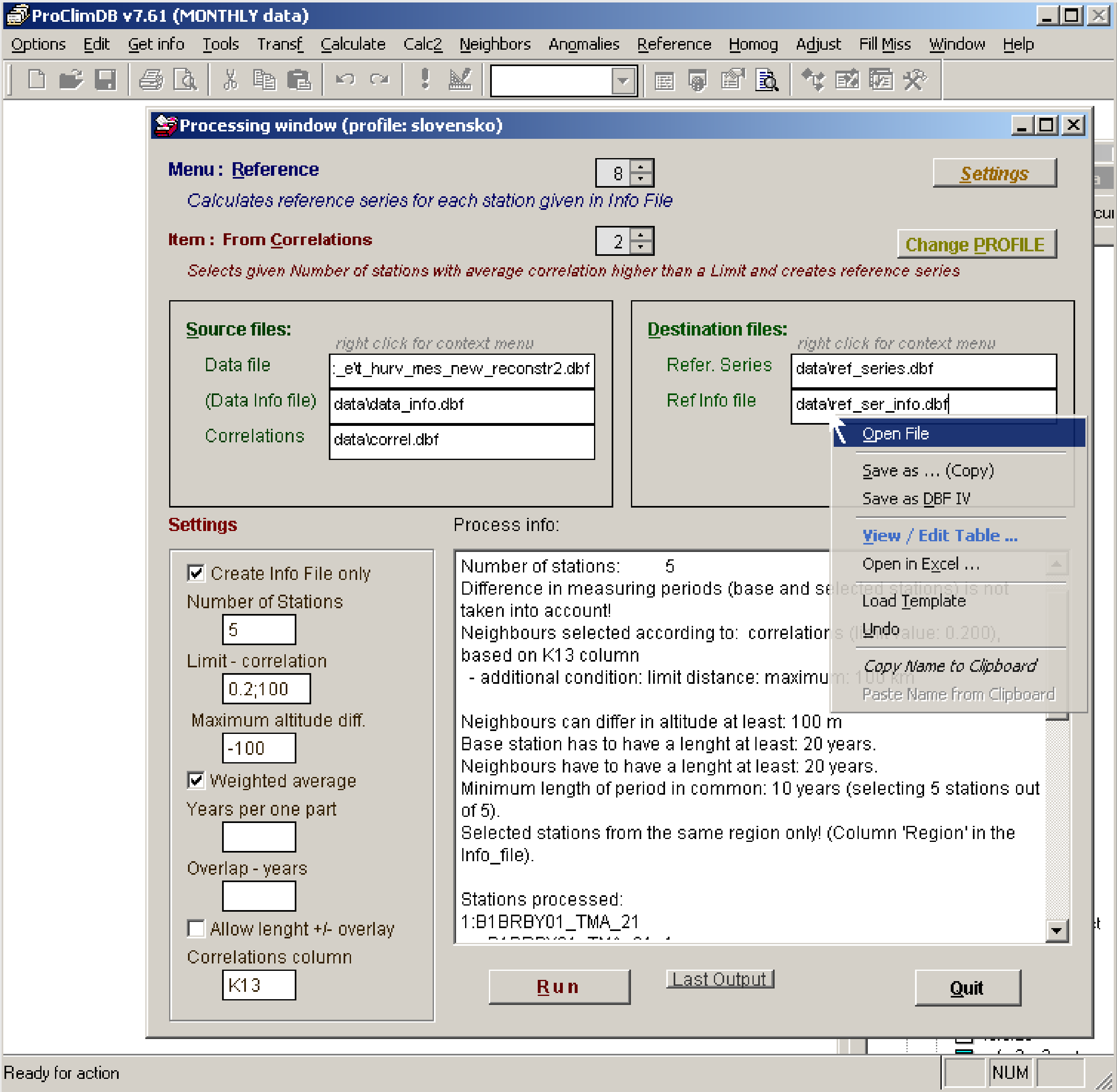
database processing



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
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00

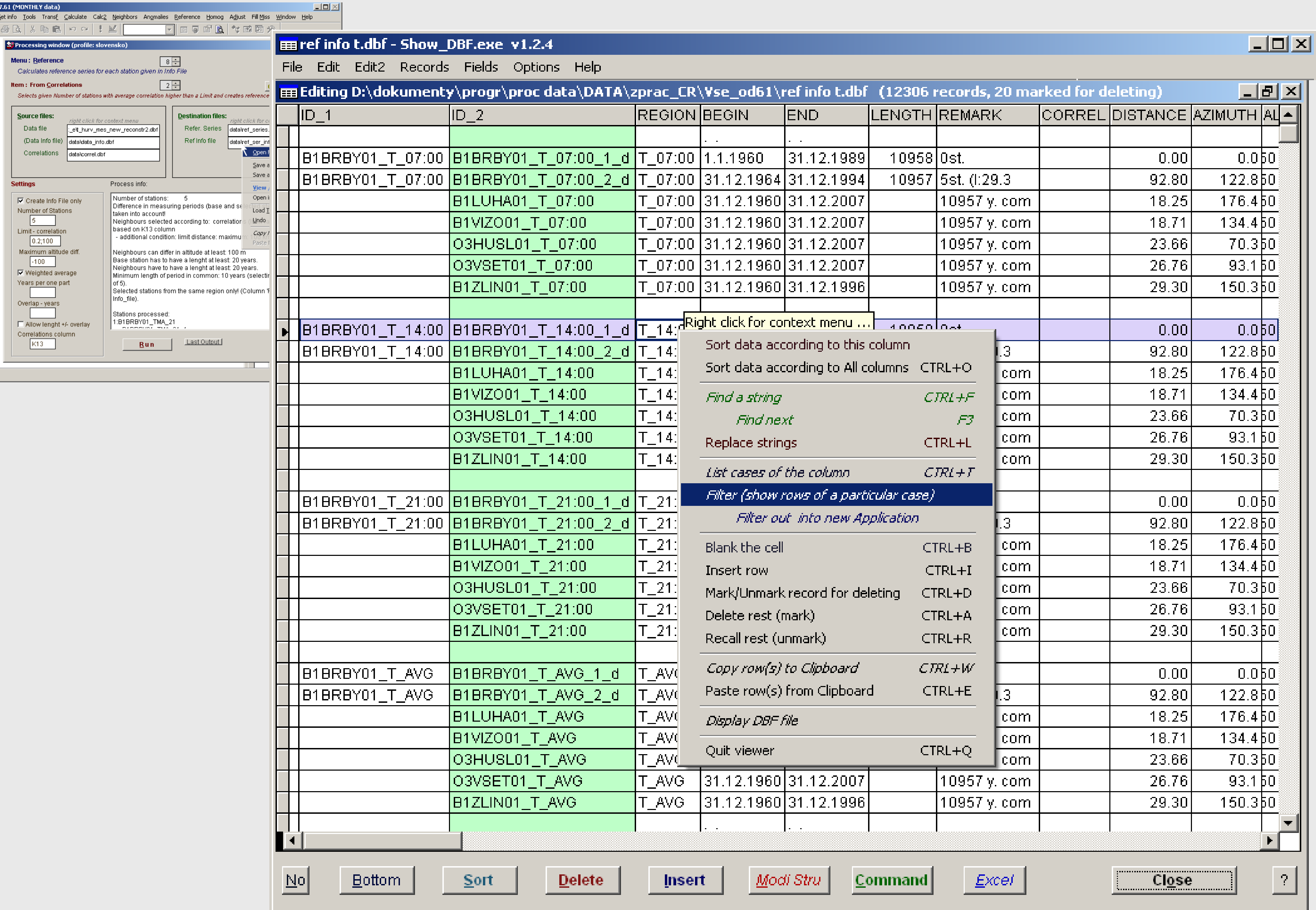
NAME	ID	B	E	L	IDXXX	III	REGION	LATITUDE	LONGITUDE	ALTITUDE	BEGIN	END	LENGTH	MISS_CN
Bystřice pod Hostýnem	B1BYSH01_SCE_07:00				B1BYSH01		SCE	17.67	49.40	315	1.1.1961	31.1.2006	46	
Bystřice pod Hostýnem	B1BYSH01_SNO_07:00				B1BYSH01		SNO	17.67	49.40	315	1.1.1961	31.1.2006	46	
Bystřice pod Hostýnem	B1BYSH01_SRA_07:00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

ProClimDB software



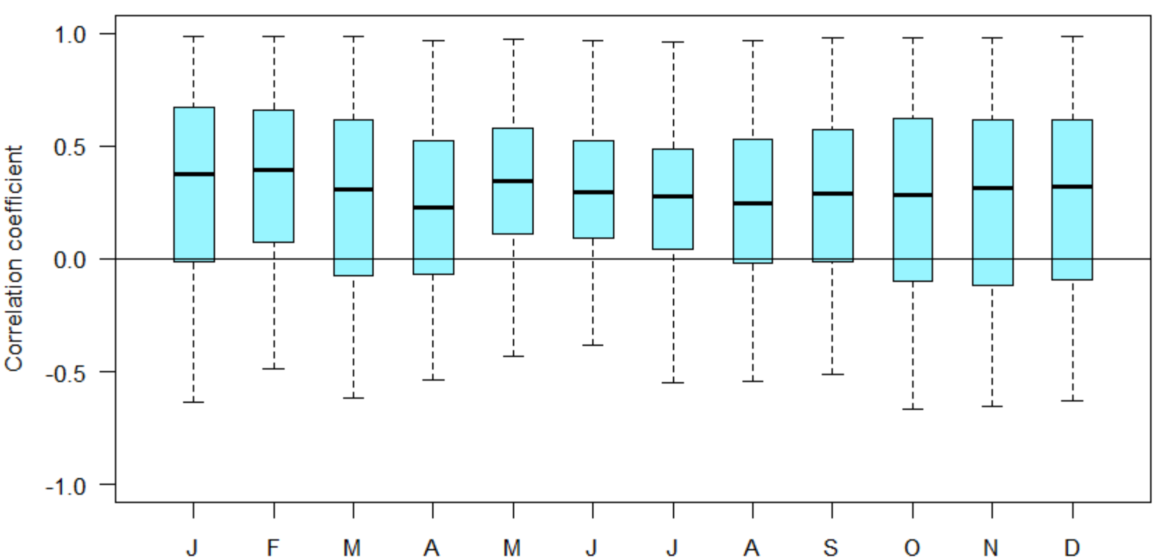
Institute

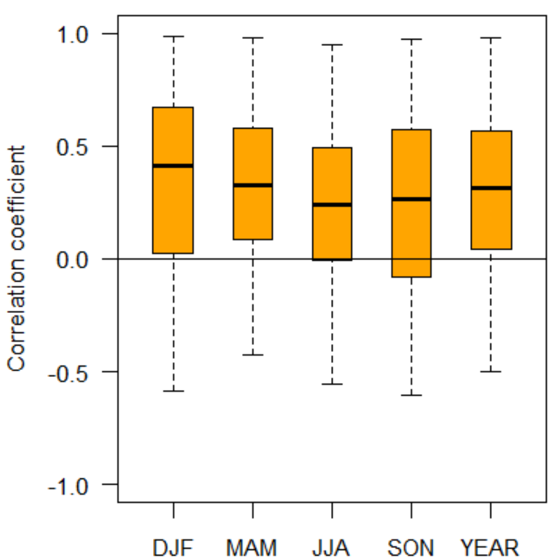
ProClimDB software



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

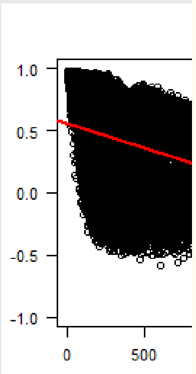


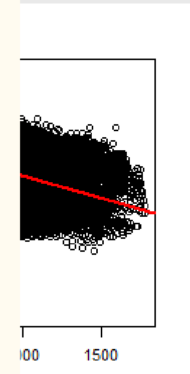


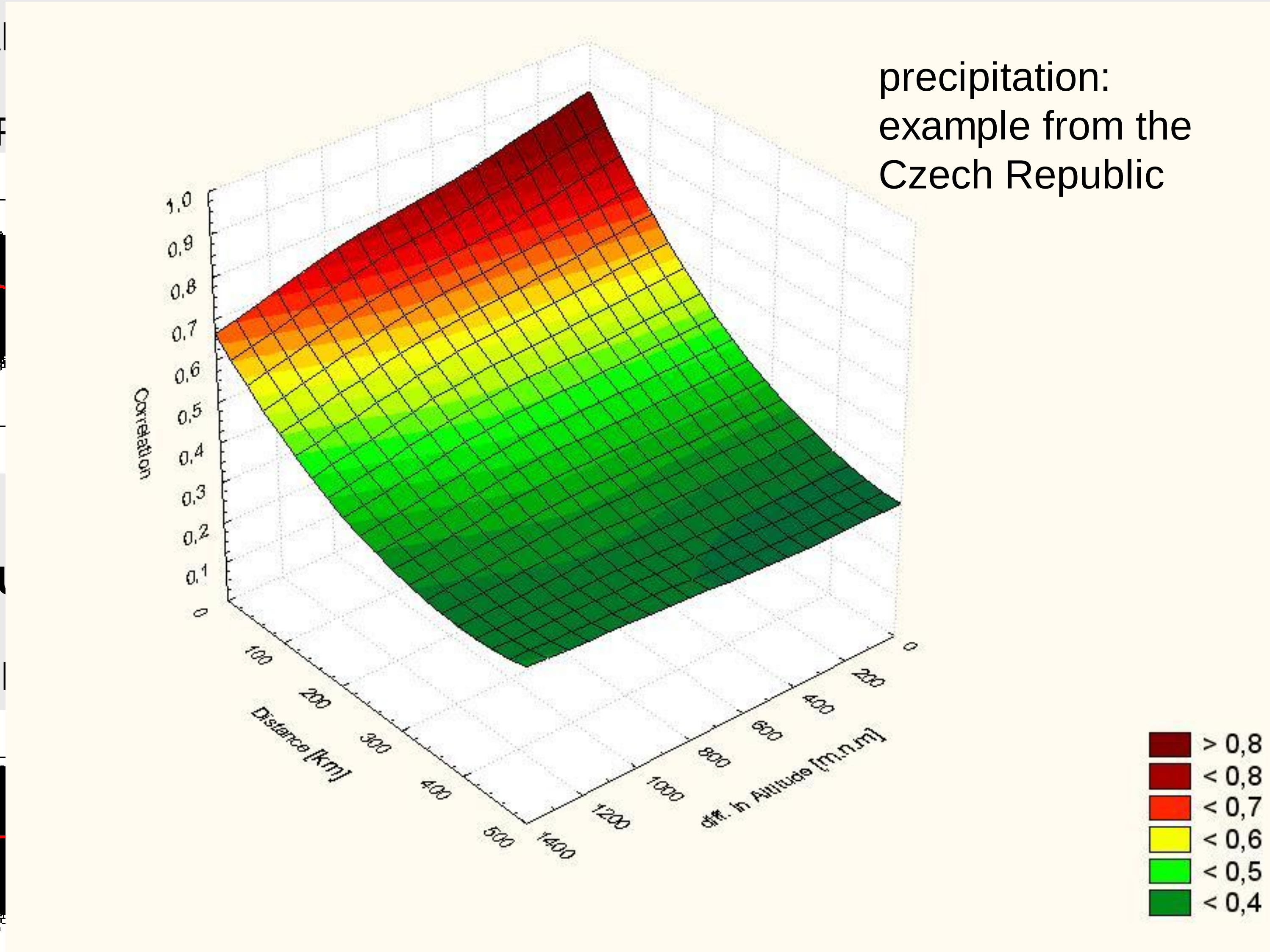
Spatial dependence

DISTA

DJF

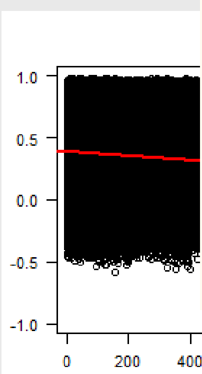


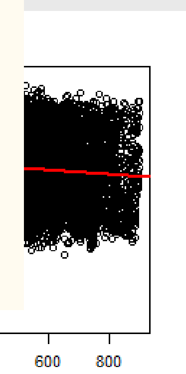




ALTITU

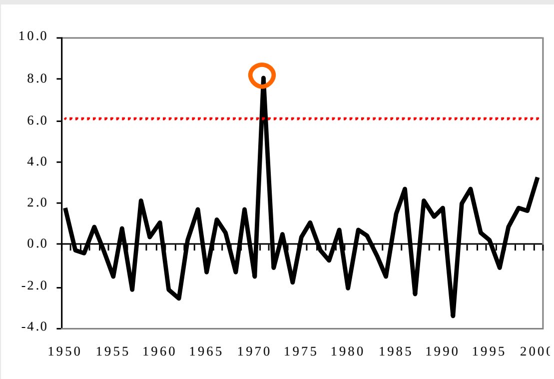
DJ





Data Quality Control

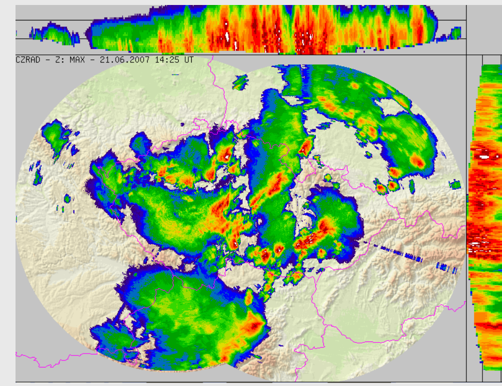
- Own approach, combination of several methods



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
REGIO	ID	YE	MONT	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	
T_03:30	**B2BTUR01_T_03:30**	**2006**	**6**	**25**	**27,30**	**17,28**		**17,30**	**16,10**	**15,50**	**15,80**	**16,10**	**-7**

Data quality control

Year/ Month

Test values

Expect values

Neighbours stations

Distance

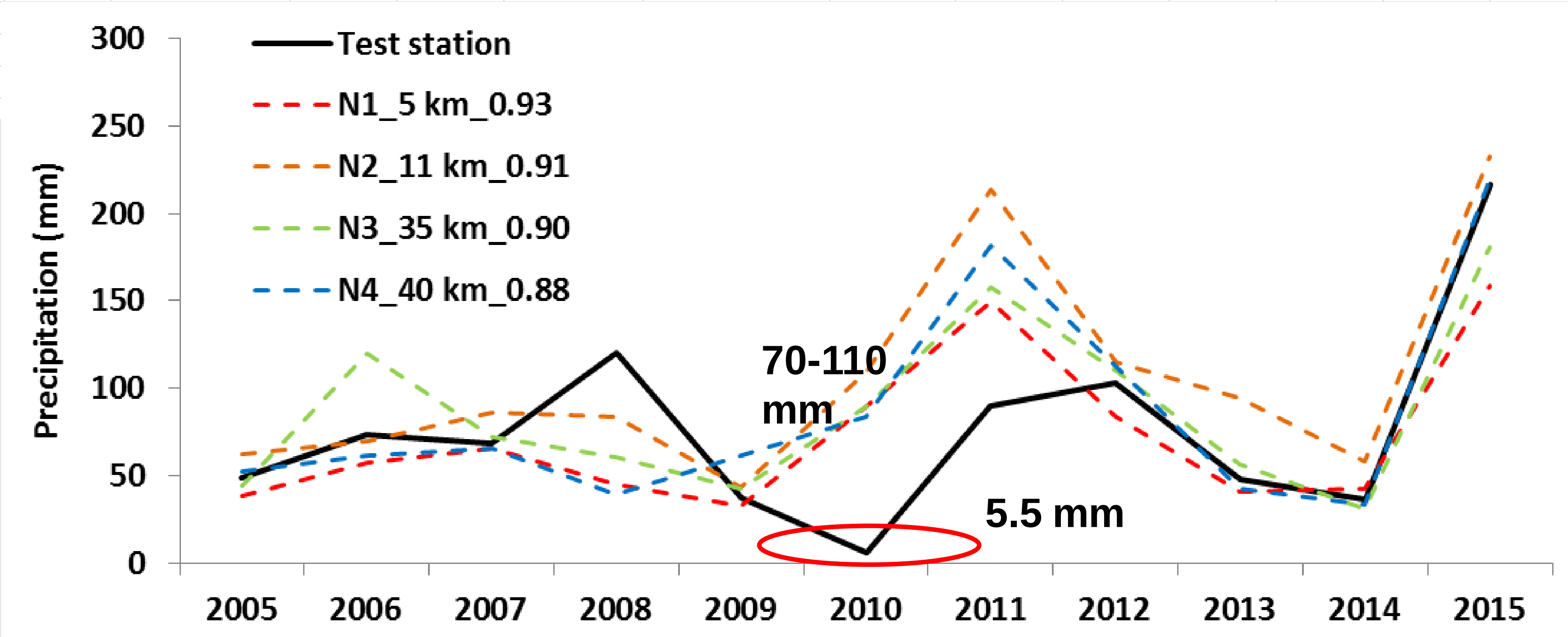
Correlation

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		
	rarrm00004050d					st_5, Co					0.9	
	rarrm00020520d					st_6, Co						0.9
	rarrm00018450d	**2010**	**9**	**5.5**	**85.407**		**89.1**	**109.5**	**89.5**	**83.5**	**71.3**	**79.7**
	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
	rarrm00029600d	**1977**	**6**	**173.8**	**82.517**		**74.2**	**70.0**	**59.6**	**137.5**	**68.3**	**92.1**
	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
	rarrm00039750d	**2011**	**9**	**49.2**	**201.865**		**307.0**	**266.5**	**223.8**	**179.0**	**289.1**	**183.5**

Data quality control

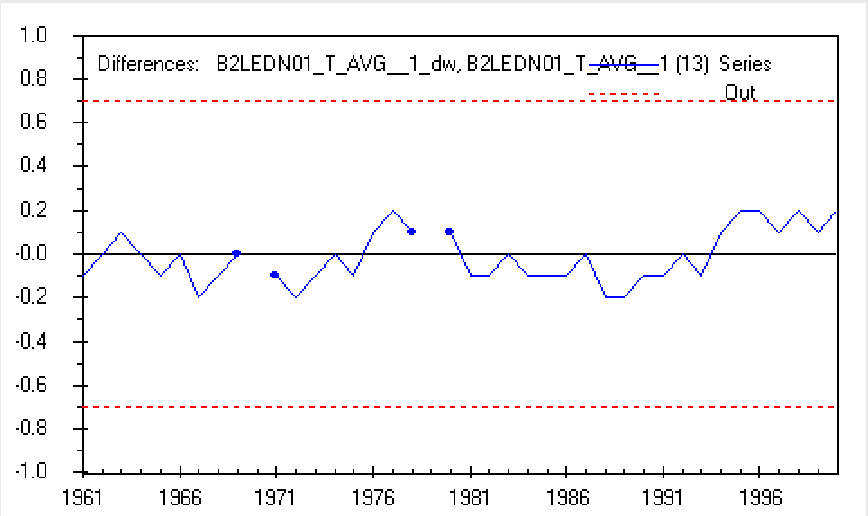
- 22 suspicious values detected in first version
- 19 suspicious values detected in second clean-up version

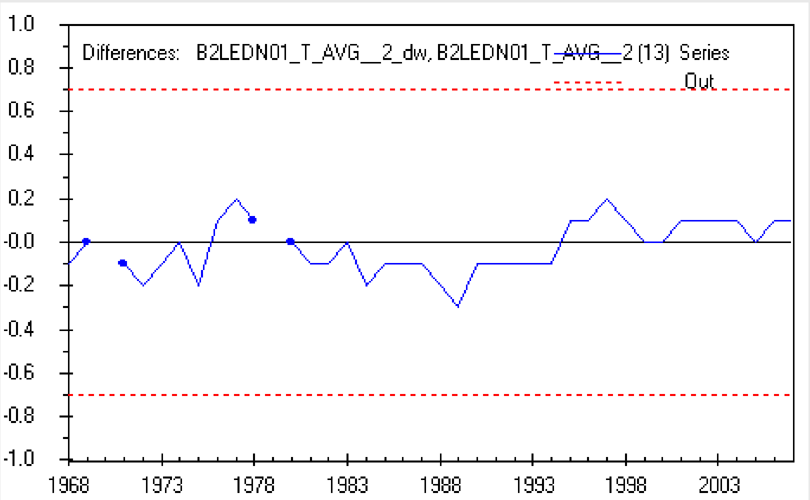
REGION	ID	YEAR	MONTH	ST_BASE	EXPECT_TRF	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5	ST_6
	rarrm00018450d			173.0		Altitude	94.0	360.0	92.0	144.0	164.0	372.0
	rarrm00018700d					st_1, Co	0.9					
	rarrm00018500d					st_2, Co		0.9				
	rarrm00017850d					st_3, Co			0.9			
	rarrm00003780d					st_4, Co				0.9		



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

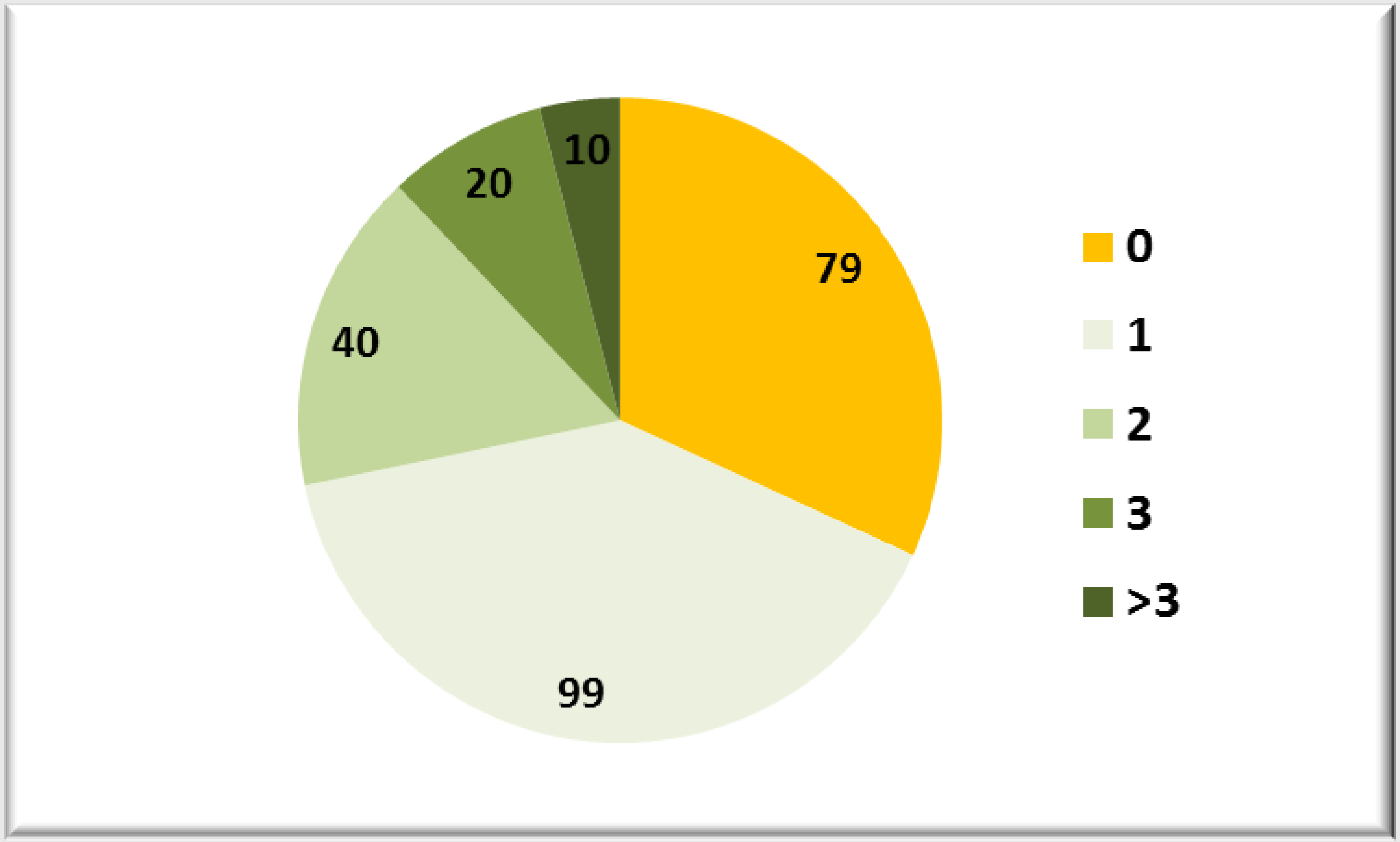
- Ensemble approach – results from many tests, for all month, seasons and annual value

Test	Ref	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Win	Spr	Sum	Aut	Year
A	**avg**	1927	1929	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
A			1930															
A	**corr**	1927	1927	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
A				1939		1938	1939	1940	1922						1937	1937		1935
A	**dist**	1927	1928	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
A			1930								1940							1918
B	**avg**	1927	1928	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
B									1922									
B	**corr**	1927	1927	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
B				1936		1938	1939	1944	1922					1935	1937	1937		1935
B									1937									
B	**dist**	1927	1928	1927	1927	1927	1928	1927	1926	1926	1926	1926	1926	1927	1927	1927	1926	1927
B		1930									1940			1931			1913	1918
V	**corr**													1927			1926	
V															1937	1922		1935
V																1937		
V	**dist**													1927	1927	1927		
V																		1918

Homogenization - results

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

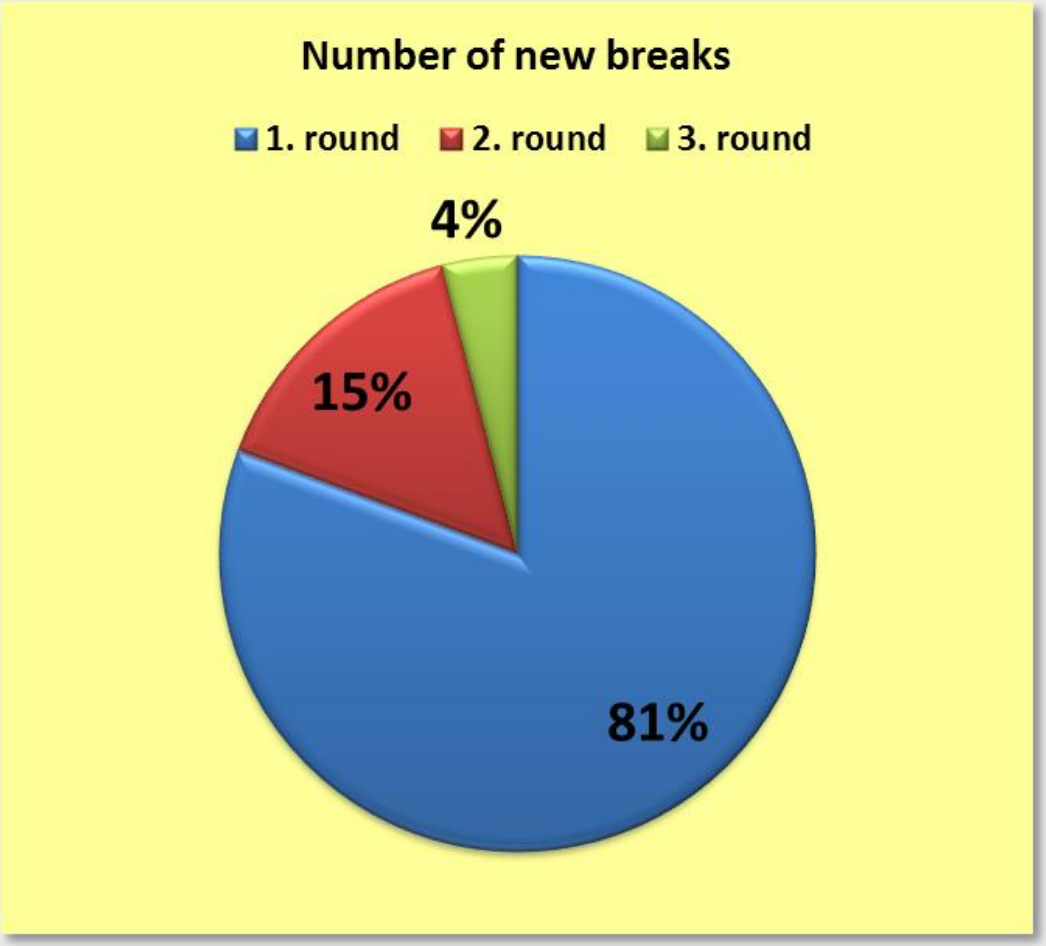
Number of stations with given number of breaks



Homogenization - results

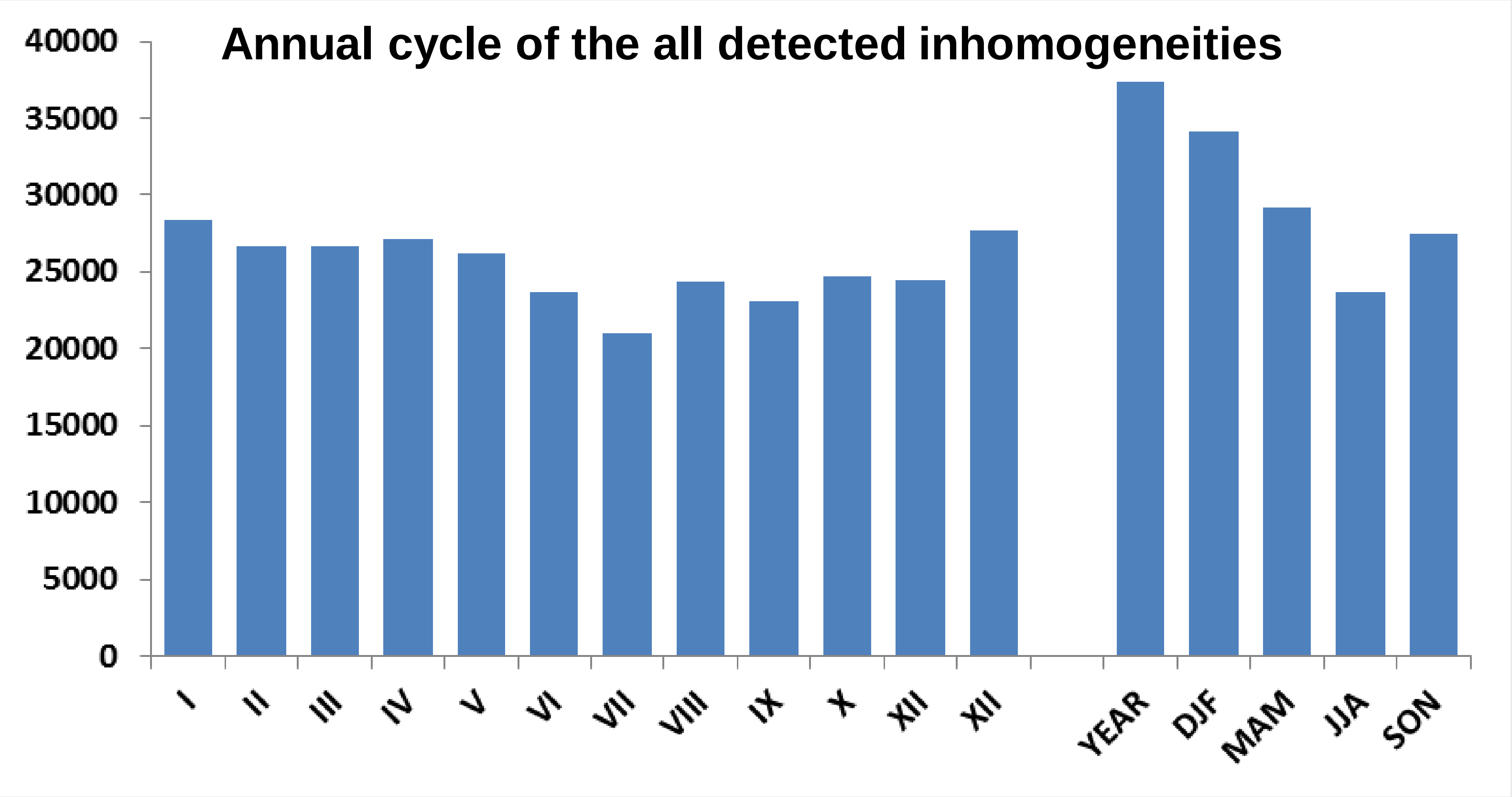
- 1. version: 245 stations and 307 detected breaks
- 2. version (cleanup): 248 stations and 222 detected breaks

Number of breaks found per iteration



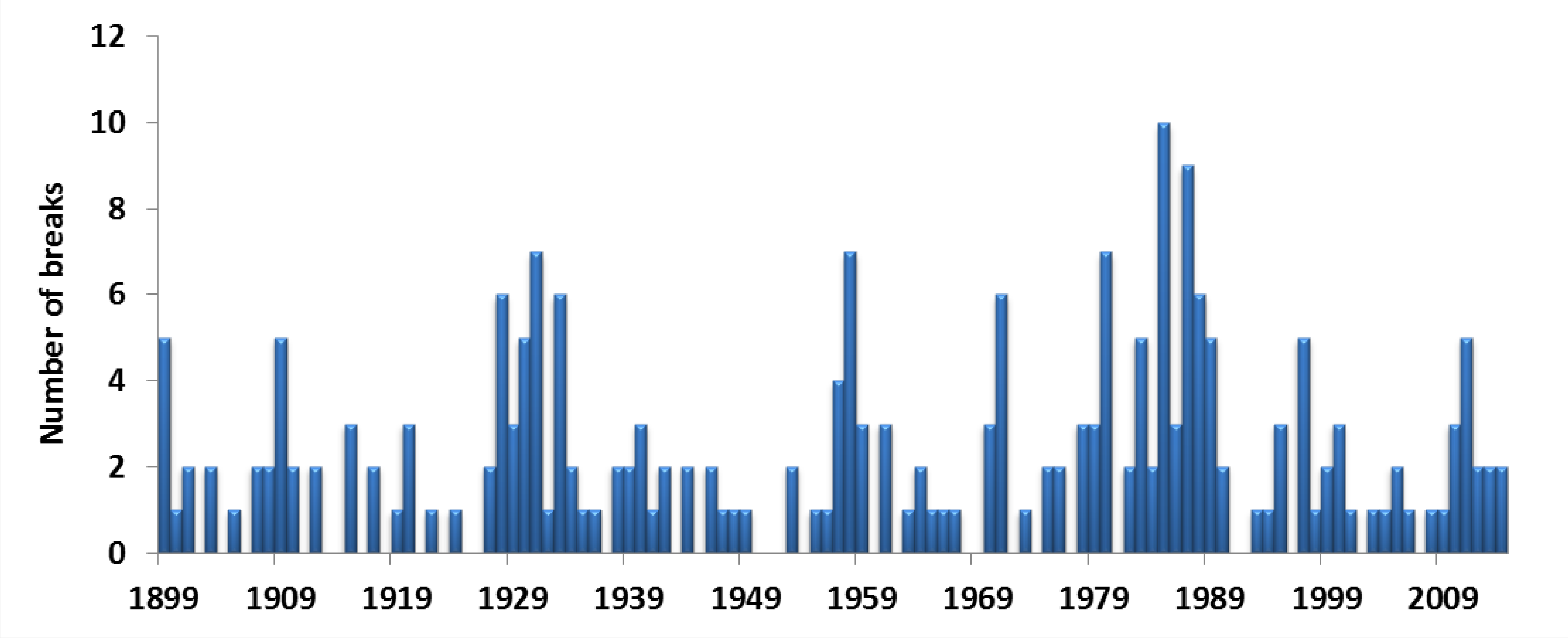
Inhomogeneities detection

- 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



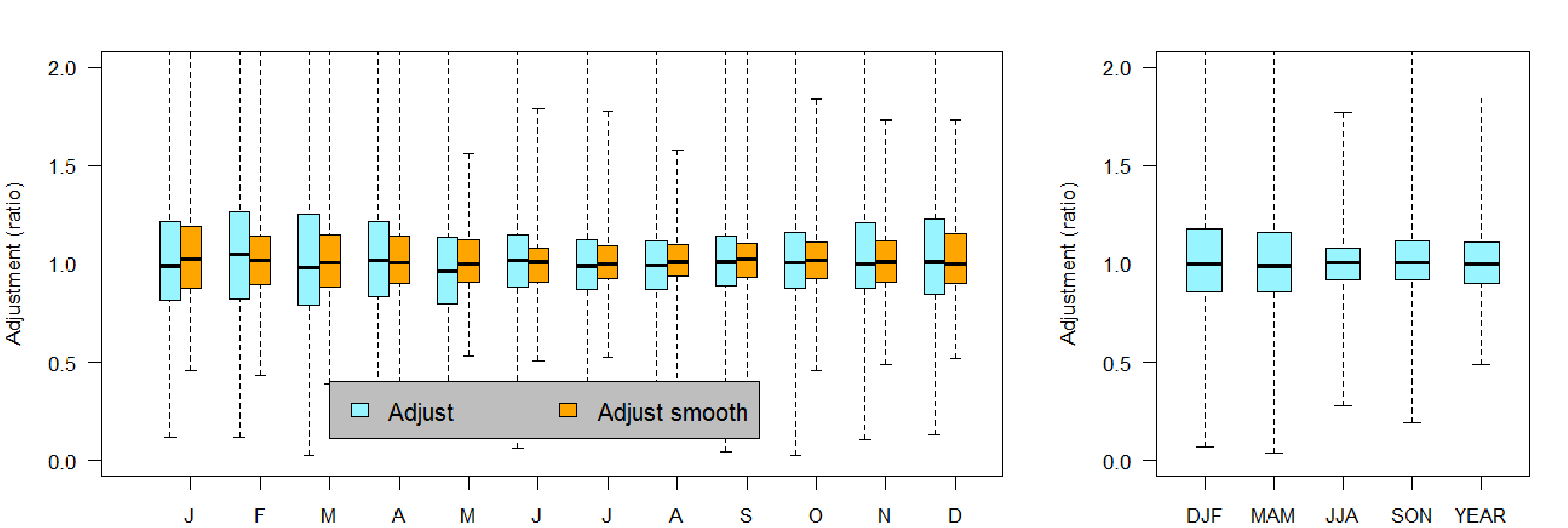
Inhomogeneities detection

Number of breaks in individual years



Inhomogeneities adjustment

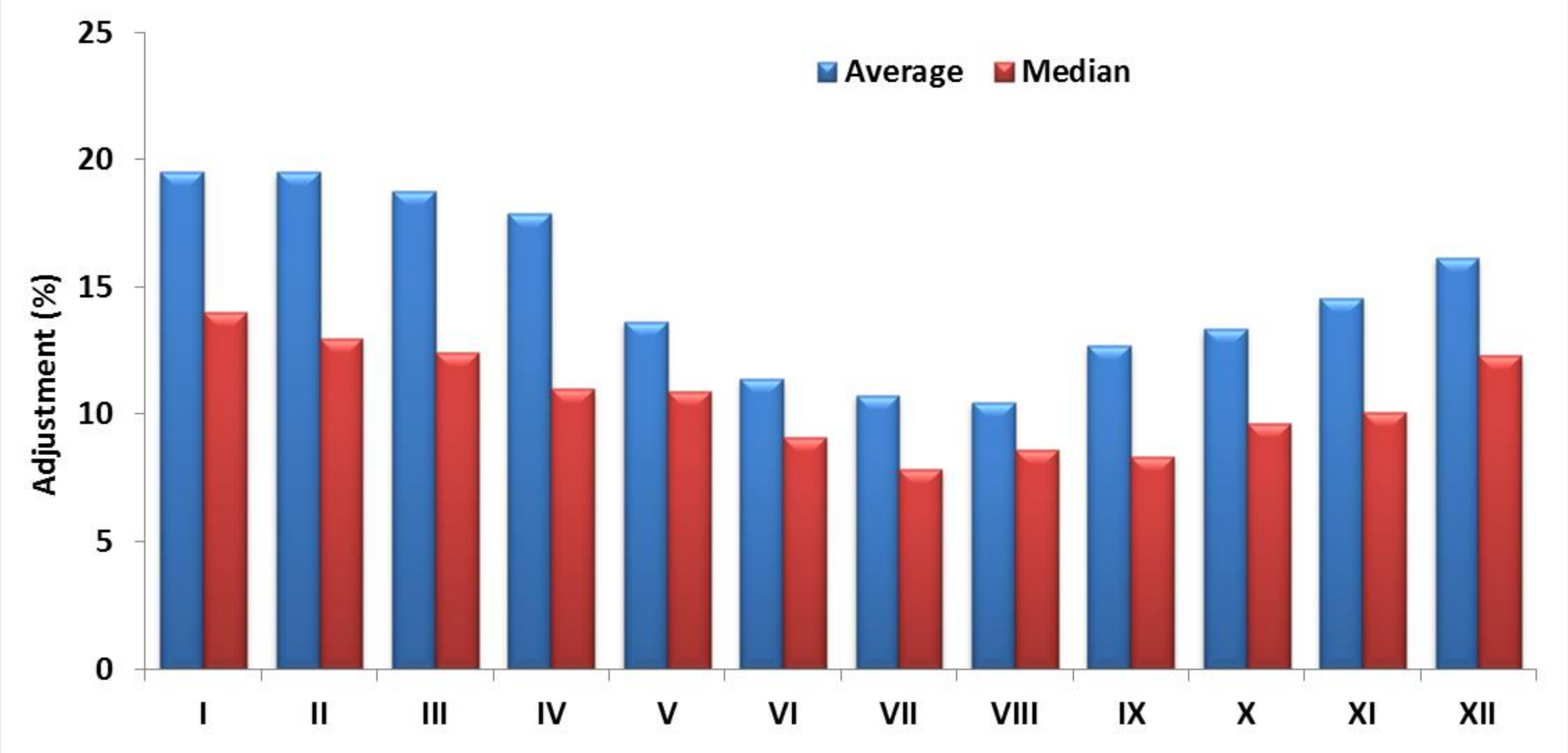
- Reference series calculated from 5 neighbours best correlated or nearest stations
- Correction factor is calculated from differences 20 years before and after breaks
- Final correction factor is smooth by Gauss low pass filter for individual month
- Normally we use own method for correcting daily data,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

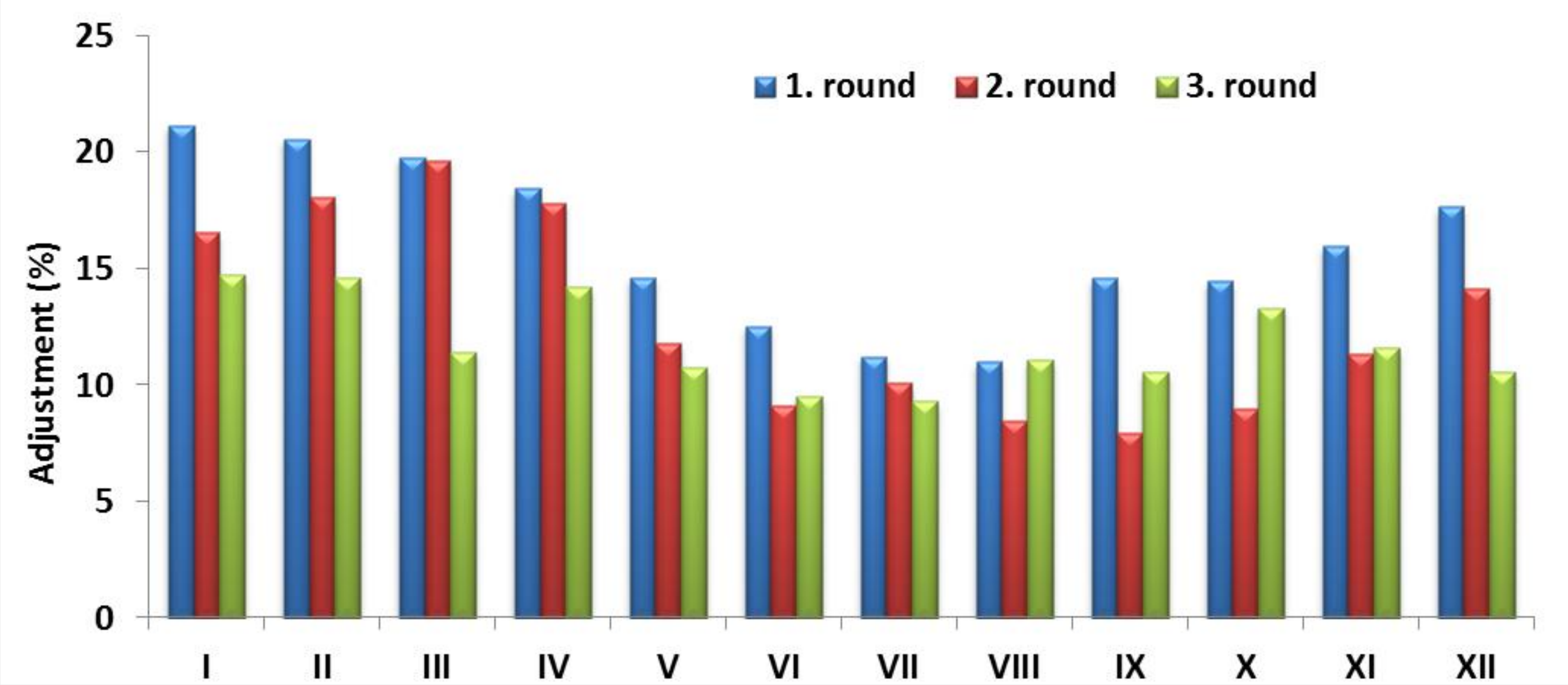
Absolute size of the adjustment (ratio)



Inhomogeneities adjustment

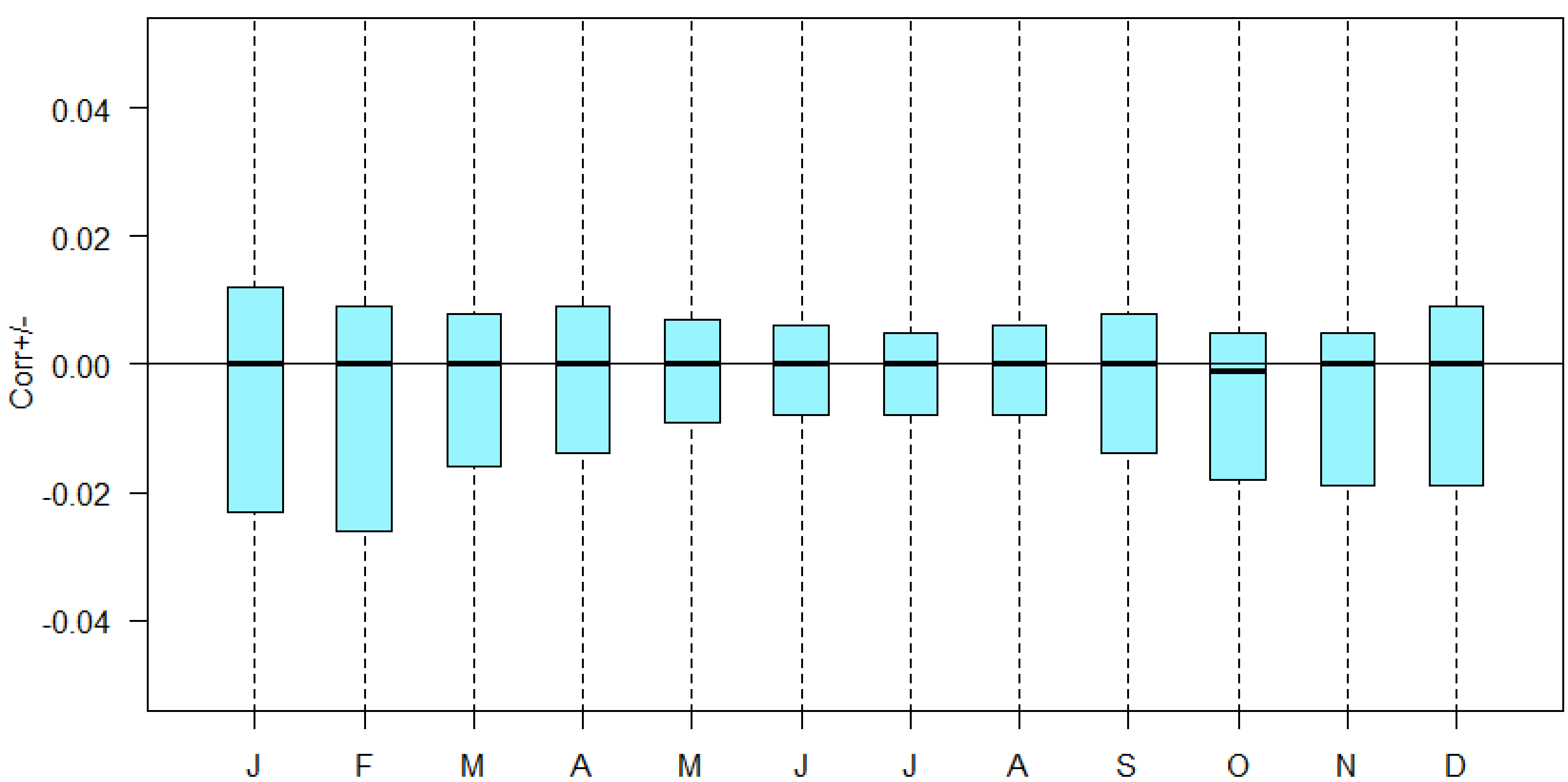
- Correction factor was higher in the 1st iteration of detection/correction (largest breaks were detected in the 1st iteration)
- 1st iteration (16.8 %), 2nd itereation (12.8 %), 3rd iteration (11.8 %)

Absolute size of the adjustment (ratio)

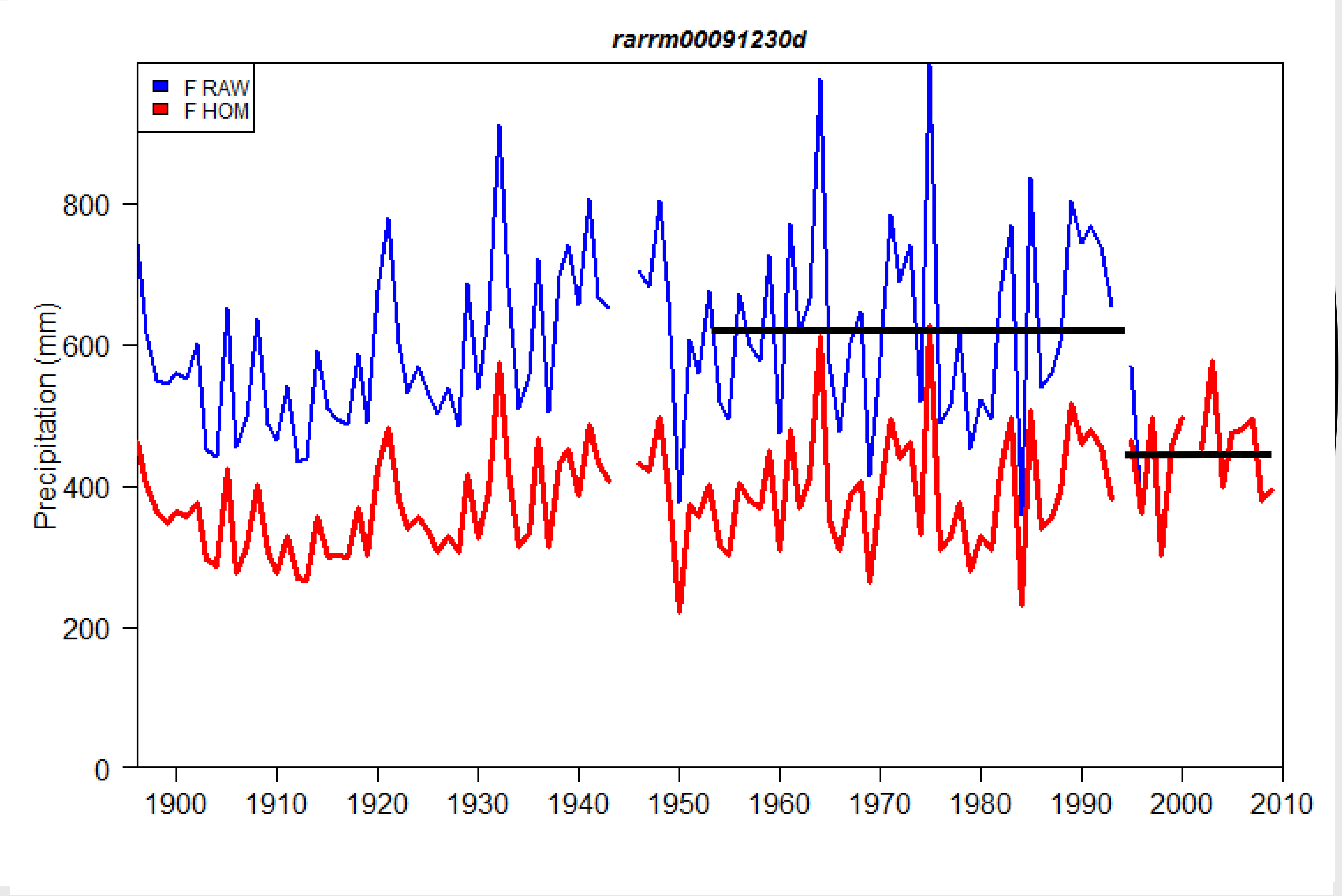


Inhomogeneities adjustment

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



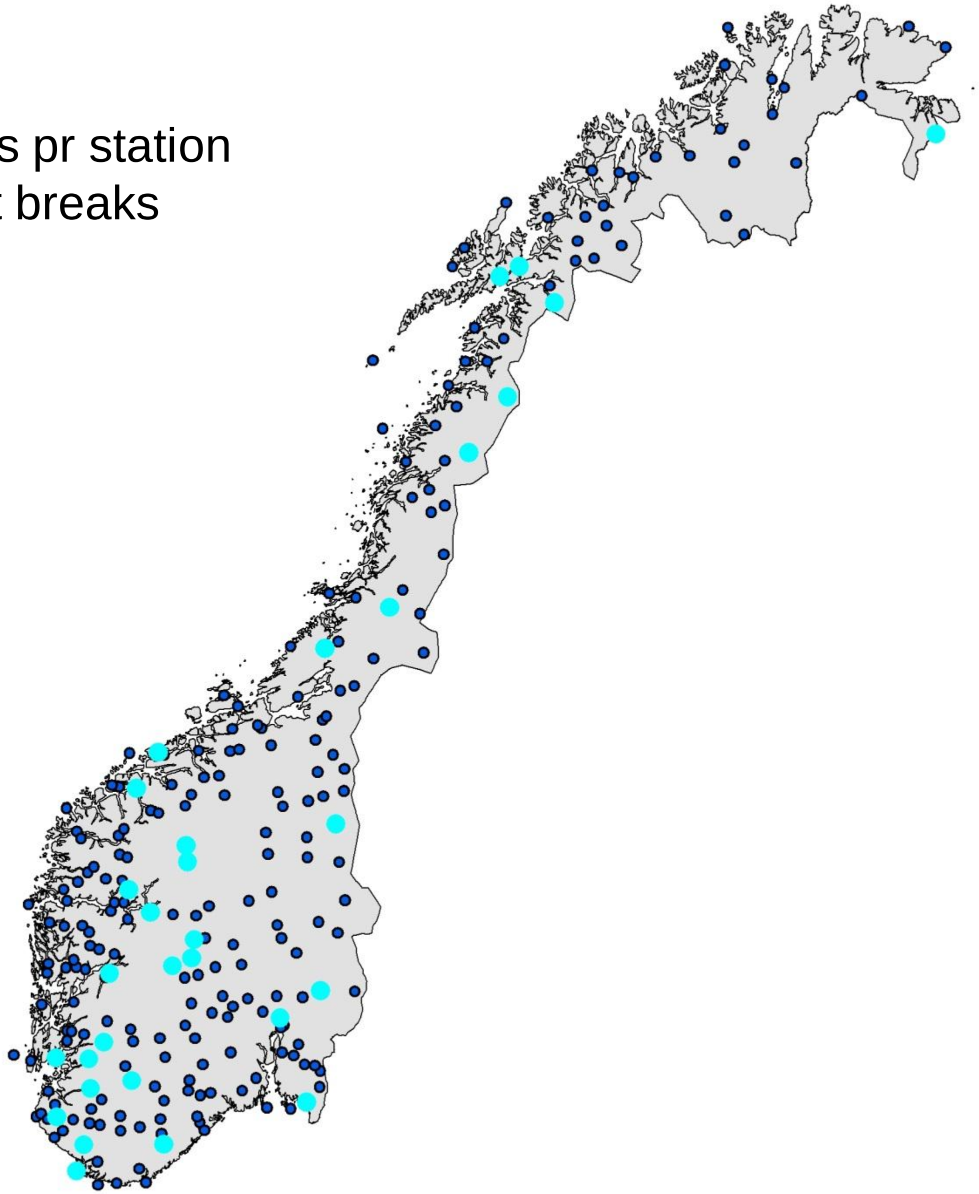
Homogenization – new series



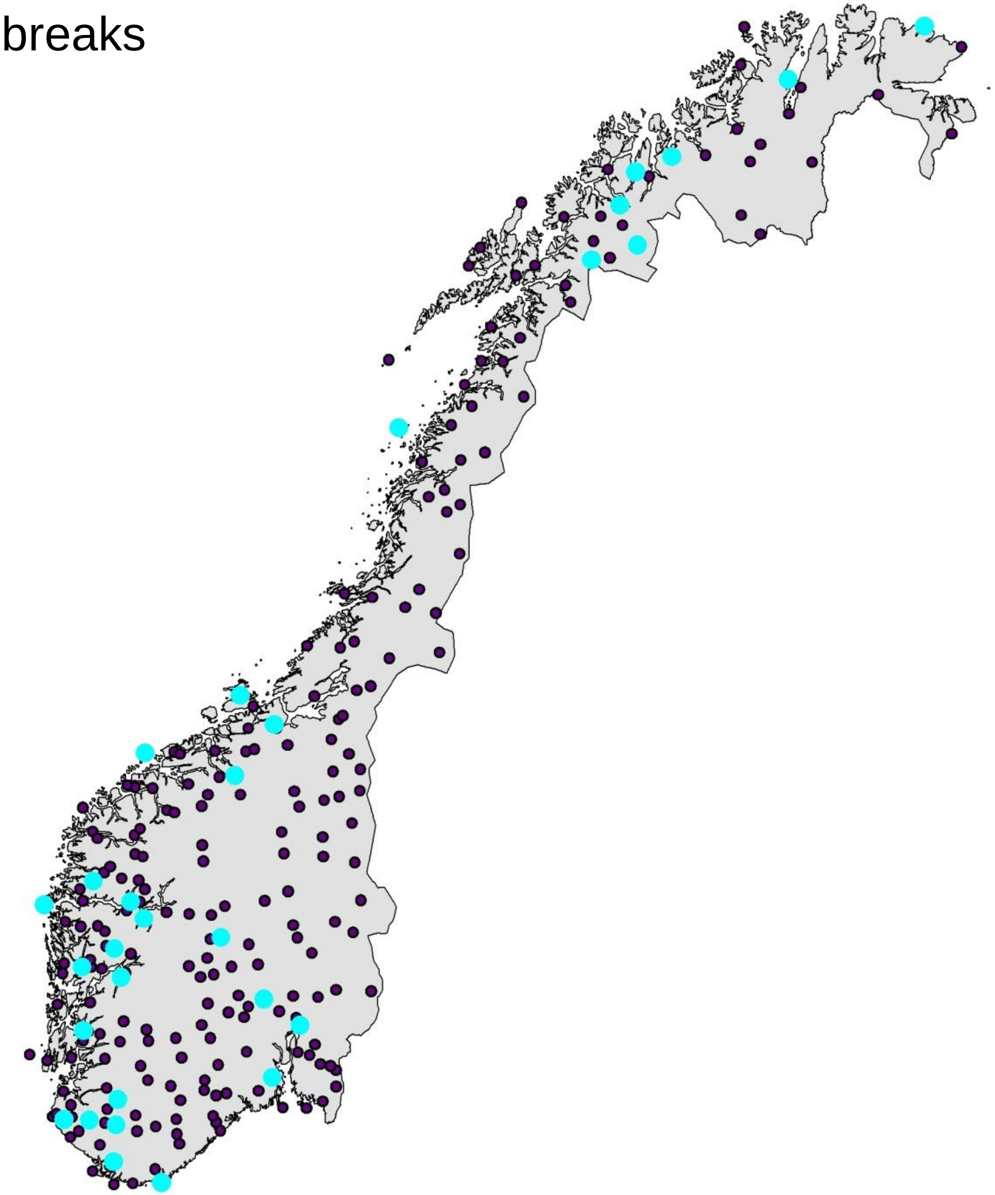
MASH homogenization results

1314 Breaks

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

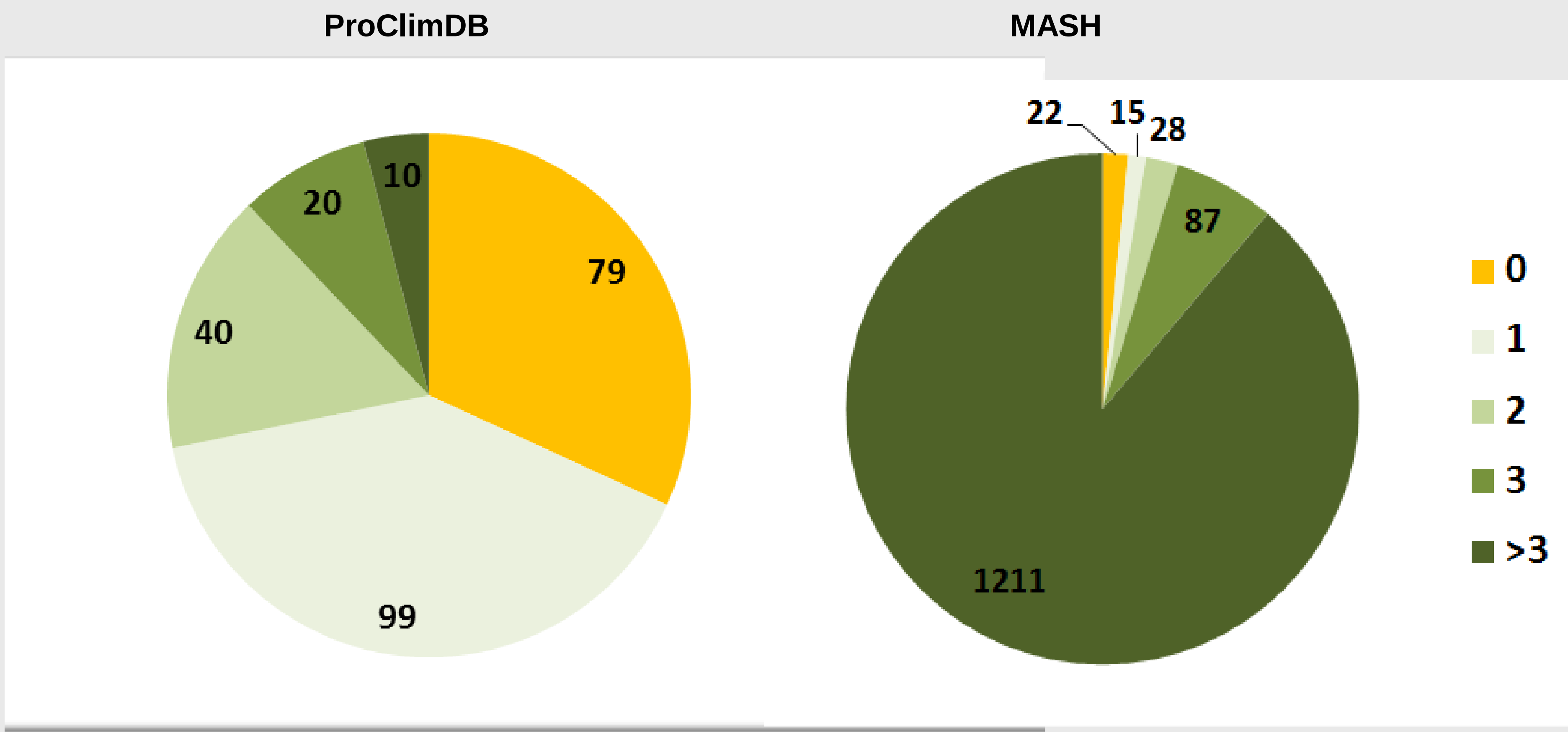


- 30 stations: > 11 breaks



Comparison with MASH results

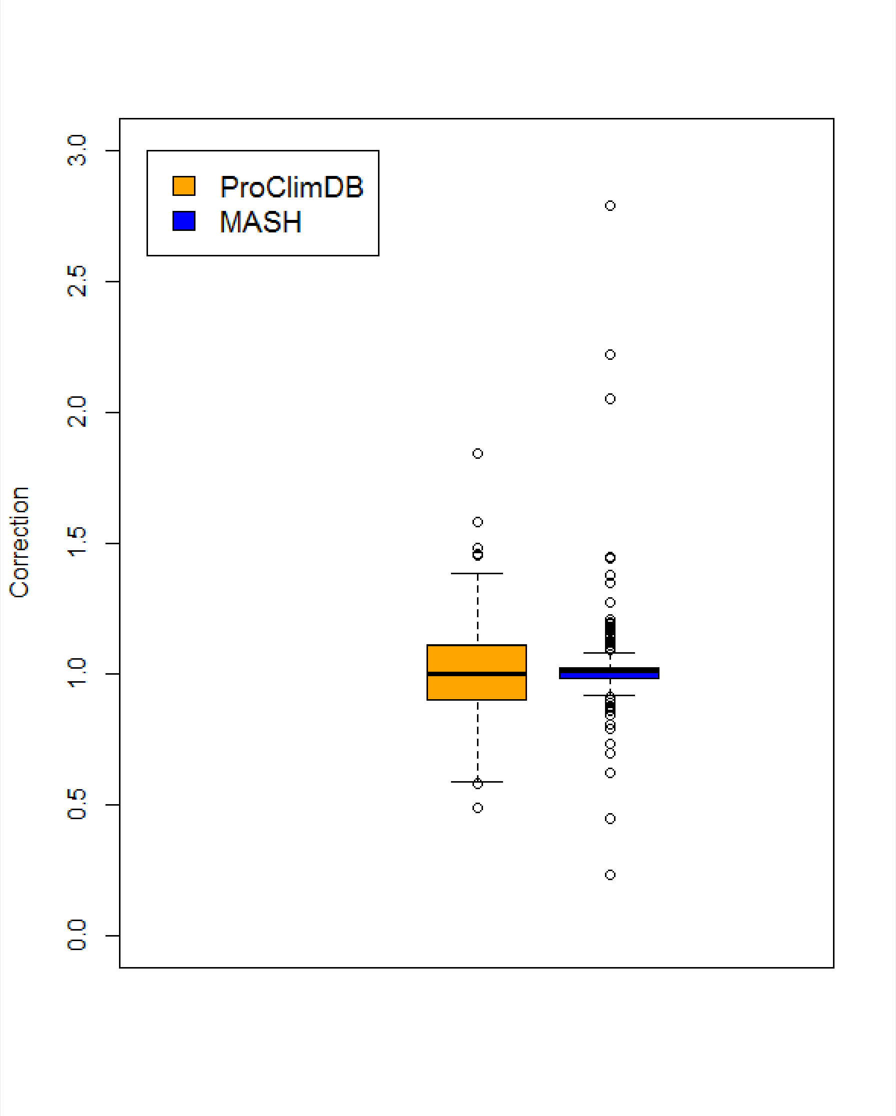
Number of stations with given number of breaks



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

Thanks for the attention!

Telephone: +47 22 96 31 69
E-mail: elinl@met.no
Twitter: @nile_2011