Homogenization of Precipitation Series of the Czech Republic

Petr Štěpánek

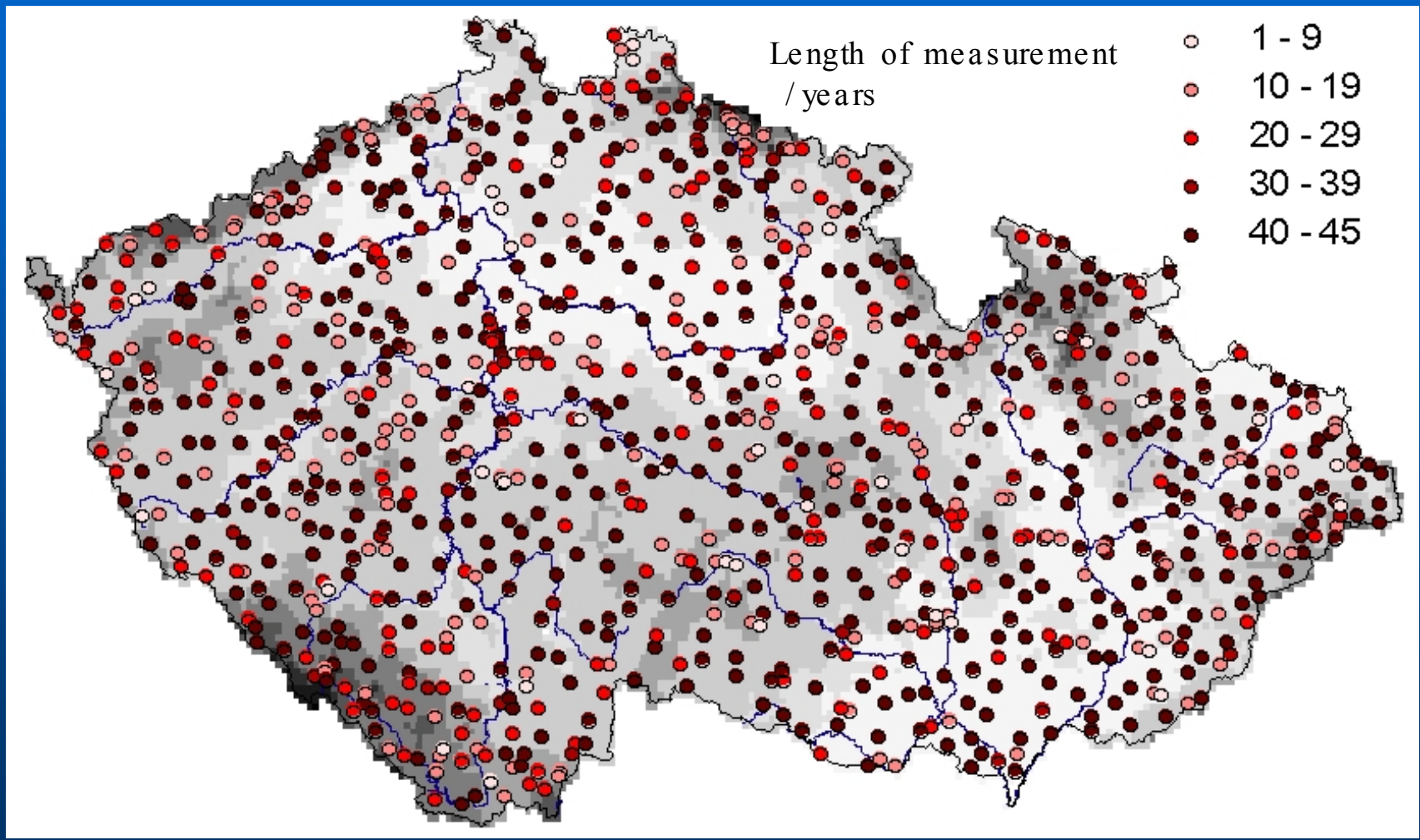
Czech Hydrometeorological Institute,
regional office Brno

e-mail: petr.stepanek@chmi.cz

Data processed

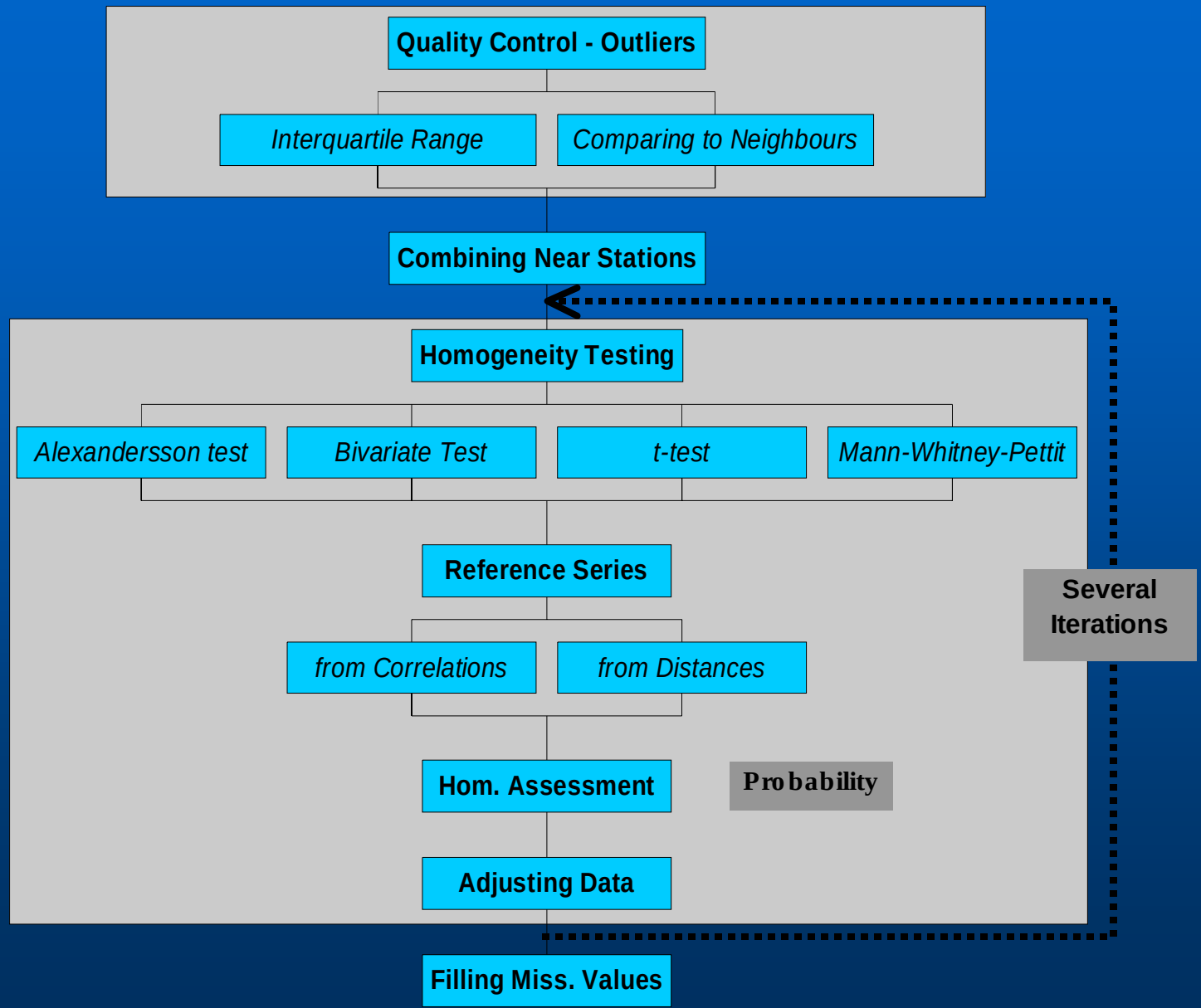
- period 1961-2003
- monthly series (+seasonal and annual sums)
- 589 stations
  - 566 series with measurements longer than 30 years
  - 23 series created by combining neighbouring stations
  - 457 stations has 40 years of measurement
- mean minimum distance: 7.5 km

Spatial distribution of precipitation stations

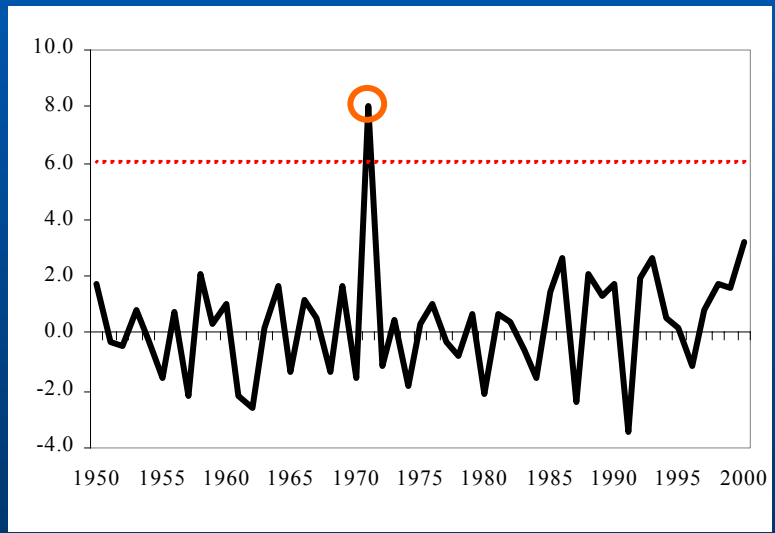


Data Processing

Monthly, Seasonal and Annual Averages

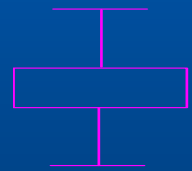


Data Quality Control
Finding Outliers



Data Quality Control
Finding Outliers

- **1. Limits derived from interquartile range**
  - **series of ratios (logarithms) of tested and reference series**
  - **reference series created as an average of 5 mostly correlated stations, max. distance 35 km**
  - **limits: coefficient (multiple) = 3.0**



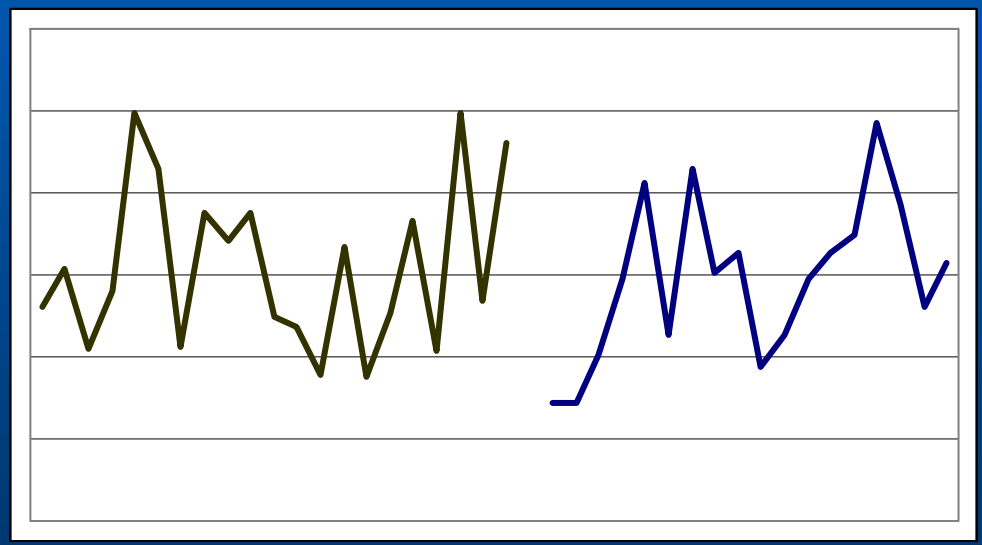
- **2. comparing values to values of neighbouring stations**
  - **comparing to 5 nearest stations**
  - **series of standardized ratios (logarithms of ratios)**
  - **number of cases exceeding 95% confidence limits**

Example:

Comparing tested series to neighbouring series

ID	YEAR	MON	ST_BASE	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5	Rat1_STND	Rat2_STND	Rat3_STND	Rat4_STND	Rat5_STND	CDF_MAX	No_sign.
B1BLAT01			**211.0**	Altitudes,li	*225.0*	*280.0*	*176.0*	*190.0*	*240.0*							
B1HLUK01				st_1, distar	*6.8*											
B1VELV01				st_2, distar		*8.9*										
B1STRZ01				st_3, distar			*10.4*									
B1BZEN01				st_4, distar				*12.2*								
B1RADE01				st_5, distar					*13.3*							
B1BLAT01	1961	1	**14.5**		21.7	16.9	15.5	23.7	19.6	1.140	-0.365	0.769	1.817	0.911	0.965	
B1BLAT01	1961	2	**39.2**		33.7	63.1	40.9	39.5	49.0	-0.646	0.467	0.233	-0.088	0.312	0.950	
B1BLAT01	1961	3	**15.1**		20.4	21.0	14.9	21.2	22.2	0.560	0.389	0.516	1.344	1.180	0.911	
B1BLAT01	1961	4	**57.7**		56.1	34.5	34.7	105.3	44.6	-0.042	-2.589	-1.295	2.145	-1.126	1.000	**2**
B1BLAT01	1961	5	**73.5**		62.6	95.9	96.3	71.1	114.6	-0.601	0.891	1.322	0.239	1.718	0.957	
B1BLAT01	1961	6	**148.3**		208.3	158.3	79.4	101.2	76.2	1.305	-0.135	-1.805	-0.915	-2.374	1.000	**1**
B1BLAT01	1961	7	**77.5**		89.2	106.9	102.3	86.0	123.2	0.475	0.988	1.549	0.604	1.658	0.951	
B1BLAT01	1961	8	**29.3**		23.4	42.8	34.2	30.9	35.6	-0.654	0.829	0.567	0.212	0.372	0.951	
B1BLAT01	1961	9	**12.4**		12.2	16.3	10.3	13.3	12.2	0.125	0.769	-0.202	0.862	0.148	0.885	
B1BLAT01	1961	10	**56.0**		51.7	77.6	74.1	81.4	82.7	-0.406	0.651	1.419	1.770	1.182	0.962	
B1BLAT01	1961	11	**60.8**		54.5	99.5	65.0	55.8	79.6	-0.643	1.751	0.775	-0.505	1.479	0.960	
B1BLAT01	1961	12	**45.5**		32.5	48.4	35.3	33.6	45.1	-1.565	-1.319	-1.066	-1.436	-0.641	0.995	
B1BLAT01	1962	1	**12.5**		26.3	8.7	12.5	11.3	13.0	2.264	-2.377	0.492	-0.493	-0.106	1.000	**2**
B1BLAT01	1962	2	**28.9**		27.3	55.4	37.1	26.6	46.7	-0.178	1.064	0.977	-0.371	1.217	0.915	
B1BLAT01	1962	3	**49.5**		47.0	55.9	43.7	44.4	49.4	-0.540	-0.427	-0.293	-0.369	-0.394	0.938	
B1BLAT01	1962	4	**44.1**		51.3	70.8	49.6	43.2	54.5	0.575	0.666	0.555	0.282	0.247	0.774	
B1BLAT01	1962	5	**113.2**		111.6	129.3	115.5	137.7	110.7	0.000	0.294	0.495	0.918	0.038	0.841	
B1BLAT01	1962	6	**29.2**		24.1	23.9	39.5	18.6	29.6	-0.504	-1.225	1.036	-1.138	0.131	0.987	
B1BLAT01	1962	7	**143.1**		157.1	103.3	84.7	177.8	115.8	0.284	-2.197	-1.579	0.947	-0.881	0.999	**1**
B1BLAT01	1962	8	**51.1**		58.4	13.9	14.1	18.8	14.9	0.614	-3.961	-3.217	-2.477	-3.306	1.000	**4**
B1BLAT01	1962	9	**39.6**		39.9	36.0	35.8	36.8	33.3	0.191	-0.815	0.145	0.061	-0.329	0.965	
B1BLAT01	1962	10	**44.5**		43.8	55.5	47.7	45.4	50.2	-0.070	0.298	0.674	0.162	0.447	0.858	

Combining measurements of neighbouring stations



Combining measurements of neighbouring stations

- Selecting neighbours till **5 km**
- gap between two series: maximum **4 years**
- resulting series: at least **30 years long**

→ **Further 23 combined series added to 566 original ones**

Example:
Combining series of neighbouring stations

ID	REMARK	BEGIN	END	LENGTH	*REMARK*	DISTANCE	ALT_1	ALT_2	ALL_MONTHS
B1POZL01		1961	1993	33					
	List of neighbours>								
	1.>B1HLHO01	1961	2000	40	*49.16;17.80*	**4.68**		340	
	2.>B1LUHA01	1961	2000	40	*49.11;17.77*	**1.33**		254	
	Correlations of Overlap>								
	1.>B1HLHO01	1961	1993	33					0.944
	2.>B1LUHA01	1961	1993	33					0.968
	Reconstruction>				*1961-2000 (40)*				
	B1LUHA01	1994	2000	7		**1.33**		254	
	B1POZL01	1961	1993	33	*0*		290		
	Better to reconstr.>								
	>B1LUHA01				*40 / 33 years*				
	Exist before&after>								
	>B1HLHO01								
	>B1LUHA01								

Merging series

Suggestions

Creating Reference Series

- **calculated as weighted average**
  - **from the most correlated stations (5), limit 0.7**
  - **from the nearest stations (5) , limit 30 km**
- **ratios tested / reference series**
  - **testing ratios**
  - **testing logarithms of ratios**

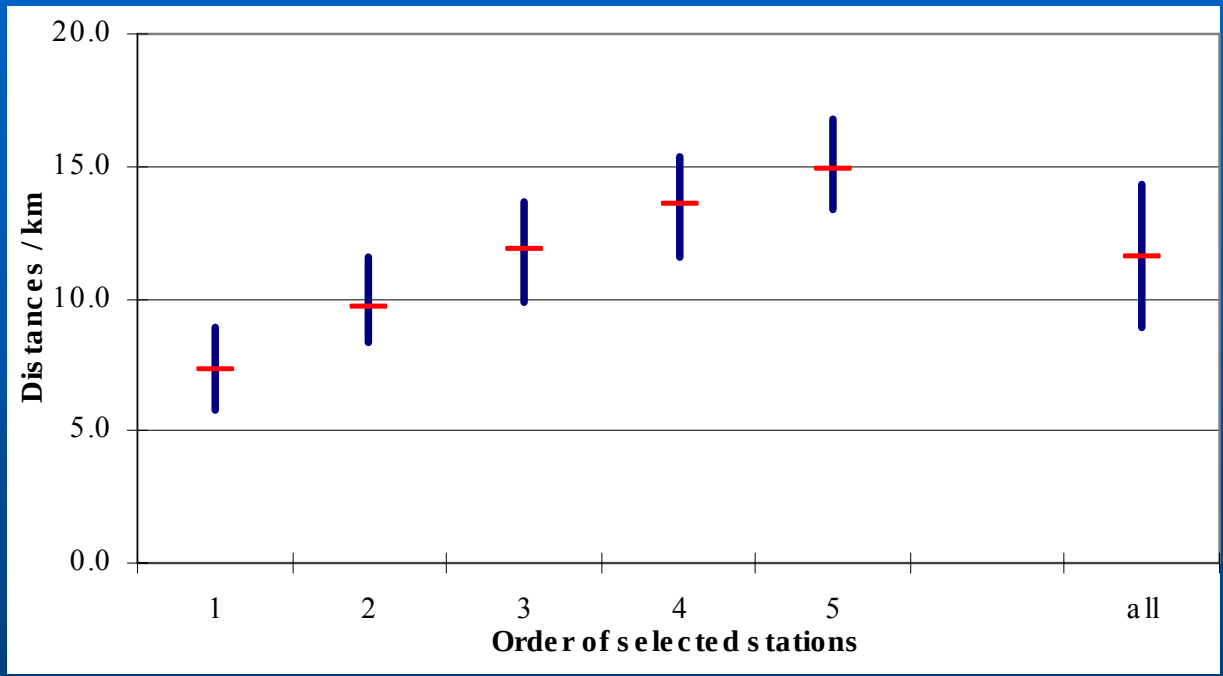
Example:

Proposed list of stations used for creating reference series

ID_1	ID_2	BEGIN	END	LEN(	REMARK	CORREL	DISTANCE	ALT_1	ALT_2
B1BLAT01		1961	2000	40	5st. (l:0.88			211	
	B1HLUK01	1961	2000		40 y. comm.p	0.931	6.78	211	225
	B1VELV01	1961	2000		40 y. comm.p	0.921	8.94	211	280
	B1STRZ01	1961	2000		40 y. comm.p	0.910	10.39	211	176
	B1UHBR01	1961	2000		40 y. comm.p	0.901	17.11	211	222
	B1RADE01	1961	2000		40 y. comm.p	0.884	13.32	211	240
B1BOJK01		1961	2000	40	5st. (l:0.89			302	
	B1STRN01	1961	2000		40 y. comm.p	0.920	16.55	302	385
	B1STHR01	1961	2000		40 y. comm.p	0.917	7.29	302	412
	B1LUHA01	1961	2000		40 y. comm.p	0.908	9.62	302	254
	B1VIZO01	1961	2000		40 y. comm.p	0.895	21.20	302	315
	B1UHBR01	1961	2000		40 y. comm.p	0.891	11.68	302	222
B1BRBY01		1961	1994	34	5st. (l:0.87			350	
	B1BOJK01	1961	2000		34 y. comm.p	0.888	16.54	350	302
	O3ZDEC01	1961	2000		34 y. comm.p	0.886	18.34	350	520
	O3HUSL01	1961	2000		34 y. comm.p	0.881	23.66	350	450
	B1HLHO01	1961	2000		34 y. comm.p	0.875	17.36	350	340
	B1STHR01	1961	2000		34 y. comm.p	0.873	18.59	350	412
B1BUCH01		1961	2000	40	5st. (l:0.86			280	
	B1STME01	1961	2000		40 y. comm.p	0.919	7.29	280	235
	B2KYJO01	1961	2000		40 y. comm.p	0.879	16.54	280	195
	B2KORC01	1961	2000		40 y. comm.p	0.873	11.72	280	305
	B1BZEN01	1961	2000		40 y. comm.p	0.869	12.44	280	190
	B1NAPA01	1961	2000		40 y. comm.p	0.869	17.08	280	205

Selection according to correlations

Creating reference series:
Statistics for selection of the nearest stations

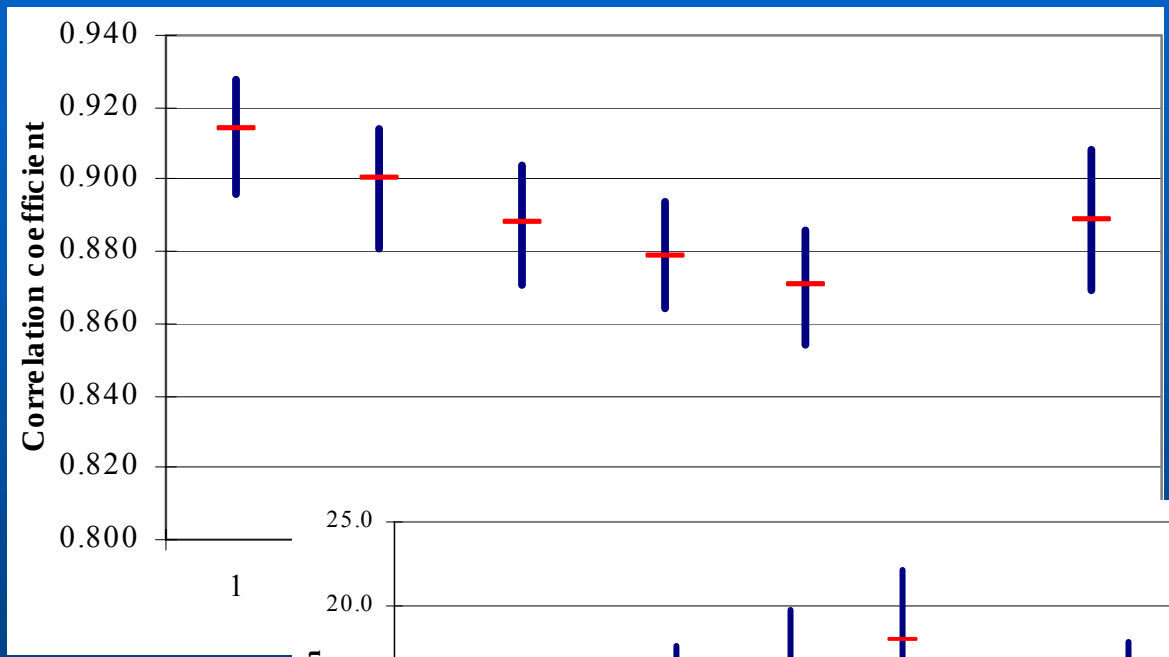


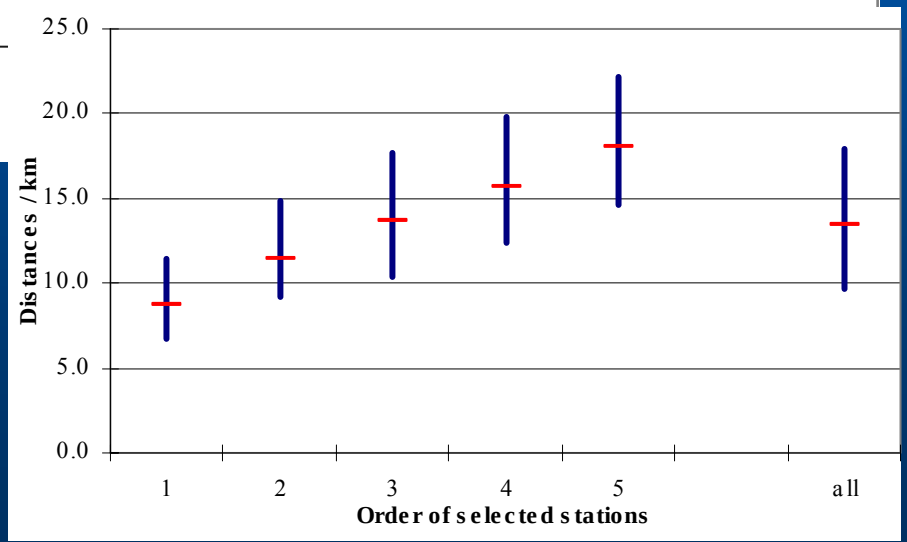
Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)

Creating reference series:
Statistics for selection of the mostly correlated stations





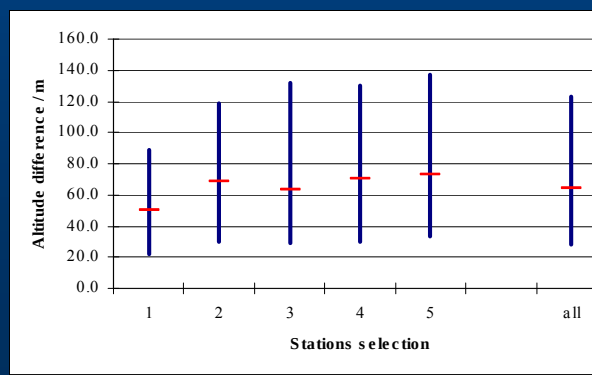
Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)

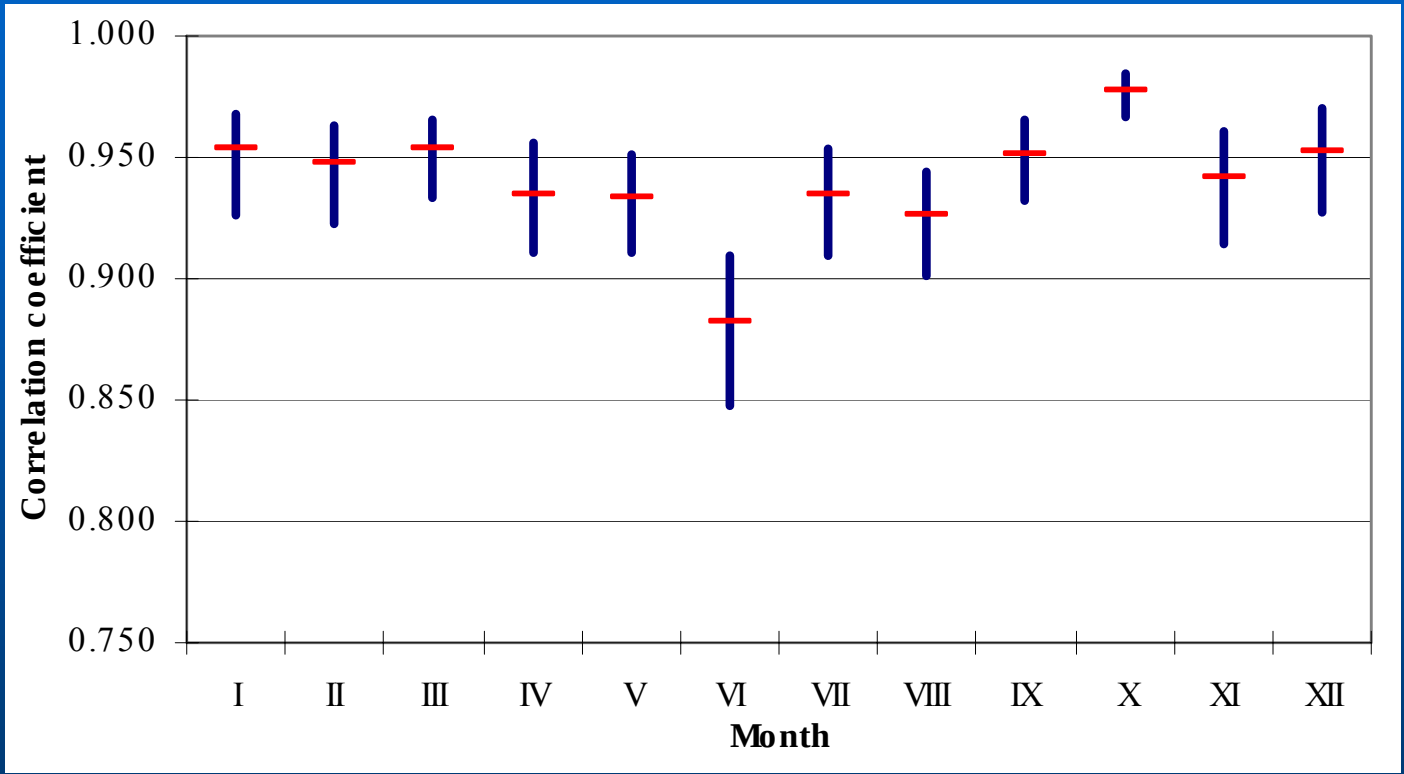
Distances:

- coincidence with selection by means of distances
- but still a little bit different



Correlations between tested and reference series

(reference series calculation based on correlations)



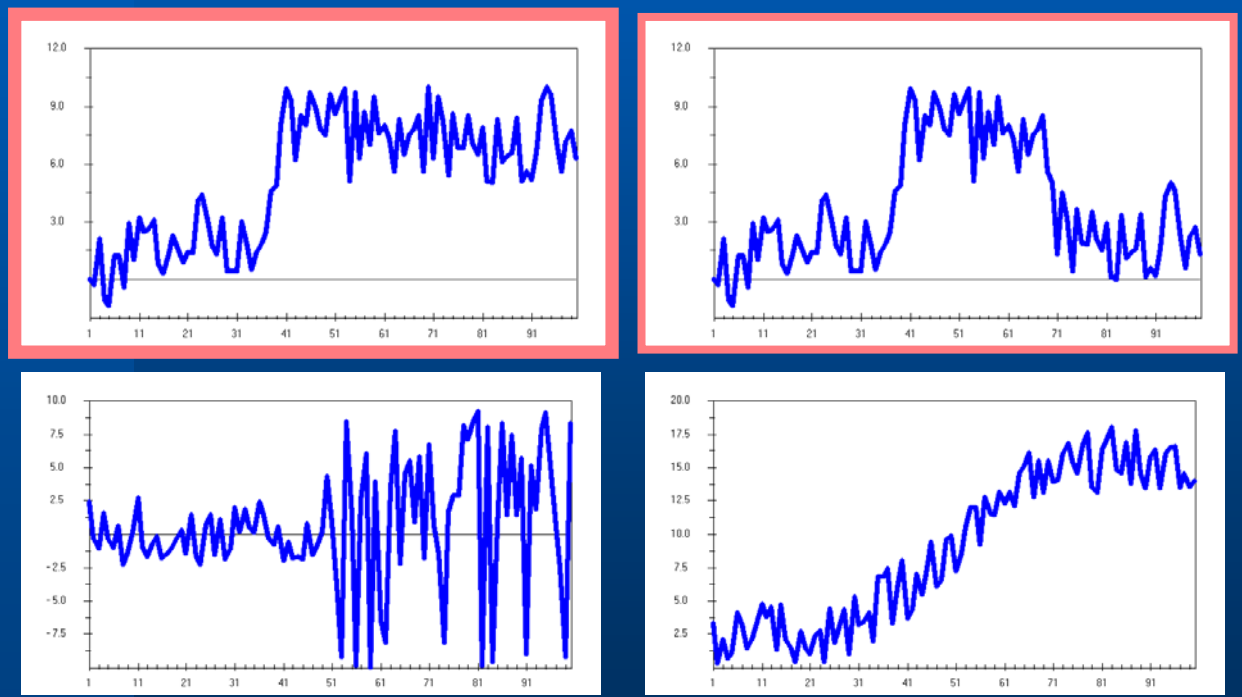
Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)

Homogeneity testing

- **Change in shift detection, p=0.05**



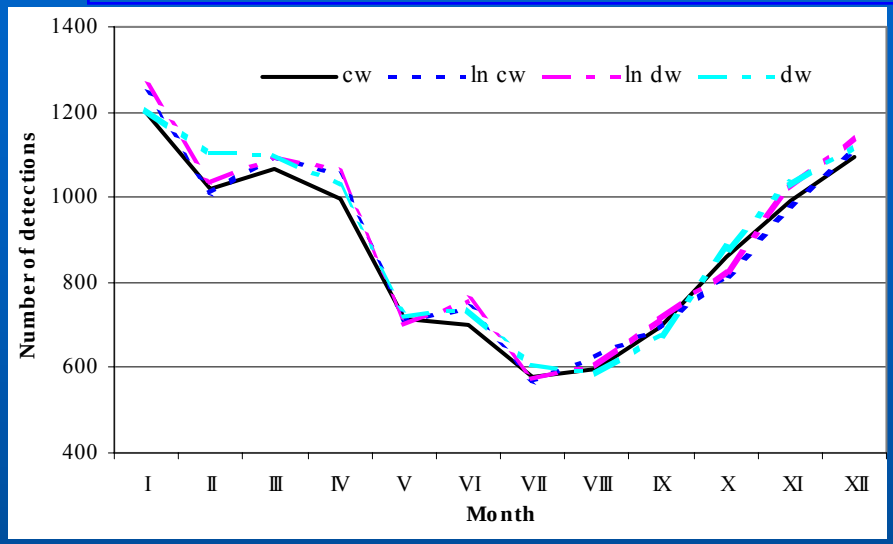
Splitting series into more parts

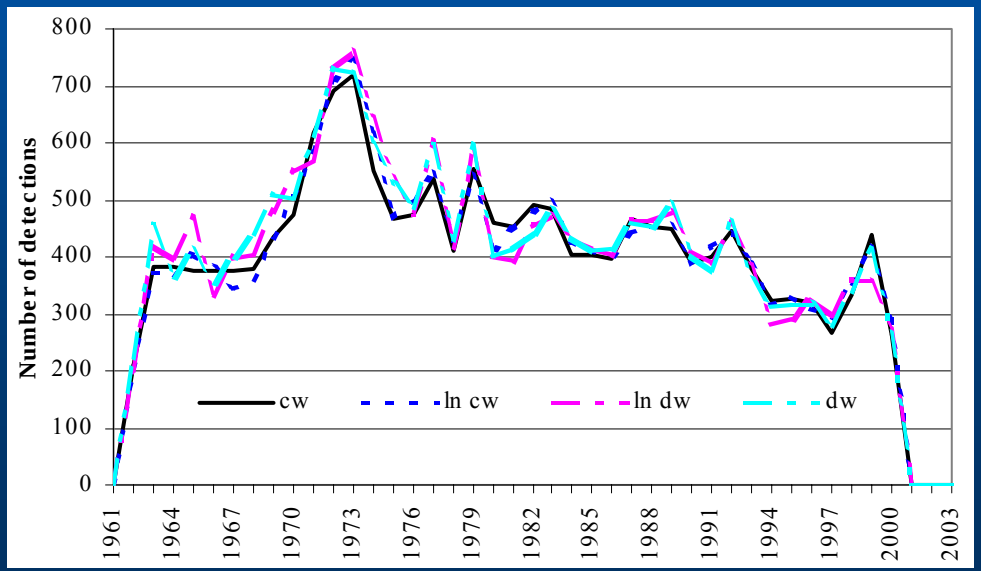
Relative homogeneity testing

- **Used tests:**
  - **Alexandersson SNHT**
  - **Bivariate test of Maronna and Yohai**
  - **Mann – Whitney – Pettit test**
  - **t-test**

Tests results

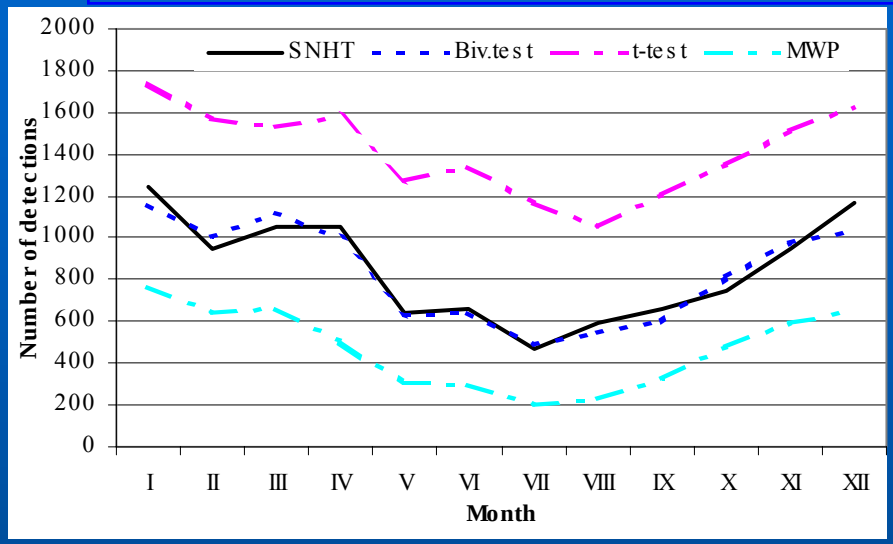
Differences among individual **reference series**

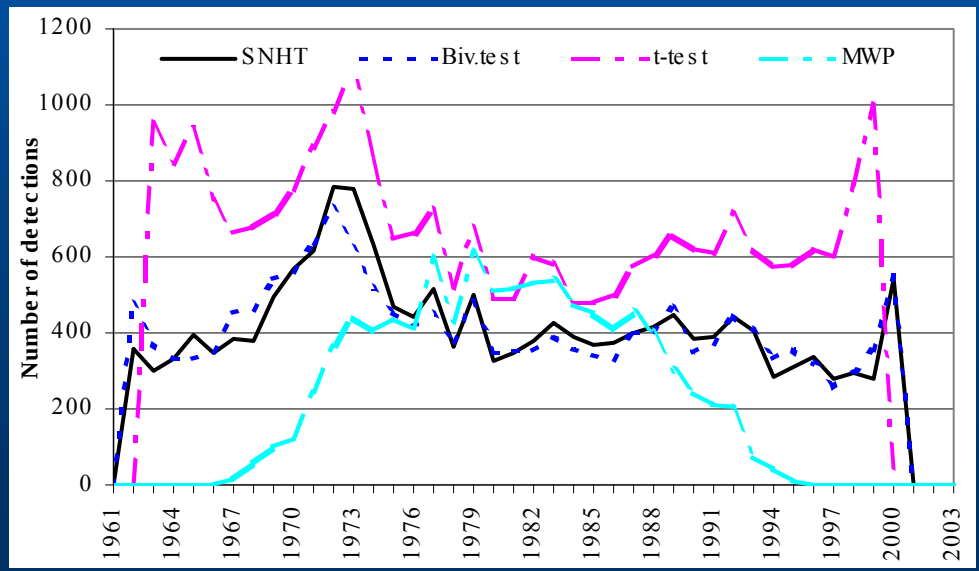




Tests results

Differences among individual **tests**





Homogeneity assessment

- Various outputs created for better inhomogeneities assessment
- Combining results with information from metadata
- Decision about „undoubted“ inhomogeneities

Example I:

Homogeneity assessment

ID	REFERENCE	ELEM	TEST	BEGIN	END	CO	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Win	Spr	Sum	Aut	Year
B1BOJK01	B1BOJK01_1_dw	x	A	1961	2000		1973	x	1985	2000	x	x	x	x	x	x	x	x					
B1BOJK01		x	A				1988*<										1988*<						
B1BOJK01	B1BOJK01_1_dw	x	As	1961	2000														1994	1985	x	x	1994
B1BOJK01		x	As																1971*<		1992*<		
B1BOJK01	B1BOJK01_1_dw	x	B	1961	2000		1973	x	1985	2000	x	x	x	x	x	x	x	x					
B1BOJK01		x	B				1988*<										1988*<						
B1BOJK01	B1BOJK01_1_dw	x	Bs	1961	2000														1997	1985	x	x	x
B1BOJK01		x	Bs																1966*<		1992*<		
B1BOJK01	B1BOJK01_1_dw	x	t_F	1961	2000		1973	1997	1985	1998	1985	1968	1966	1963	x	x	x	1996 <					
B1BOJK01	B1BOJK01_1_dw	x	t_Fs	1961	2000														1994	1985	x	x	1994
B1BOJK01	B1BOJK01_1_dw	x	Uk	1961	2000		1973	x	1985	x	x	x	x	x	x	x	x	x					
B1BOJK01	B1BOJK01_1_dw	x	Uks	1961	2000														x	1985	x	x	1985
B1BOJK01	B1BOJK01_1_cw	x	A	1961	2000		1973	x	1985	x	x	2000	1966	x	x	x	x	x					
B1BOJK01		x	A				1988*<										1988*<						
B1BOJK01	B1BOJK01_1_cw	x	As	1961	2000														1994	1985	x	x	x
B1BOJK01		x	As																1971*	1981*<			1981*
B1BOJK01	B1BOJK01_1_cw	x	B	1961	2000		x	x	1985	x	x	x	x	x	x	x	x	x					
B1BOJK01		x	B				1988*<							1982*<			1985*<						
B1BOJK01	B1BOJK01_1_cw	x	Bs	1961	2000														1994	1985	x	x	x
B1BOJK01		x	Bs																1971*<				1981*
B1BOJK01	B1BOJK01_1_cw	x	t_F	1961	2000		1973	1997	1985	1998	x	1992	1966	x	1999	x	x	1993 <					
B1BOJK01	B1BOJK01_1_cw	x	t_Fs	1961	2000														1994	1985	x	x	1985
B1BOJK01	B1BOJK01_1_cw	x	Uk	1961	2000		1973	x	1985	x	x	x	x	x	x	x	x	x					
B1BOJK01	B1BOJK01_1_cw	x	Uks	1961	2000														1988	1985	x	x	1985

Example II:

Homogeneity assessment

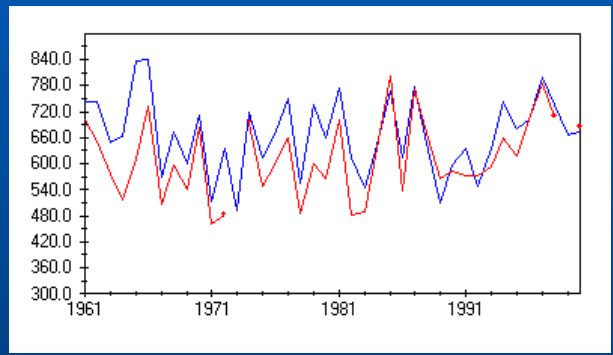
ID	BEGIN	END	LENGTH	YEAR_INHOM	YEAR_COUNT	Y_PORTION	Y_POSSIBL
B1KROM01	1961	2000	40		104	36.24	36.11
B1KROM01				1977	18	6.27	6.25
B1KROM01				1978	13	4.53	4.51
B1KROM01				1975	11	3.83	3.82
B1KROM01				1974	10	3.48	3.47
B1KROM01				1983	8	2.79	2.78
B1KROM01				1987	8	2.79	2.78
B1KROM01				1989	7	2.44	2.43
B1KROM01				1988	5	1.74	1.74
B1KROM01				1971	4	1.39	1.39
B1KROM01				1962	3	1.05	1.04
B1KROM01				1982	3	1.05	1.04
B1KROM01				1972	3	1.05	1.04
B1KROM01				1964	3	1.05	1.04
B1KROM01				1973	2	0.70	0.69
B1KROM01				1986	2	0.70	0.69
B1KROM01				1963	1	0.35	0.35
B1KROM01				1984	1	0.35	0.35
B1KROM01				1965	1	0.35	0.35
B1KROM01				1995	1	0.35	0.35
B1KROM01	1962	1965	4		8	2.79	2.78
B1KROM01	1971	1975	5		30	10.45	10.42
B1KROM01	**1977**	**1978**	**2**		**31**	**10.80**	**10.76**
B1KROM01	1982	1984	3		12	4.18	4.17
B1KROM01	1986	1989	4		22	7.67	7.64
B1KROM01	1961	1970	10		8	2.79	2.78
B1KROM01	1971	1980	10		61	21.25	21.18
B1KROM01	1981	1990	10		34	11.85	11.81
B1KROM01	1991	2000	10		1	0.35	0.35

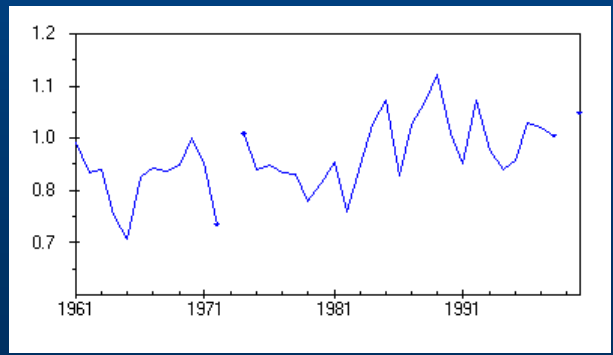
Summed numbers of detections for individual years

Homogeneity assessment

- **Deciding which years to adjust for inhomogeneities (using metadata, plots, …)**

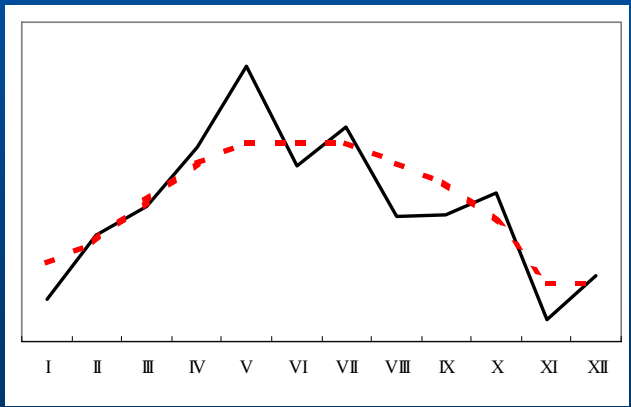
	ID	EL	YEAR	BEGIN	END	YEAR_COUN	Y_POSSIBL	YEA	MIS	X_BEGIN_D	X_END_DA	X	X	L	L	A	B	REMARK	C	C
x	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.3.2003	#	#					change		
	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.12.9999	#	#					obs	V	B
	B1BYSH01	x	1978			37	12.85													
?	B1BYSH01	x	1979			33	11.46													
?	B1BYSH01	x	1980			43	14.93													
?	B1HLHO01	x	1965			31	10.76	4	1											
	B1HOLE01	x	1976			33	11.46													
	B1KROM01	x		1977	1978	31	10.76													
x	B1RADE01	x	1994			44	15.28		2	1.1.1994	31.12.9999	#	#					change		
	B1RADE01	x	1994			44	15.28		2	1.1.1994	31.12.9999	#	#					obs	J	B
x	B1RYCH01	x	1973			49	17.01			1.5.1973	28.2.1991	#	#					change		
	B1RYCH01	x	1973			49	17.01			1.9.1972	28.2.1991	#	#					obs	M	B
xx?	B1STRZ01	x	1987			53	18.40													
	B1STRZ01	x	1988			30	10.42													
	B1UHBR01	x	1983			31	10.76			18.2.1984	31.1.1999	#	#					change		
	B1UHBR01	x	1983			31	10.76			18.2.1984	12.5.1993	#	#					obs	J	B
x	B1UHBR01	x	1984			77	26.74			18.2.1984	31.1.1999	#	#					change		
	B1UHBR01	x	1984			77	26.74			18.2.1984	12.5.1993	#	#					obs	J	B
	B1VELI01	x	1978			31	10.76													
?	B1VELI01	x		1977	1978	44	15.28													
?	B1VKLO01	x	1984			29	10.07													
x	B1VYSK01	x	1999			32	11.11	-1		1.4.1998	31.12.9999	#	#					change		
	B1VYSK01	x	1999			32	11.11	-1		1.4.1998	31.12.9999	#	#					obs	V	B
	B2BOSK01_1	x	1968			33	11.46													
	B2BREC01	x	1968			35	12.15													
	B2BRUM01	x	1989			51	17.71			1.2.1989	31.3.1994	#	#					change		
	B2BRUM01	x	1989			51	17.71			1.2.1989	31.3.1994	#	#					obs	M	B





Adjusting data

- using reference series based on correlations
- adjustment: from ratios 20 years before and after a change, monhtly
- smoothing monthly adjustments (low-pass filter for adjacent values)



Several iterations

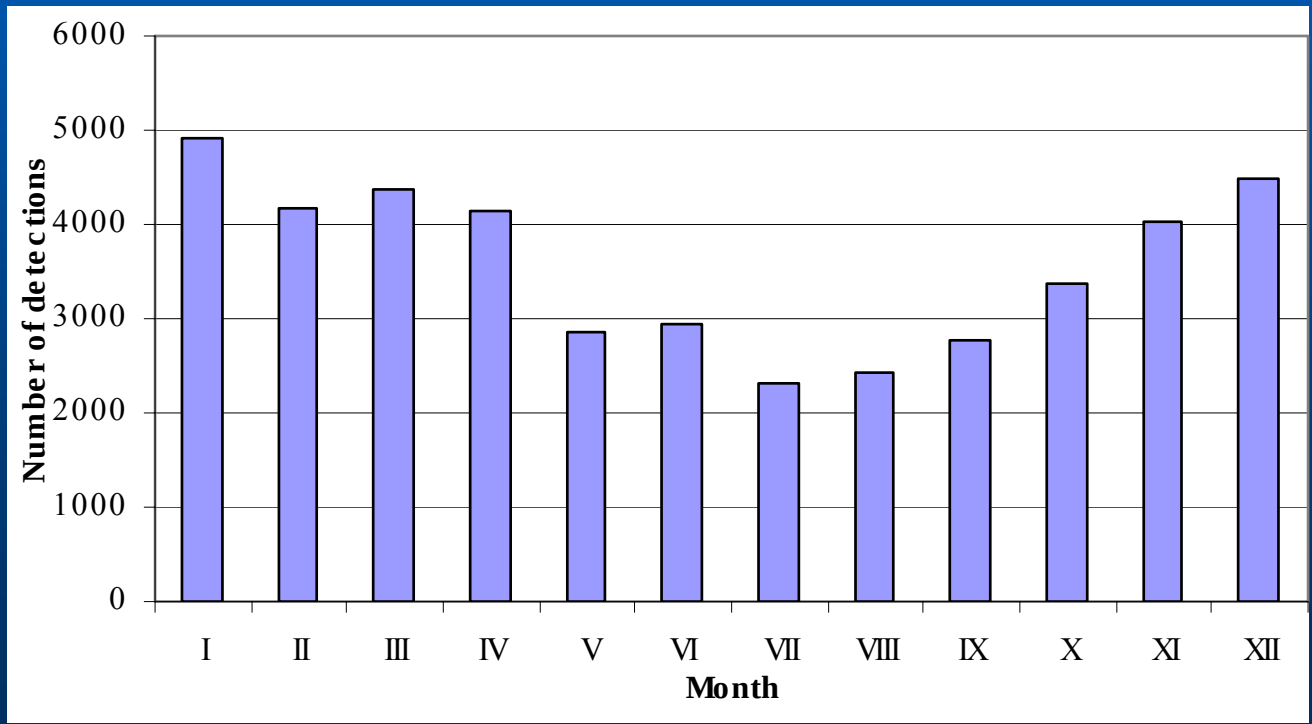
- **several iterations of homogeneity testing and series adjusting (3 iterations should be sufficient)**
- **question of homogeneity of reference series is thus solved:**
  - **possible inhomogeneities should be eliminated by using averages of several neighbouring stations**
  - **if this is not true: in next iteration neighbours should be already homogenized**

Filling missing values

- **linear regression (tested and reference series), monthly**
- **10 years before and after filled value**

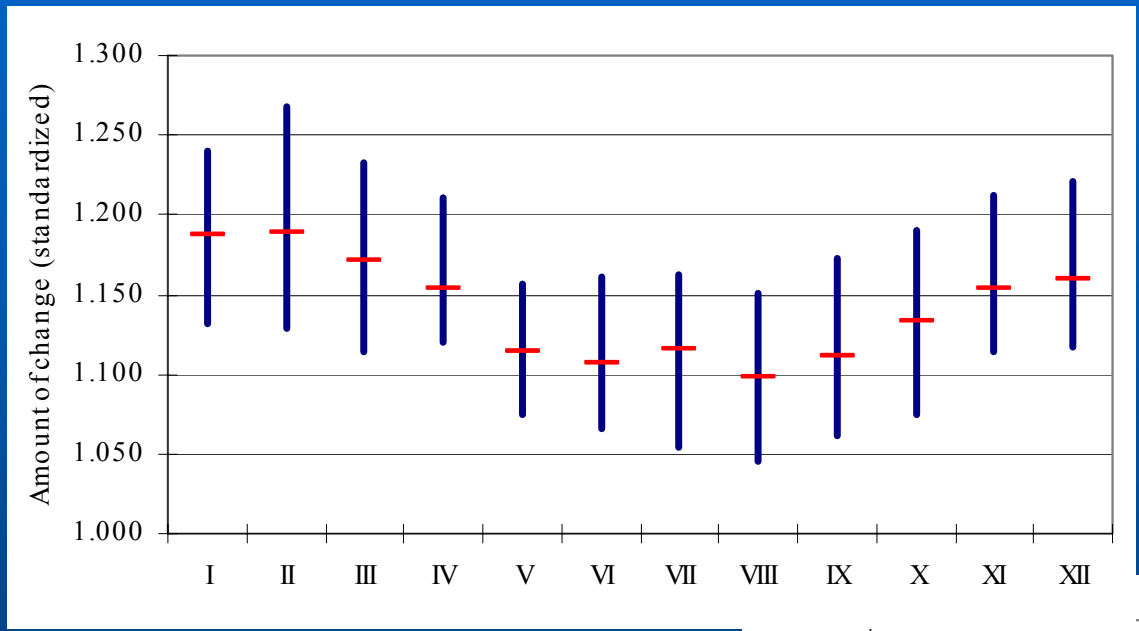
Homogeneity testing results

- 4 tests, 4 reference series, 12 months + 4 seasons and year
- Number of detected inhomogeneities (significant)



Amount of change (ratios – standardized to be >1.0)

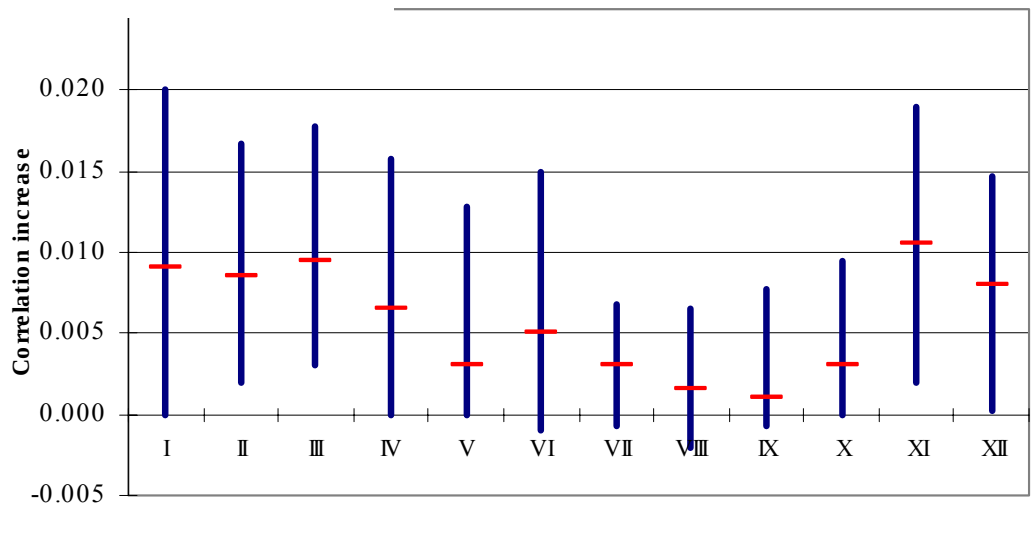
(reference series calculation based on correlations)



Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)



Final remarks

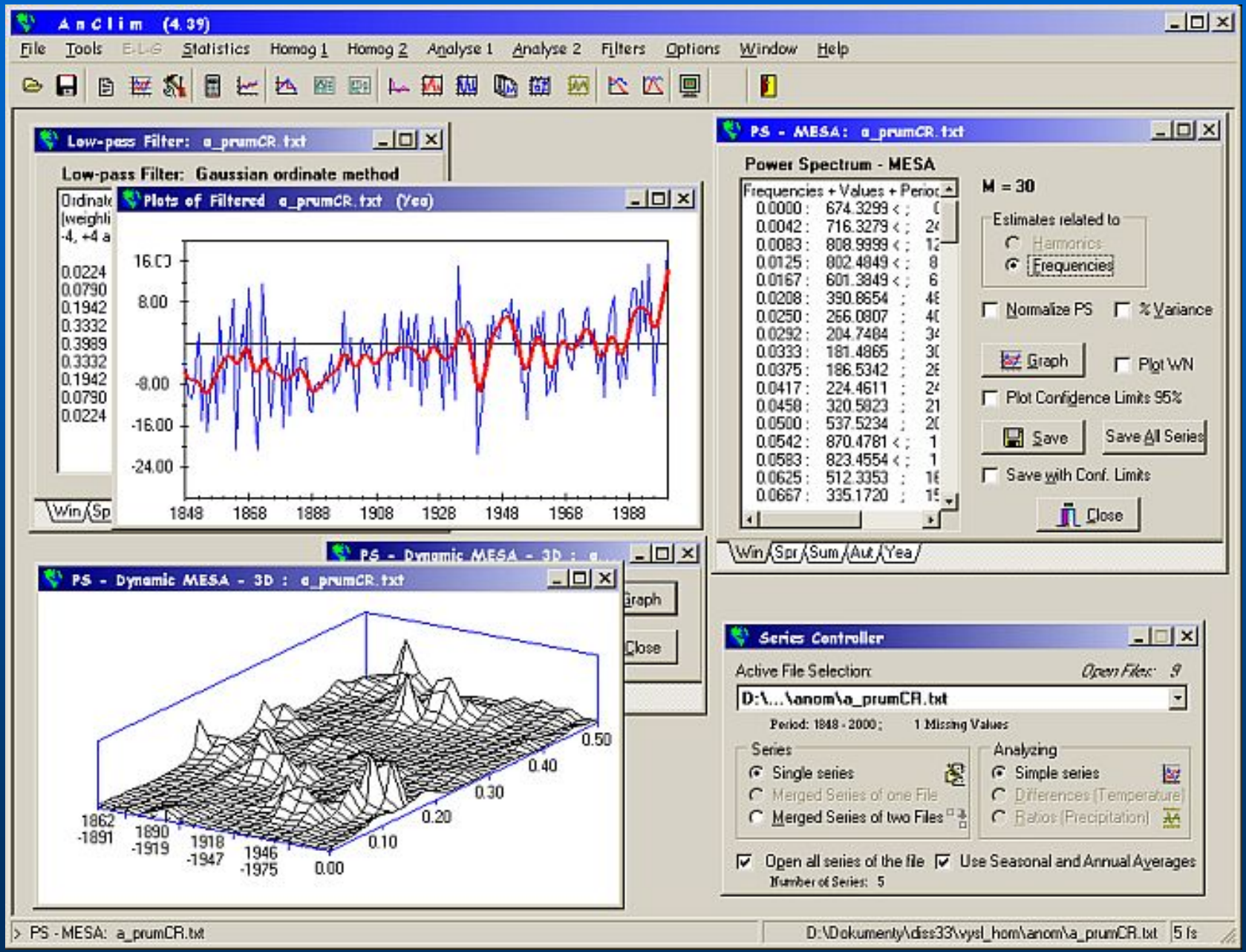
- problem: detection of inhomogeneites near ends of series …
- open question: impact of automatization … (introduced only few years ago)

Software used for data processing

- **Application for downloading data from central database (e.g. Oracle)**
- **ProClimDB software for processing whole dataset (finding outliers, combining series, creating reference series, preparing data for homogeneity testing, …)**
- **AnClim software for homogeneity testing**

http://www.klimahom.com/software

AnClim software



ProcData software