Homogenization of monthly and daily climatological time series

Petr Štěpánek

Pavel Zahradníček

Czech Hydrometeorological Institute, Czech Republic

E-mail: petr.stepanek@chmi.cz

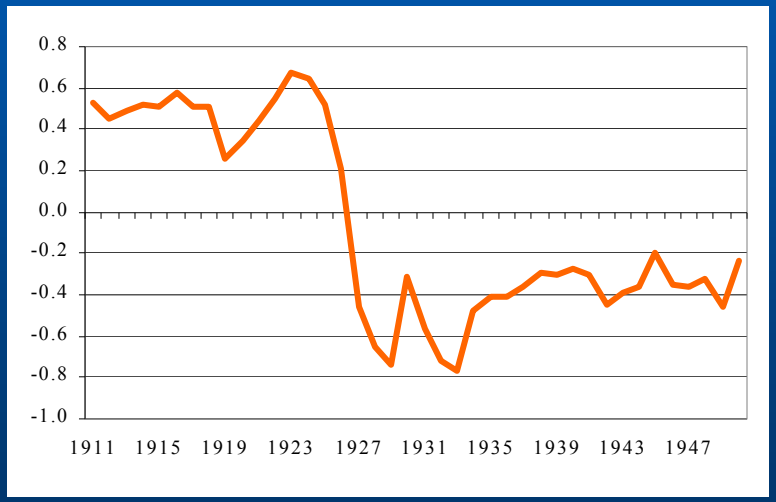
1st Joint DMCSEE-JRS Workshop on Drought Monitoring, Ljubljana, Slovenia, 21.-25. Sept. 2009

Climatological studies

- **Measuring and collecting data**

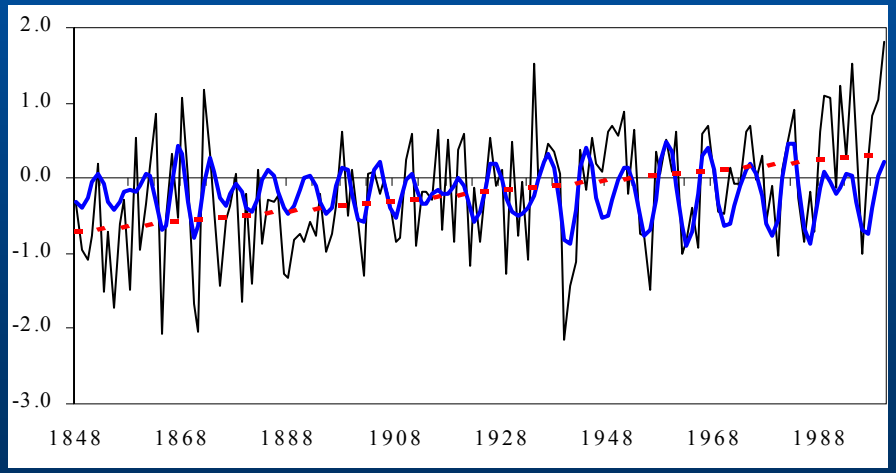
Climatological studies

- **Measuring data**
- **Data quality control and Homogenization**

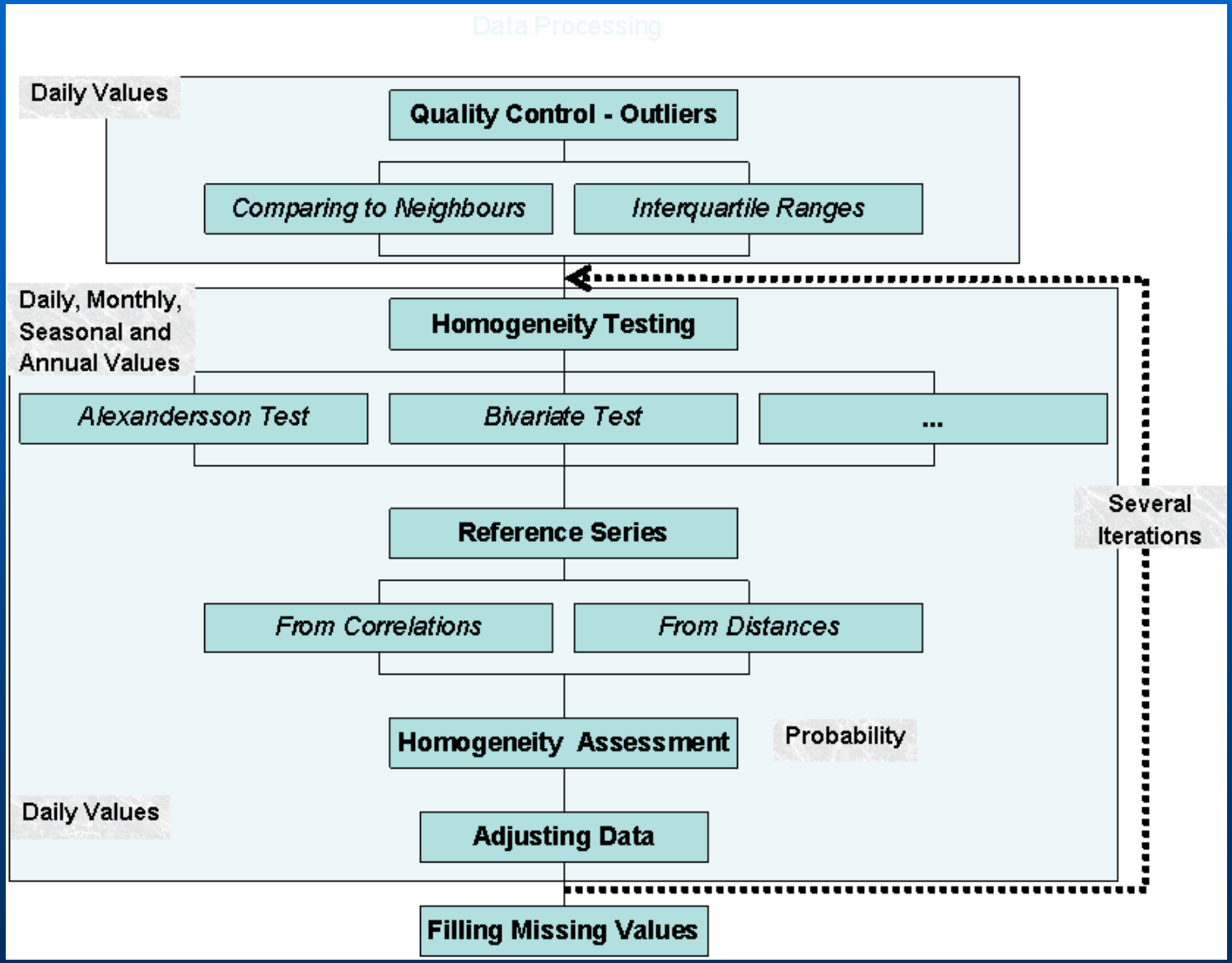


Climatological studies

- **Measuring data**
- **Homogenization**
- **Data Analysis**



Processing before any data analysis

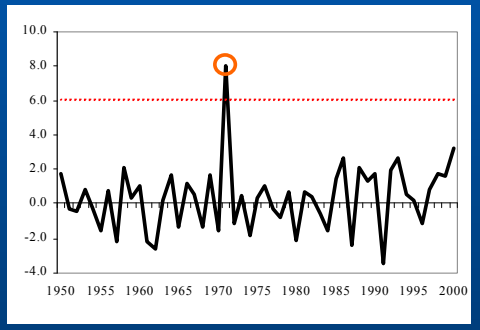


Software AnClim, ProClimDB

Data Quality Control
Finding Outliers

Two main approaches:

- **Using limits derived from interquartile ranges (time series)**

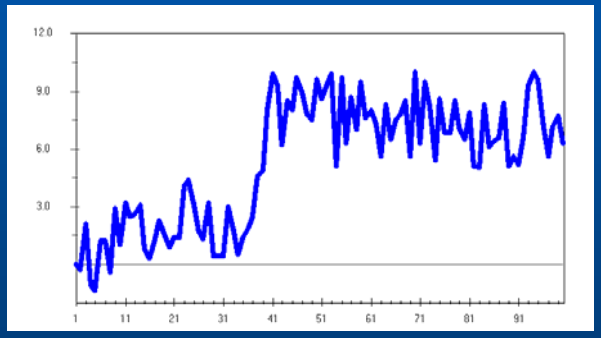


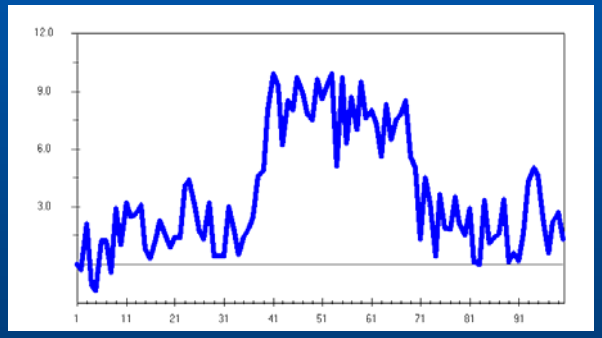
- **comparing values to values of neighbouring stations (spatial analysis)**

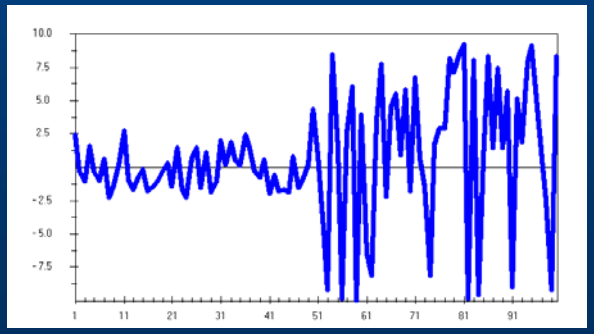
Homogenization

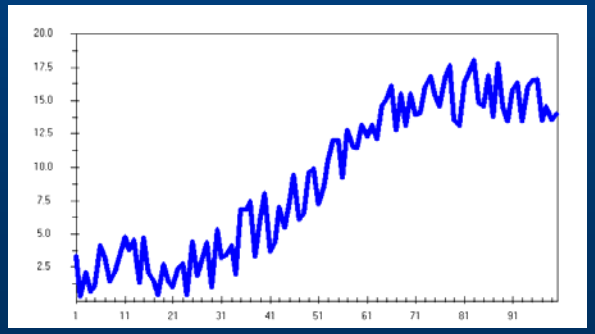
- **Change of measuring conditions**

  ⟶ inhomogeneities



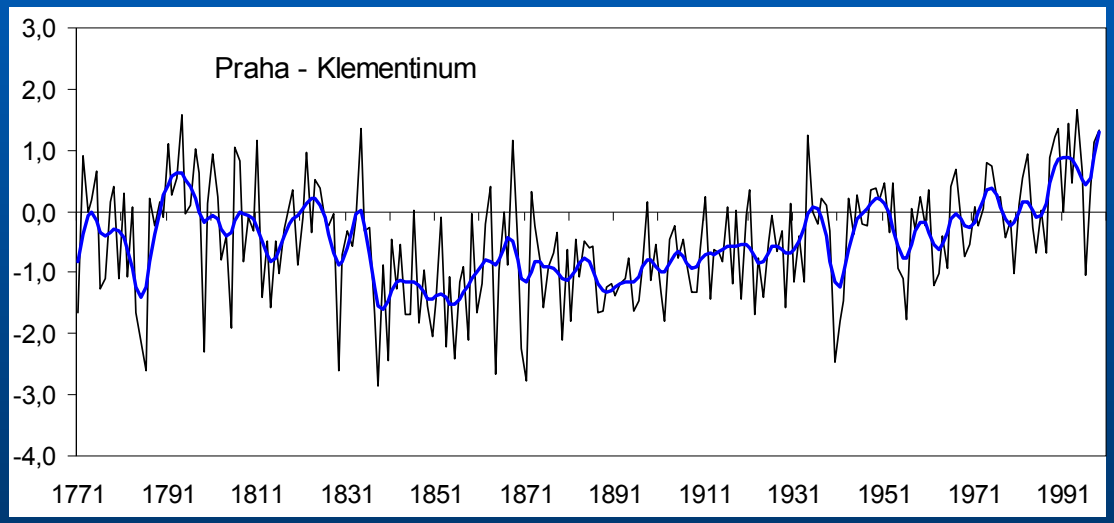






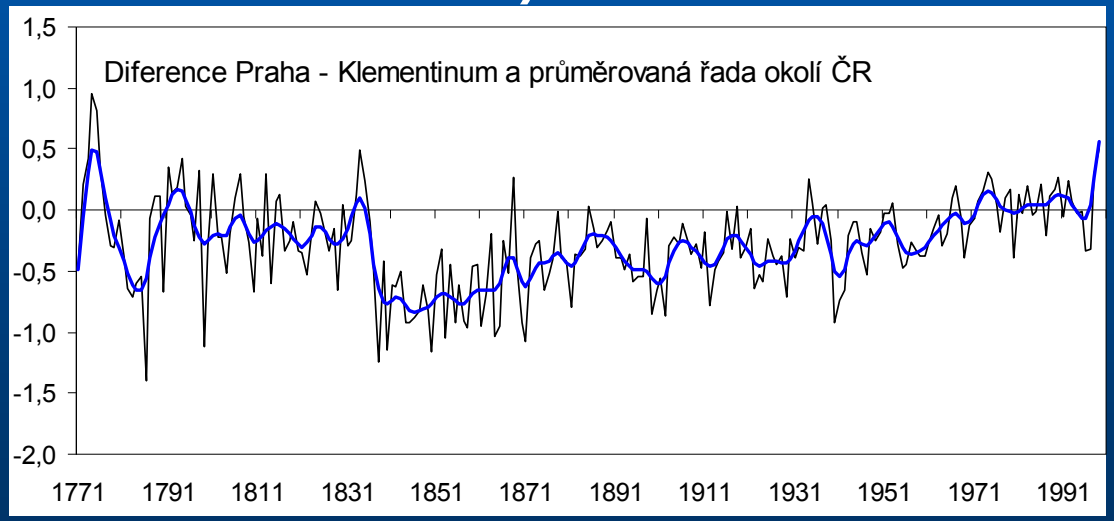
Inhomogeneity Detection

- **Absolute Homogeneity Testing**



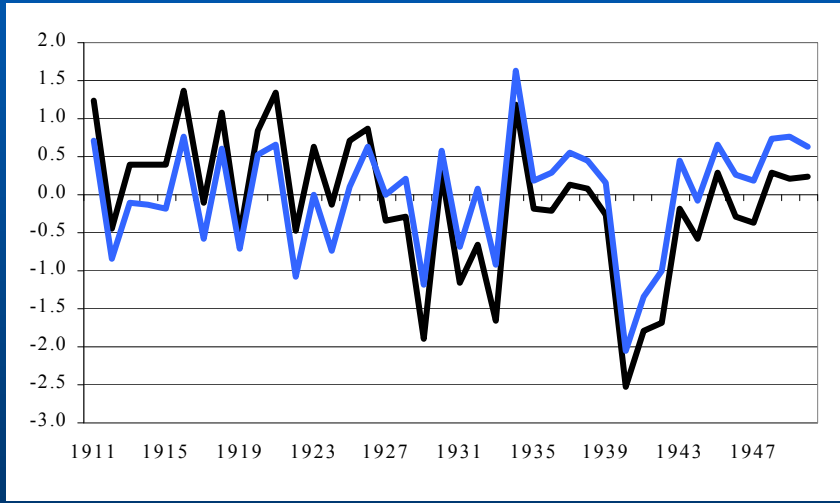
Inhomogeneity Detection

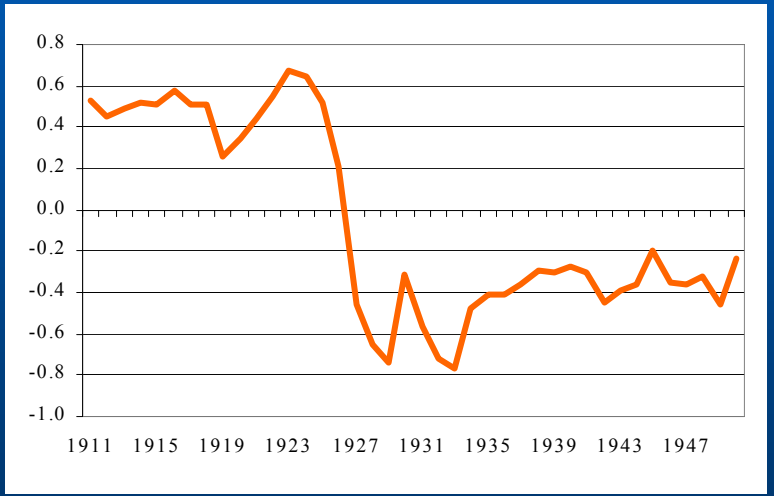
- **Absolute Homogeneity Testing**
- **Relative Homogeneity Testing**
  (differences with reference series)



Relative Homogeneity Testing

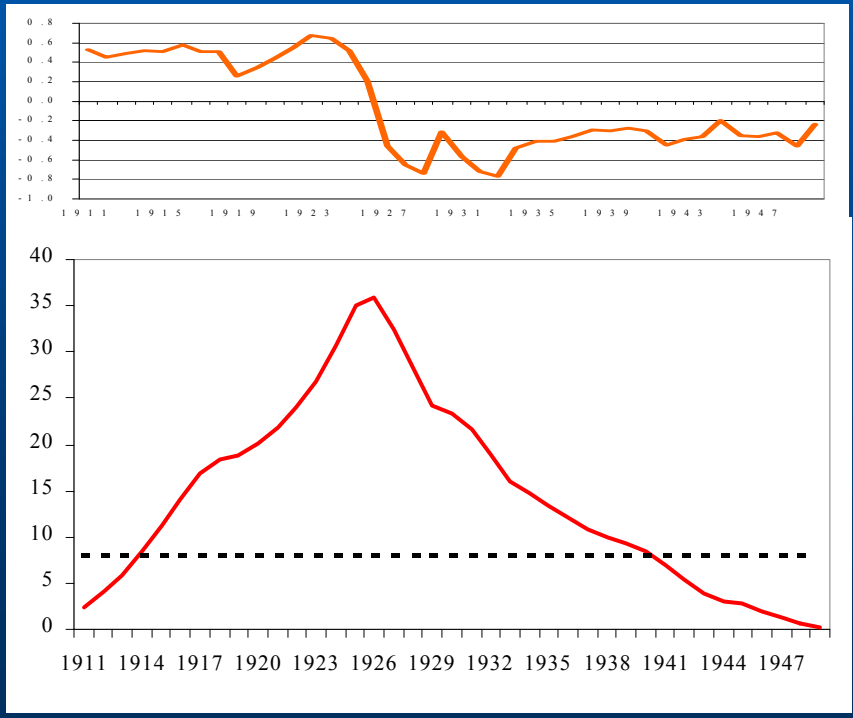
- **Creating reference Series**





Relative Homogeneity Testing

- **Creating reference Series**
- **Tests of homogeneity**



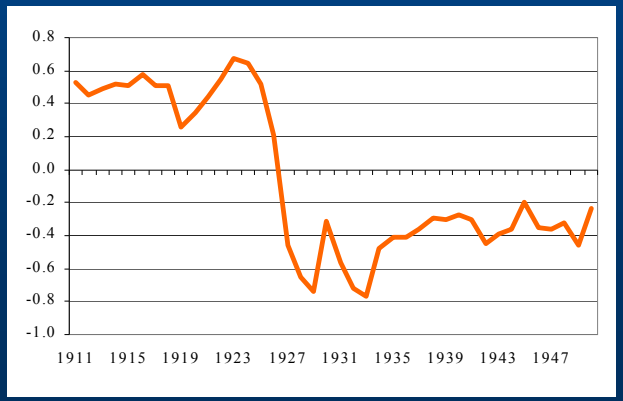
Relative Homogeneity Testing

- **Creating reference Series**
- **Tests of homogeneity**
- **Assessing homogeneity**

- Metadata

- Physically justified

- “undoubted” inhomogeneity



Relative Homogeneity Testing

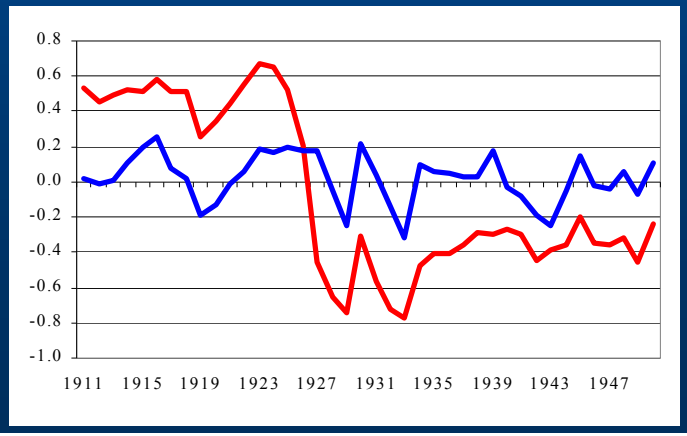
- **Creating reference Series**
- **Tests of homogeneity**
- **Assessing homogeneity**

  - Metadata

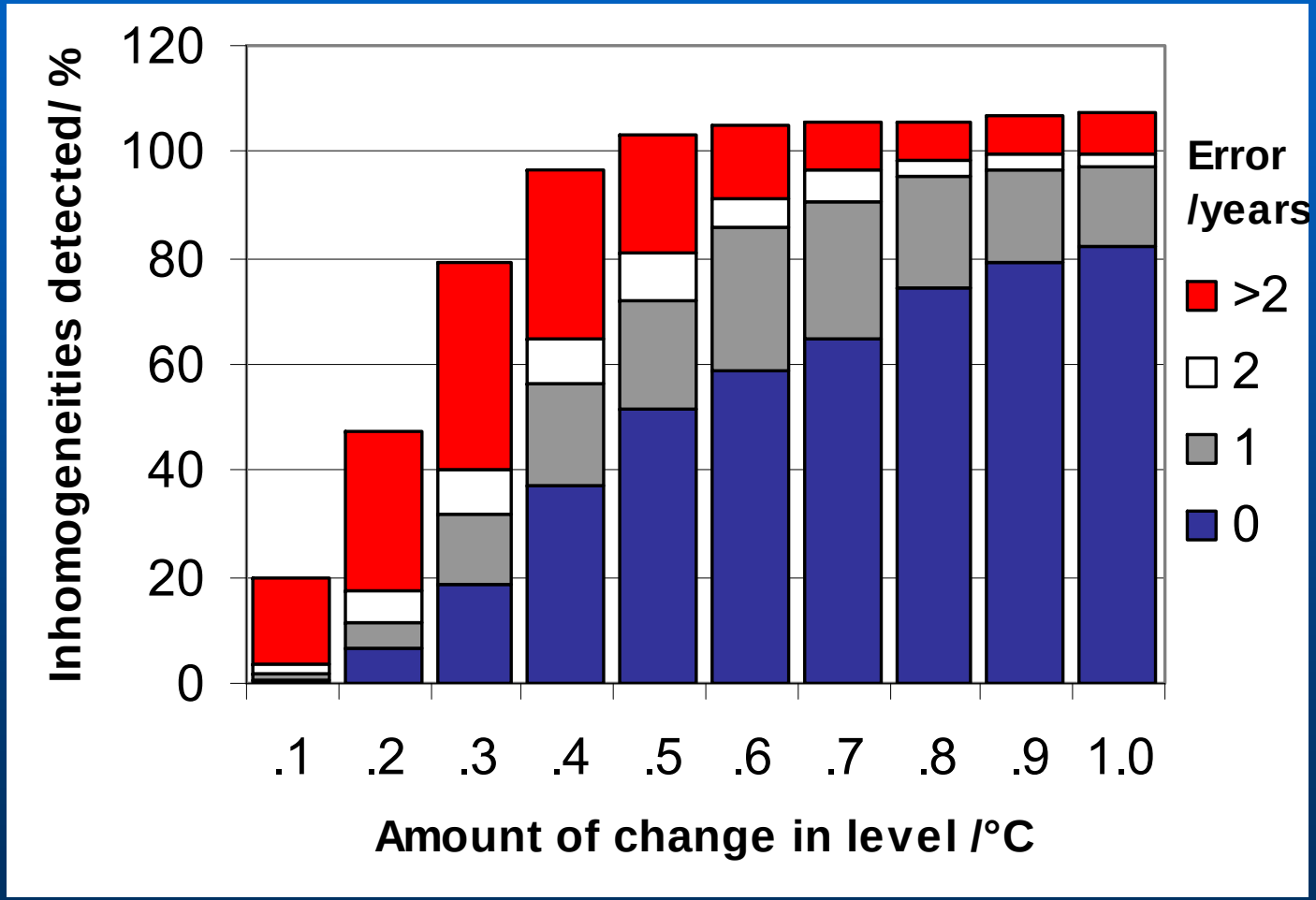
  - Physically justified

  - "undoubted" inhomogeneity

- **Adjusting Series**



Inhomogeneity Detecting by SNHT (p=0.05, 950 series)



Assessing Homogeneity – Problems:

- **most of metadata incomplete**

→ **we depend upon statistical tests results**

Assessing Homogeneity – Problems:

- **most of metadata incomplete**

  → **we depend upon statistical tests results**

- **uncertainty in test results**
  - right inhomogeneity detection is problematic
  (for smaller amount of change)

Proposed solution

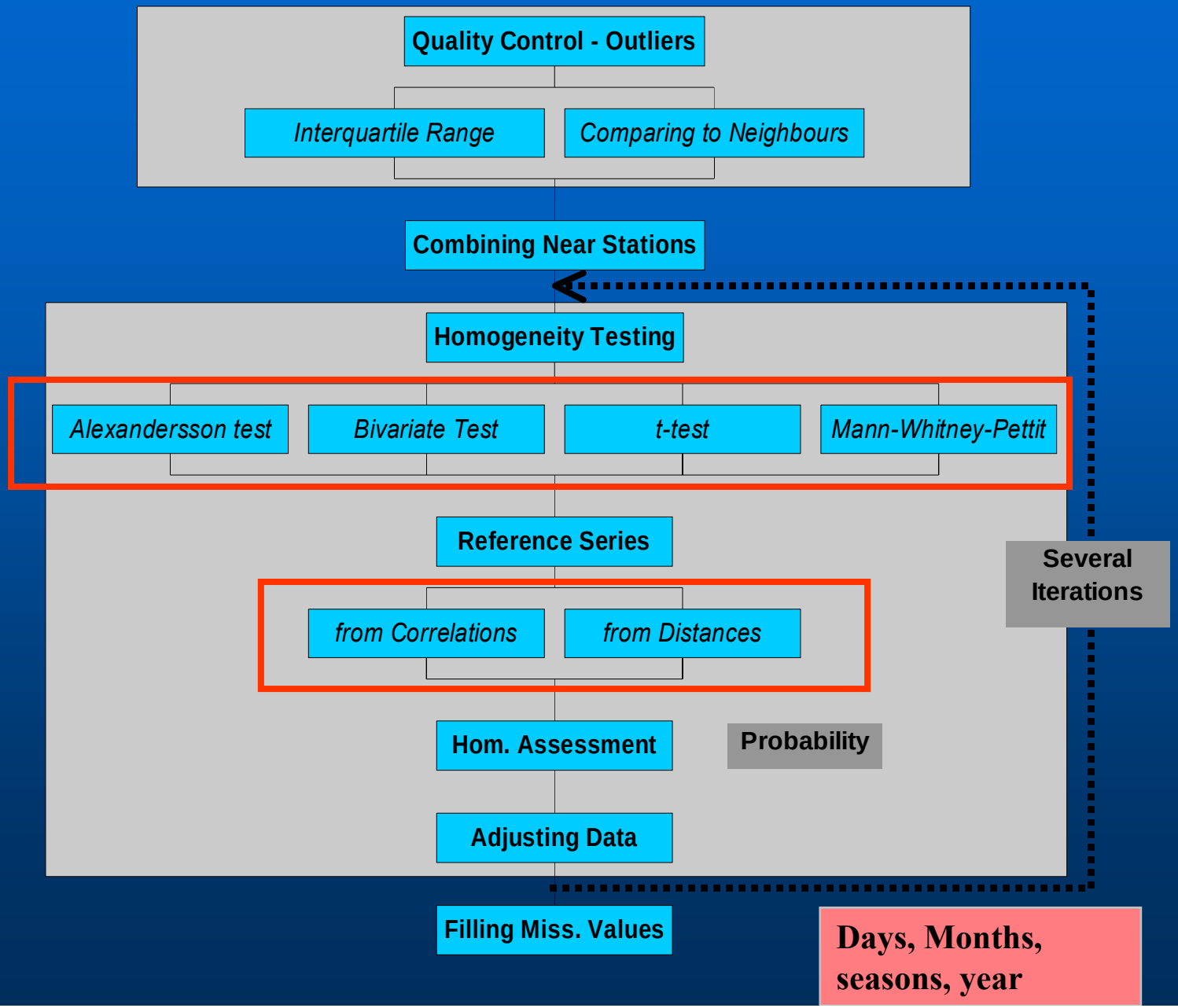
- To get as many test results for each candidate series as possible

→ **„Ensemble“ approach - processing of big amount of test results for each individual series**

Adventages of the „Ensemble“ approach

- **we know relevance (probability) of each inhomogeneity**
- **we can easily assess quality of measurements for series as a whole**

How to increase number of test results



Creating Reference Series

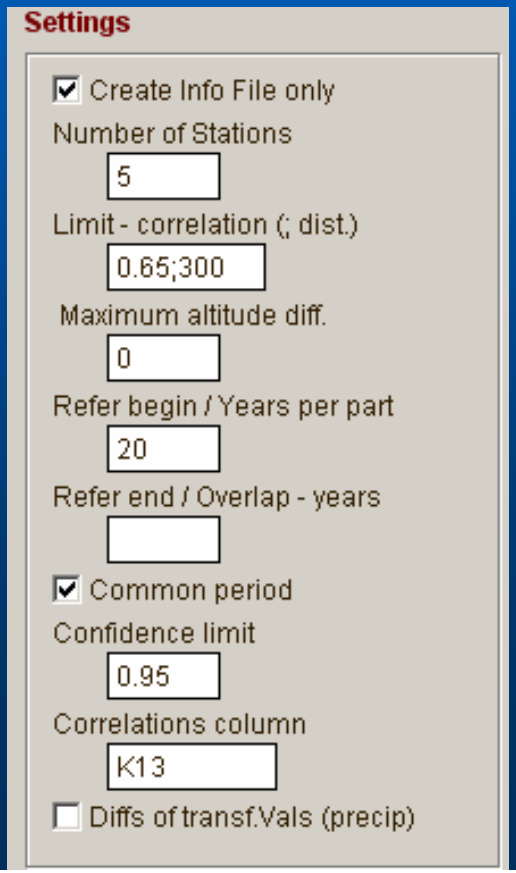
- for monthly, daily data (each month individually)
- weighted/unweighted mean from neighbouring stations
- criterions used for stations selection (or combination of it):
  - best correlated / nearest neighbours

    (correlations – from the first differenced series)
  - limit correlation, limit distance
  - limit difference in altitudes

- neighbouring stations series should be standardized to test series
  AVG and / or STD

  (temperature - elevation, precipitation - variance)

  - **missing data are not so big problem then**



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, distances and altitudes

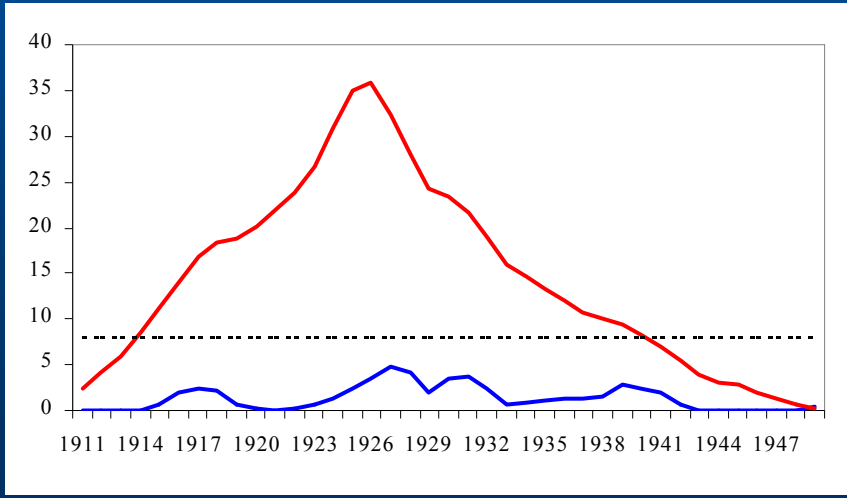
Relative homogeneity testing

- **Available statistical tests:**
  - **Alexandersson SNHT**
  - **Bivariate test of Maronna and Yohai**
  - **Mann – Whitney – Pettit test**
  - **t-test**
  - **Easterling and Peterson test**
  - **Vincent method**
  - **…**

20 year parts of the daily series **(40 for monthly series with 10 years overlap),**
in SNHT splitting into subperiods in position of detected significant changepoint
(30-40 years per one inhomogeneity)

Homogeneity Tests

Alexandersson's SNHT



Alexandersson Standart Normal Homogeneity Test

(Single shift test)

Reference series:

$$q_i = Y_i / \{[\sum_{j=1}^{k} \rho_j{}^2 X_{ji} \bar{Y} / \bar{X}_j] / \sum_{j=1}^{k} \rho_j{}^2\}$$

$$q_i = Y_i - \{\sum_{j=1}^{k} \rho_j{}^2 [X_{ji} - \bar{X}_j + \bar{Y}] / \sum_{j=1}^{k} \rho_j{}^2\}$$

Null and alternative hypothesis:

$H_0 : z_i \in N(0,1), \quad i \in \{1,..,n\}$.

$H_1 : z_i \in N(\mu_1,1), \quad i \in \{1,..,a\}$,

$z_i \in N(\mu_2,1), \quad i \in \{a+1,..,n\}$,

for $1 ? a < n$ a $\mu_1 ? \mu_2$.

$z_i = (q_i - \bar{q})/s_q$, $z_i \in N(0,1)$

Test statistic:

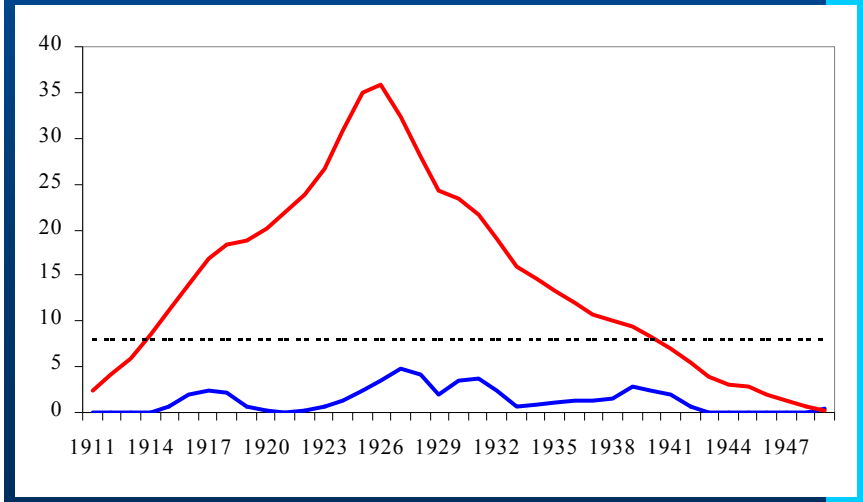
$$T_0 = \max_{1 \le a < n-1} \{T_a\} = \max_{1 \le a < n-1} \{a\bar{z}_1^2 + (n-a)\bar{z}_2^2\}$$

where $\bar{z}_1 = \frac{1}{a}\sum_{i=1}^{a} z_i$, $(\bar{z}_1 ? \mu_1)$,

$\bar{z}_2 = \frac{1}{(n-a)}\sum_{i=a+1}^{n} z_i$, $(\bar{z}_2 ? \mu_2)$.

Homogeneity Tests

Bivariate Test of Maronna and Yohai



Bivariate Test

Null and alternative hypothesis:

H_0 : vectors $\{x_i, y_i\}$ bivariate normal distributed
$N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$

H_1 : pro $0<i_0<n$ a d ? 0 -
$N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ pro i ? i_0
$N(\mu_x, \mu_y+d, \sigma_x^2, \sigma_y^2, \rho)$ pro $i > i_0$.

Test statistic:

$$T_0 = \max_{i<n}\{T_i\}$$

where: $X_i = 1/i\sum_{j=1}^{i} x_j$, $Y_i = 1/i\sum_{j=1}^{i} y_j$, $\overline{X} = X_n$, $\overline{Y} = Y_n$

$$S_x = \sum_{j=1}^{n}(x_j - \overline{X})^2 \text{ , } S_y = \sum_{j=1}^{n}(y_j - \overline{Y})^2 \text{ , } S_{xy} = \sum_{j=1}^{n}(x_j - \overline{X})(y_j - \overline{Y}) \text{ ,}$$

$$F_i = S_x - (X_i - \overline{X})^2 ni/(n-i) \text{ , } i<n \text{ ,}$$

$$D_i = S_x(\overline{Y} - Y_i) - S_{xy}(\overline{X} - X_i)n/\left[(n-i)F_i\right],$$

$$T_i = \left[i(n-i)D_i^2 F_i\right]/(S_x S_y - S_{xy}^2)$$

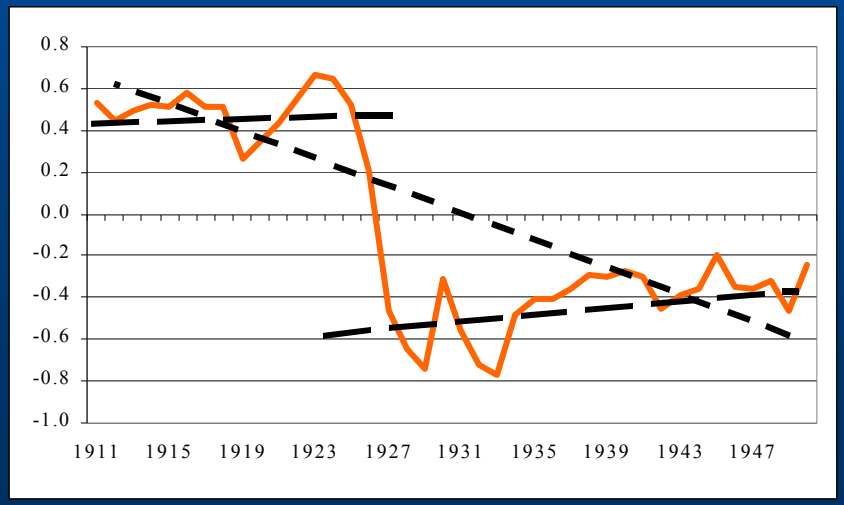
Homogeneity Tests

Two-phase linear regression
(Vincent Technique,
Easterling and Peterson test)

Easterling and Peterson

Test statistic: $U = [(RSS_1 - RSS_2)/3]/[RSS_2/(n-4)]$? F(3,n-4)

t-test: differences of levels before and after a discontinuity



Homogeneity assessment

- **Various outputs created for better inhomogeneities assessment**
- **Combining results with information from metadata whenever possible**
- **Decision about „undoubted“ inhomogeneities (without metadata) – coincidence of test results**

Homogeneity assessment

Output example: Station Čáslav, 3rd segment, 1911-1950, n=40

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

Homogeneity assessment, Output II example:

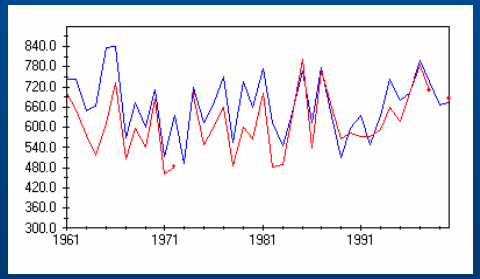
Begin	End	Length	InHomogeneity	Number	% detected inhom	% possible inhom	End	Missing
1911	1950	40		140	100	120		
			1927	60	43	51		
			1926	37	26	32		
			1928	9	6	8		4
			1937	7	5	6		
			1922	4	3	3		
			1935	4	3	3		
			1918	3	2	3		
			1930	3	2	3		
			1939	3	2	3		
			1940	3	2	3		2
			1938	2	1	2		
			1913	1	1	1	3	3
			1929	1	1	1		
			1931	1	1	1		
			1936	1	1	1		
			1944	1	1	1		
1926	1927	2		97	69	83		
1926	1931	6		111	79	95		
1935	1940	6		20	14	17		
1911	1920	10		4	3	3		
1921	1930	10		114	81	97		
1931	1940	10		21	15	18		
1941	1950	10		1	1	1		

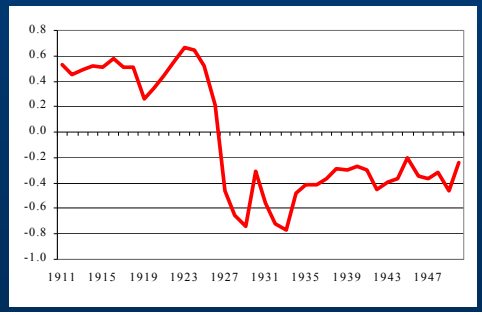
Summed numbers of detections for individual years

Homogeneity assessment

- **combining several outputs (sums of detections in individual years, metadata, graphs of differences/ratios, …)**

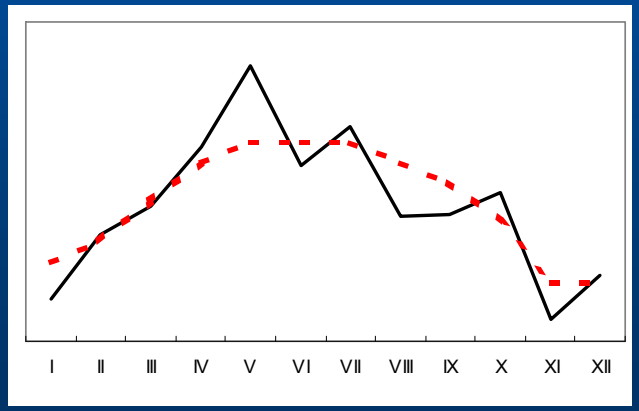
	ID	EL	YEAR	BEGIN	END	YEAR_COUN	Y_POSSIBL	YEA	MIS	X_BEGIN_D	X_END_DA	X	X	L	L	A	B	REMARK	C	C
x	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.3.2003	#	#				B	change		
	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.12.9999	#	#					obs	V	B
	B1BYSH01	x	1978			37	12.85													
?	B1BYSH01	x	1979			33	11.46													
?	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	#	#				F	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	#	#				V	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	#	#				U	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	#	#				U	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	#	#				V	change		
	B1VYSK01	x	1999			32	11.11	-1		1.4.1998	31.12.9999	#	#					obs	V	B
	B2BOSK01_	x	1968			33	11.46													
	B2BREC01	x	1968			35	12.15													
	B2BRUM01	x	1989			51	17.71			1.2.1989	31.3.1994	#	#				B	change		
	B2BRUM01	x	1989			51	17.71			1.2.1989	31.3.1994	#	#					obs	M	B





Adjusting monthly data

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



Example:

Adjusting values - evaluation

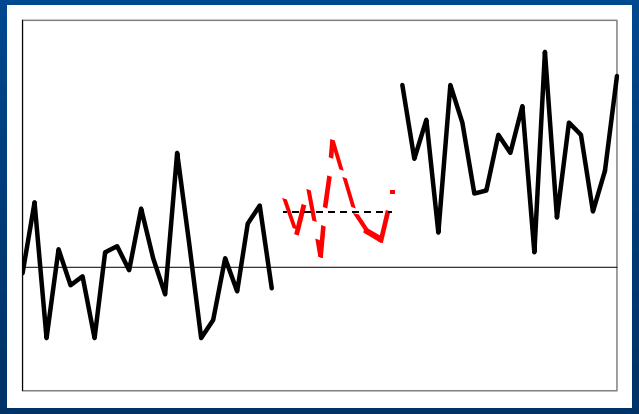
ID_1	pII	BEGIN	END	YEAR	MONTH	REMARK	C(	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12
B1RYCH01	E	1961	1992	1973	5	ADJust		1.135	1.197	1.155	1.333	1.149	1.070	1.088	1.354	1.145	1.116	1.136	1.265
B1RYCH01						DIFF1		0.905	0.875	0.912	0.813	0.906	0.956	0.896	0.786	0.912	0.956	0.908	0.855
B1RYCH01						DIFF2		1.027	1.048	1.053	1.084	1.041	1.024	0.975	1.064	1.045	1.067	1.032	1.081
B1RYCH01						corr		0.964	0.930	0.963	0.915	0.888	0.870	0.866	0.927	0.961	0.952	0.956	0.875
B1RYCH01						corr+		0.007	0.017	0.006	0.026	0.014	0.006	0.008	-0.001	-0.002	0.017	0.010	0.033
B1RYCH01						t		1.904	2.144	2.443	3.897	1.957	0.936	0.874	3.424	1.937	1.507	2.252	3.415
B1RYCH01						t_crit		2.042	2.048	2.045	2.045	2.045	2.045	2.042	2.042	2.042	2.042	2.042	2.045
B1RYCH01						Std_1		0.171	0.184	0.108	0.216	0.206	0.168	0.274	0.146	0.241	0.255	0.139	0.159
B1RYCH01						Std_2		0.178	0.235	0.181	0.169	0.175	0.209	0.232	0.256	0.146	0.164	0.157	0.185
B1RYCH01						t2		1.923	2.252	2.730	3.685	1.884	0.985	0.837	3.904	1.718	1.351	2.325	3.569
B1RYCH01						t2_crit		1.960	1.961	1.960	1.961	1.961	1.960	1.961	1.960	1.961	1.961	1.960	1.960
B1RYCH01						No_1		12	12	12	12	12	12	12	12	12	12	12	11
B1RYCH01						No_2		20	18	19	19	19	19	20	20	20	20	20	20
B1RYCH01						b1_1		-0.015	-0.016	0.002	0.017	0.028	0.002	-0.035	0.002	0.035	0.040	0.015	-0.012
B1RYCH01						b1_2		-0.007	-0.024	-0.002	0.001	-0.008	0.018	-0.022	-0.002	-0.007	-0.016	-0.014	-0.024
B1RYCH01		**> 2n:0.479,0.233**		**1973**	**5**	**ADJ_sm**		**1.180**	**1.178**	**1.206**	**1.238**	**1.172**	**1.107**	**1.149**	**1.229**	**1.185**	**1.138**	**1.162**	**1.199**
B1RYCH01						corr		0.964	0.930	0.963	0.915	0.888	0.870	0.866	0.927	0.961	0.952	0.956	0.875
B1RYCH01						corr+(AD		0.007	0.016	0.003	0.026	0.014	0.006	0.009	0.010	-0.005	0.019	0.009	0.030

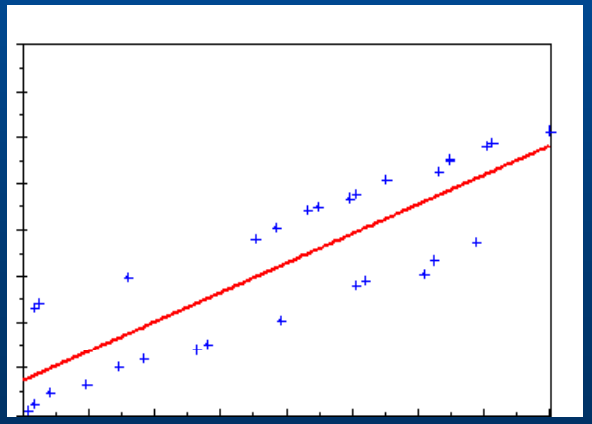
Iterative homogeneity testing

- **several iteration of testing and results evaluation**
  - **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

- **Before homogenization: influence on right inhomogeneity detection**
- **After homogenization: more precise - data are not influenced by possible shifts in the series**





Dependence of tested series on reference series

ACTION COST-ES0601: Advances in homogenisation methods of climate series: an integrated approach (HOME)

- 03/05/2007 - End date: 02/05/2011, Year: 3

- Inventory of existing detection and correction methods
- Compilation of a benchmark dataset with (un)known inhomogeneities
- Selection, comparison and evaluation of existing detection and correction (including those not traditionally used in climatology)
- Objective analysis of advantages and/or disadvantages of existing methods (benchmark)
- Investigation in further improvements of methods

- Documentation of practical recommendations
- Presentation and release of a new common method

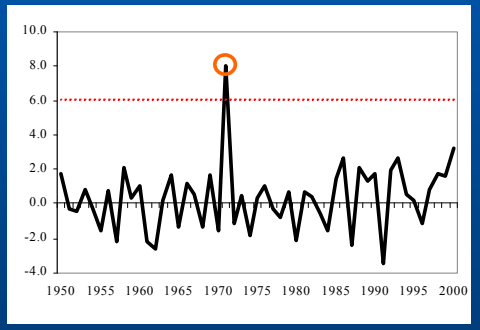
- http://www.homogenisation.org/

Quality control on example of series in the Czech Republic

Data Quality Control
Finding Outliers

Two main approaches:

- **Using limits derived from interquartile ranges (time series)**



- **comparing values to values of neighbouring stations (spatial analysis)**

Example:

Comparing base station to its neighbours

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	

Example of outputs for outliers assessment

Suspicious values

Expected value

Neighbour stations values

List of neighbours

Altitudes and distances of neighbours

	B	C	D	E	F	G	H	I	J	K	L	M	
	ID	YE	MONT	DA	ST_BASE	EXPECT_	REMAR	ST_1	ST_2	ST_3	ST_4	ST_5	D
0	B2BTUR01_T_03:30				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
0	B2BZAB01_T_03:30						st_1, di	11,58					
0	B1PROT01_T_03:30						st_2, di		36,85				
0	O3PRER01_T_03:30						st_3, di			59,12			
0	O2OLOM01_T_03:30						st_4, di				62,88		
0	O1CERV01_T_03:30						st_5, di					91,95	
0	**B2BTUR01_T_03:30**	**2006**	**6**	**25**	**27,30**	**17,28**		**17,30**	**16,10**	**15,50**	**15,80**	**16,10**	
5	B2BTUR01_T_03:45				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
5	B2BZAB01_T_03:45						st_1, di	11,58					
5	B1PROT01_T_03:45						st_2, di		36,85				
5	O3PRER01_T_03:45						st_3, di			59,12			
5	O2OLOM01_T_03:45						st_4, di				62,88		
5	O1CERV01_T_03:45						st_5, di					91,95	
5	**B2BTUR01_T_03:45**	**2006**	**6**	**25**	**26,50**	**17,26**		**17,30**	**16,30**	**15,80**	**15,60**	**16,20**	
0	B2BTUR01_T_04:00				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
0	B2BZAB01_T_04:00						st_1, di	11,58					
0	B1PROT01_T_04:00						st_2, di		36,85				
0	O3PRER01_T_04:00						st_3, di			59,12			
0	O2OLOM01_T_04:00						st_4, di				62,88		
0	O1CERV01_T_04:00						st_5, di					91,95	
0	**B2BTUR01_T_04:00**	**2006**	**6**	**25**	**26,30**	**17,41**		**17,30**	**16,50**	**16,50**	**15,90**	**16,20**	
0	B2BTUR01_T_05:00				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
0	B2BZAB01_T_05:00						st_1, di	11,58					
0	B1PROT01_T_05:00						st_2, di		36,85				
0	O3PRER01_T_05:00						st_3, di			59,12			
0	O2OLOM01_T_05:00						st_4, di				62,88		
0	O1CERV01_T_05:00						st_5, di					91,95	
0	**B2BTUR01_T_05:00**	**2006**	**6**	**25**	**24,70**	**17,52**		**17,30**	**17,20**	**17,30**	**16,30**	**17,20**	

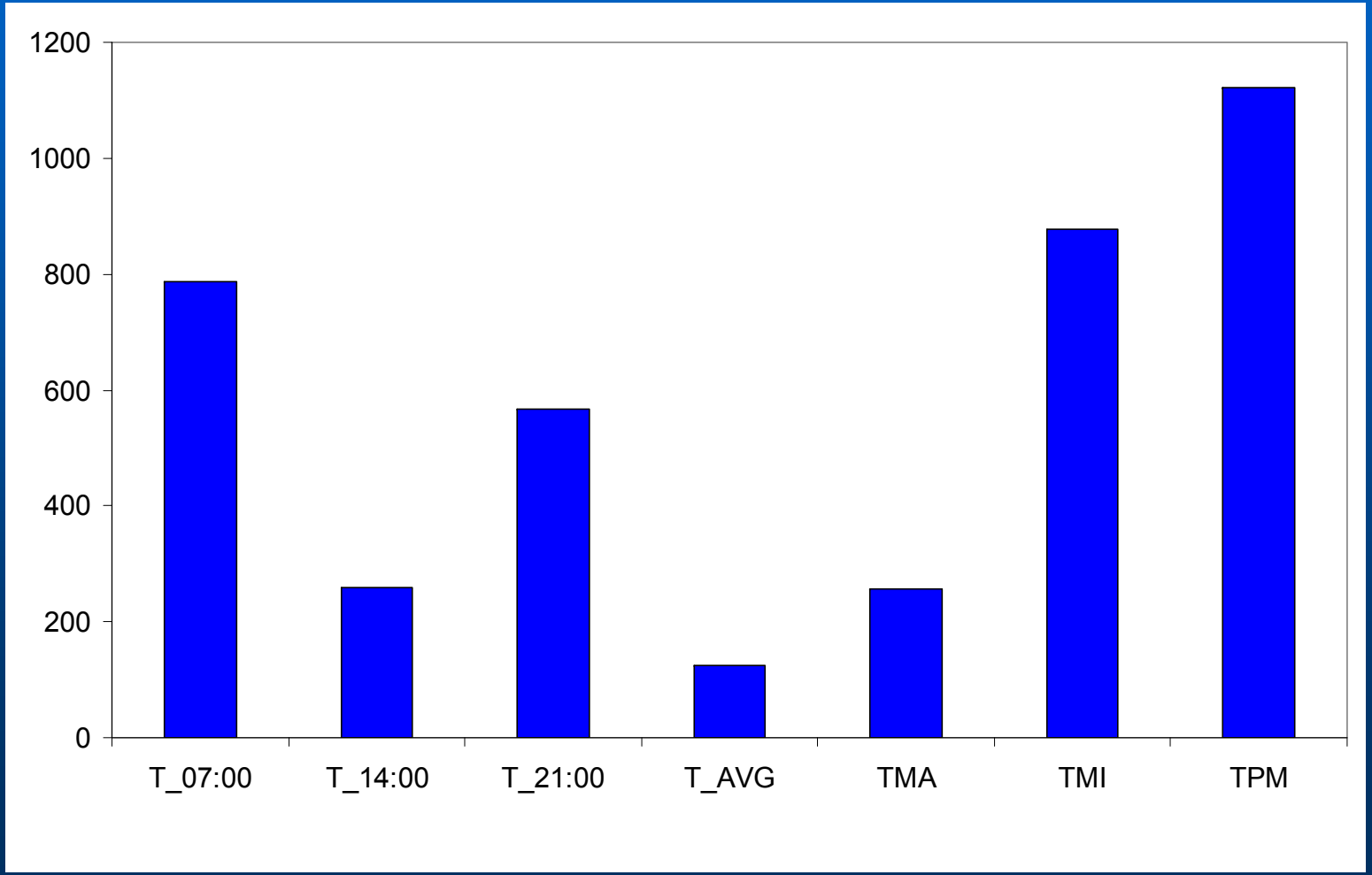
Quality control

- Run for period 1961-2007, daily data (measured values in observation hours)
- All stations (200 climatological stations, 800 precipitation stations)
- All meteorological elements (T, TMA, TMI, TPM, SRA, SCE, SNO, E, RV, H, F) – parameters set individually

- Historical records will follow now

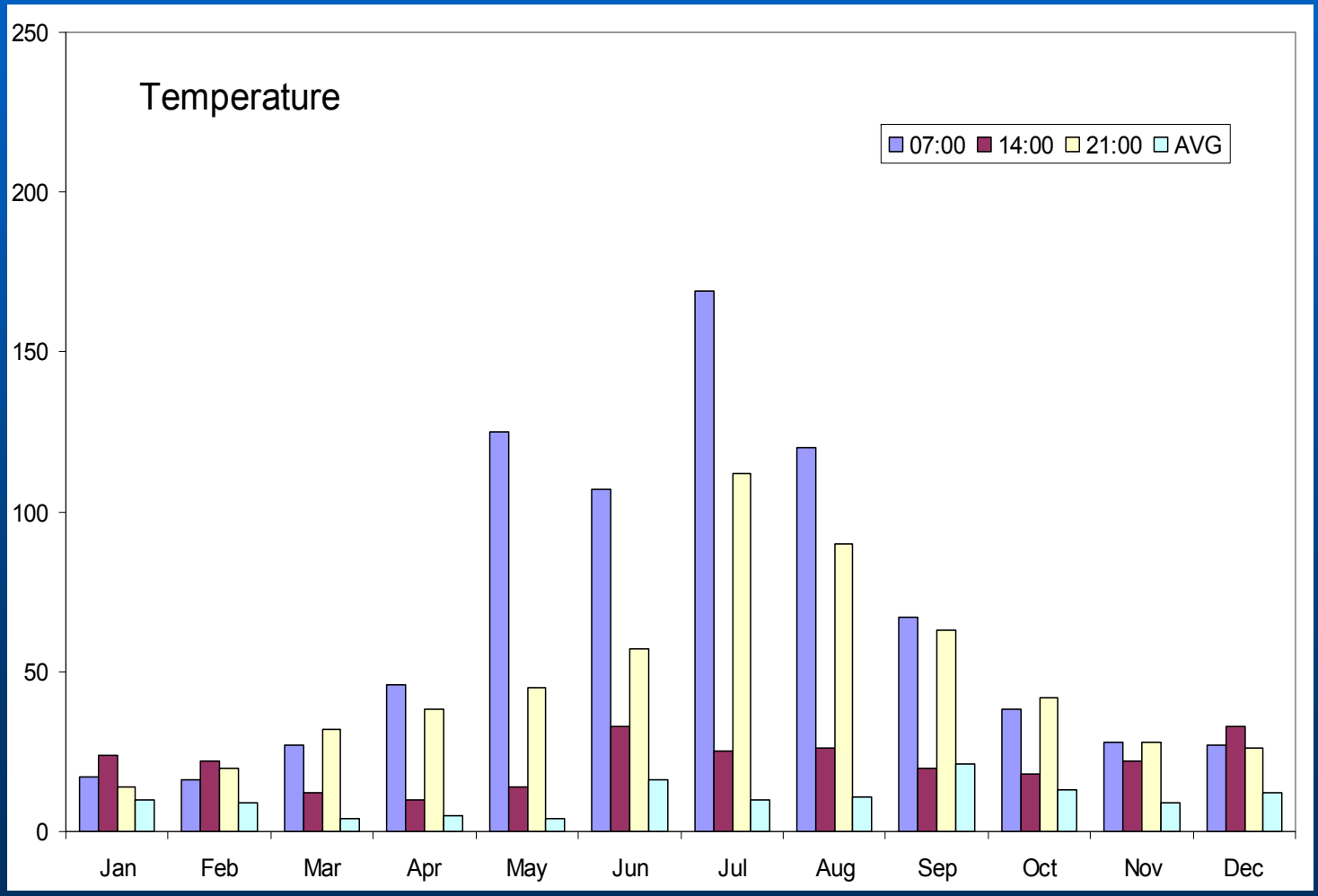
Air temperature, number of outliers 1961-2007, from 3.431.000 station-days

T – air temperature at obs. hour, TMA – daily maximum temp., TMI – daily min. temp., TPM – daily ground minimum temp.



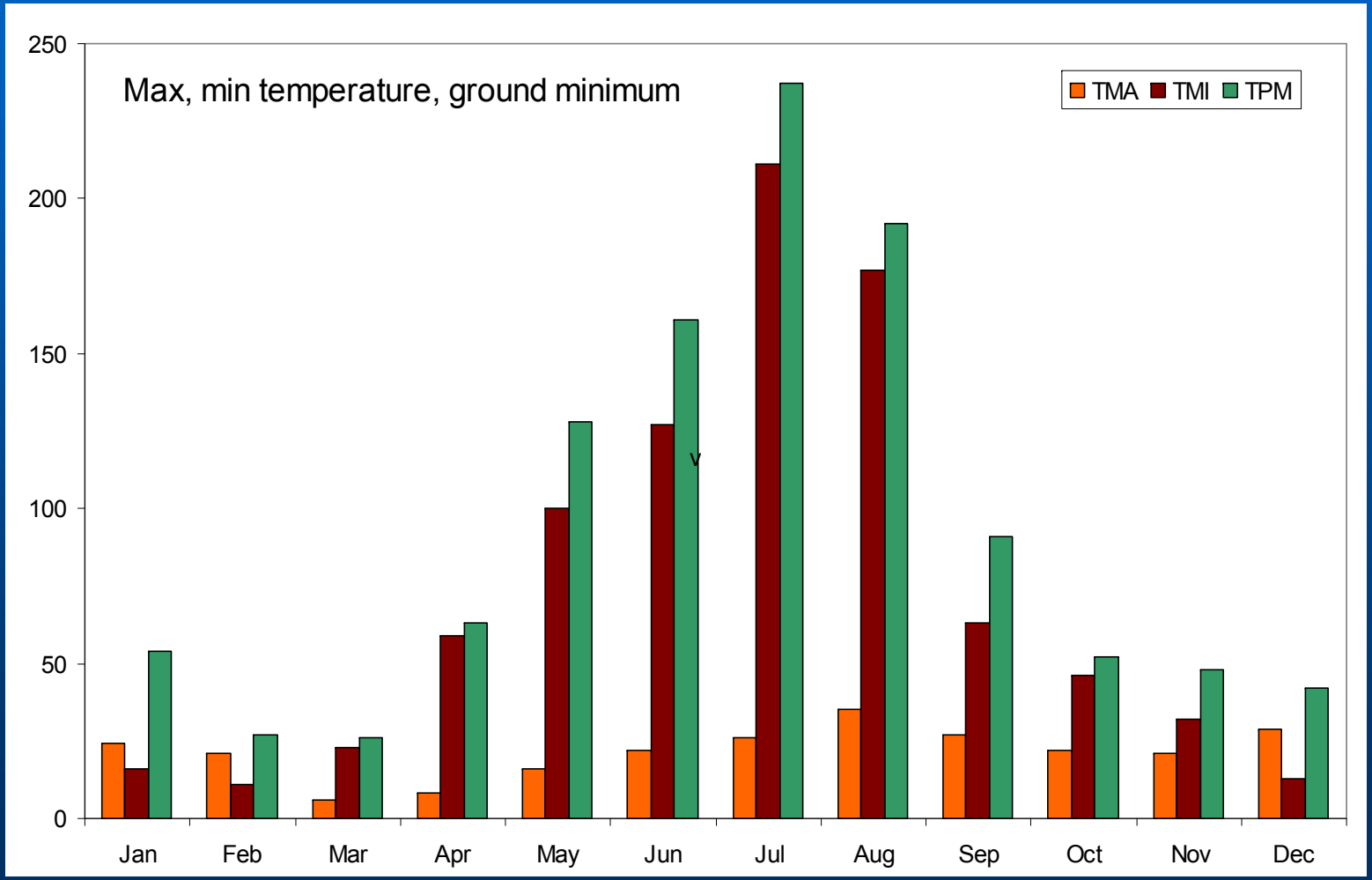
Air temperature, number of outliers 1961-2007, from 3.431.000 station-days

Air temperature at obs. hour, AVG – daily average temp.



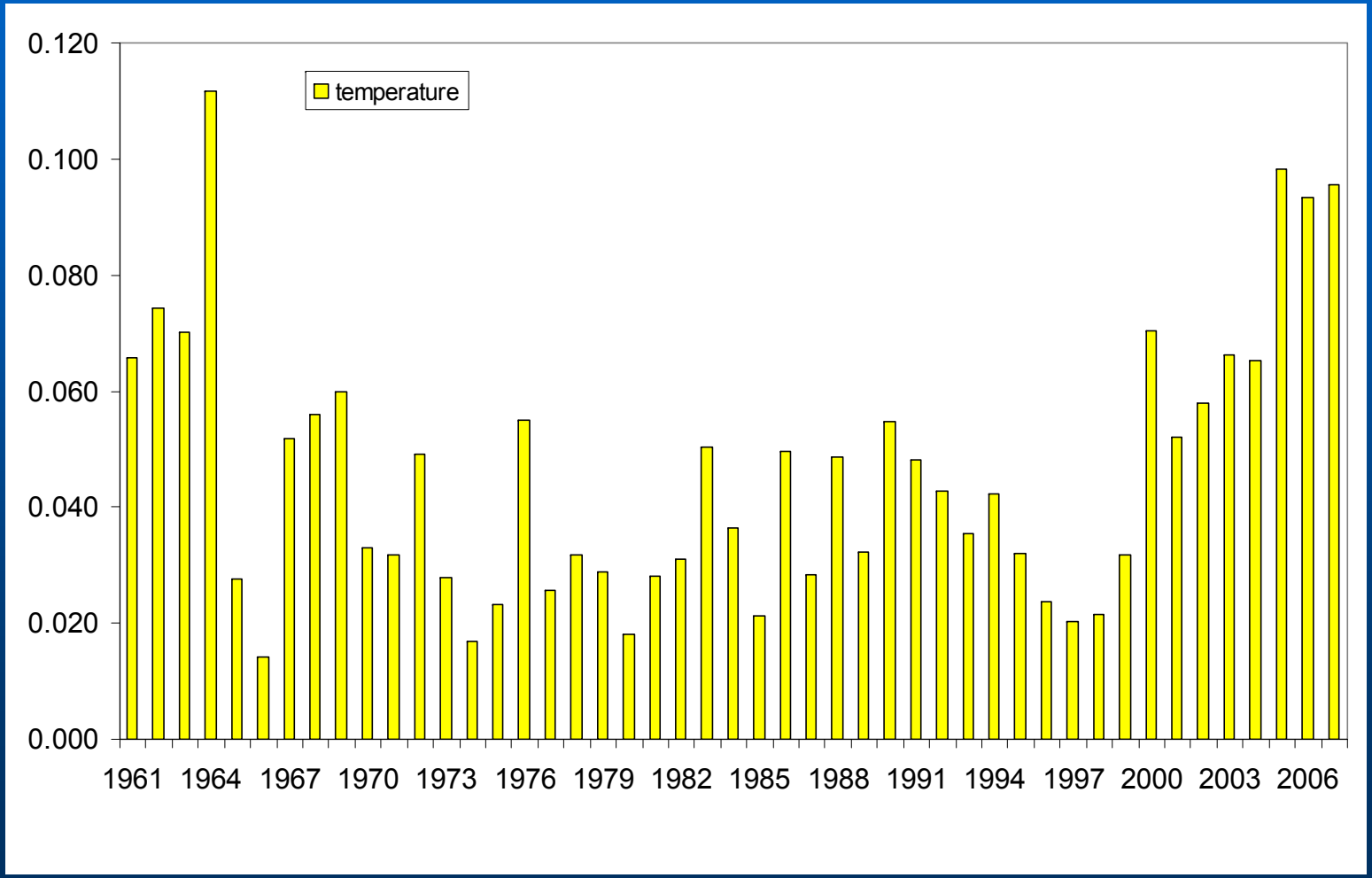
Air temperature, number of outliers 1961-2007, from 3.431.000 station-days

TMA – daily maximum temp., TMI – daily min. temp., TPM – daily ground minimum temp.

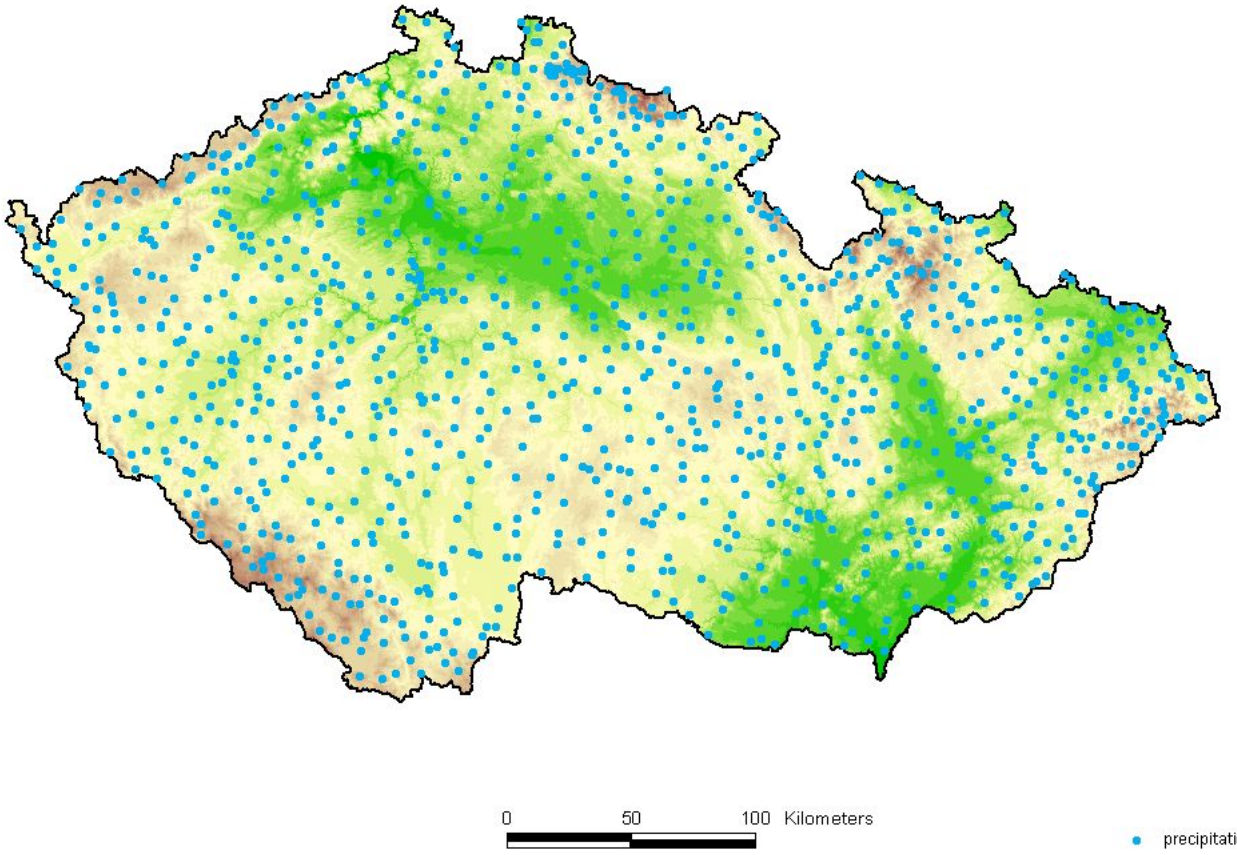


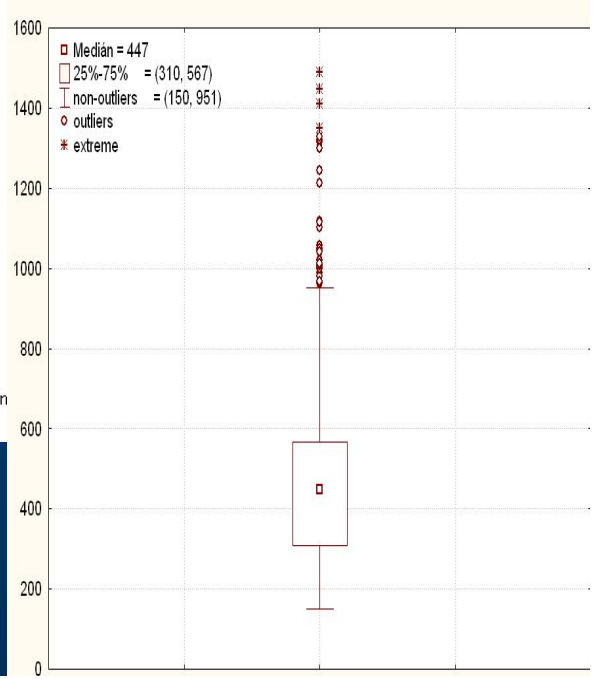
Air temperature, number of outliers 1961-2007,

Number of outliers per one station (all observation hours, AVG)



Spatial distribution of precipitation stations



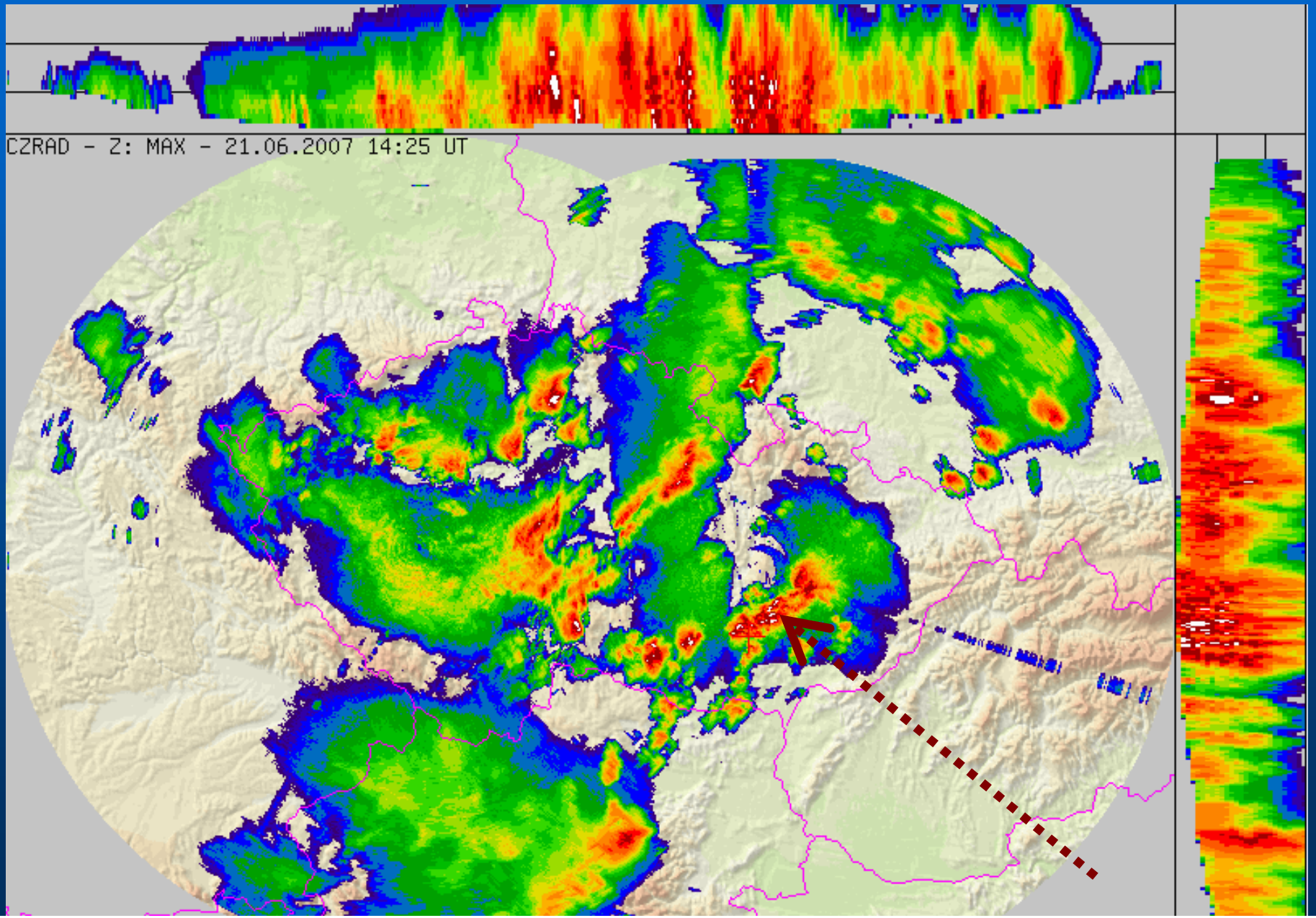


- period 1961-2007
- 600 stations
- mean minimum distance: 7.5 km

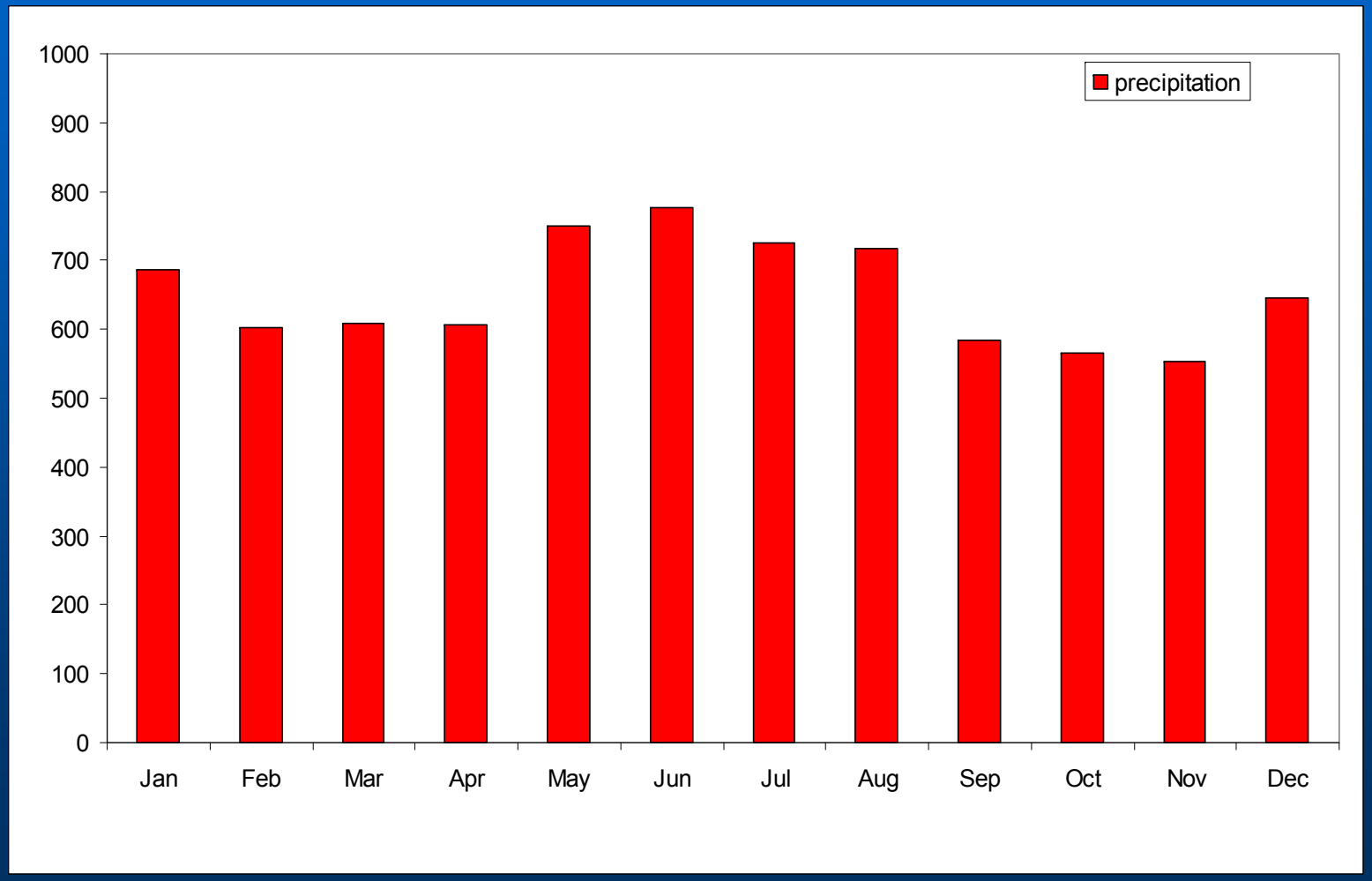
Problematic detections (heavy rainfall)

ID	YEAR	MONTH	DAY	ST_BASE	EXPECT_VAL	REMARK	ST_1	ST_2	ST_3	ST_4	ST_5	D
B2BTUR01_SRA3H_16:00				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
B2BZAB01_SRA3H_16:00						st_1, di	11,58					
B1PROT01_SRA3H_16:00						st_2, di		36,85				
O3PRER01_SRA3H_16:00						st_3, di			59,12			
O2OLOM01_SRA3H_16:00						st_4, di				62,88		
O1CERV01_SRA3H_16:00						st_5, di					91,95	
B2BTUR01_SRA3H_16:00	**2005**	**4**	**6**	**10,00**	**1,47**		**1,50**	**0,00**	**0,20**	**0,00**	**0,30**	
B2BTUR01_SRA3H_16:00	**2006**	**7**	**14**	**8,70**	**0,32**		**0,30**	**0,50**	**0,20**	**0,00**		
B2BTUR01_SRA3H_16:00	**2006**	**8**	**13**	**7,00**	**0,13**		**0,10**	**0,70**	**0,00**	**0,00**	**0,00**	
B2BTUR01_SRA3H_16:00	**2007**	**6**	**21**	**21,70**	**0,66**		**0,70**		**3,00**	**4,70**	**0,10**	
B2BTUR01_SRA3H_16:00	**2007**	**7**	**11**	**9,40**	**0,04**		**0,00**	**0,60**	**0,00**	**0,00**	**1,40**	
B2BTUR01_SRA3H_19:00				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
B2BZAB01_SRA3H_19:00						st_1, di	11,58					
B1PROT01_SRA3H_19:00						st_2, di		36,85				
O3PRER01_SRA3H_19:00						st_3, di			59,12			
O2OLOM01_SRA3H_19:00						st_4, di				62,88		
O1CERV01_SRA3H_19:00						st_5, di					91,95	
B2BTUR01_SRA3H_19:00	**2005**	**5**	**23**	**8,00**	**0,03**		**0,00**	**0,20**	**0,00**	**0,00**	**0,00**	
B2BTUR01_SRA3H_19:00	**2005**	**7**	**23**	**7,00**	**1,73**		**1,80**	**1,00**	**0,00**	**0,00**	**0,00**	
B2BTUR01_SRA3H_19:00	**2006**	**5**	**13**	**4,40**	**0,02**		**0,00**	**0,00**	**0,00**	**0,00**	**0,10**	
B2BTUR01_SRA3H_19:00	**2006**	**7**	**8**	**13,70**	**-0,04**		**0,00**	**0,00**	**0,00**	**0,00**	**0,00**	
B2BTUR01_SRA3H_19:00	**2006**	**8**	**7**	**5,90**	**0,25**		**0,20**	**0,90**	**0,90**	**0,00**	**0,00**	
B2BTUR01_SRA3H_19:00	**2007**	**1**	**1**	**3,40**	**0,69**		**0,70**	**0,60**	**0,30**	**0,00**	**1,10**	
B2BTUR01_SRA3H_19:00	**2007**	**6**	**14**	**9,00**	**0,03**		**0,00**	**0,00**	**0,30**	**0,00**	**0,00**	
B2BTUR01_SRA3H_22:00				241,00		Altitude	235,00	670,00	203,00	210,00	749,00	
B2BZAB01_SRA3H_22:00						st_1, di	11,58					
B1PROT01_SRA3H_22:00						st_2, di		36,85				
O3PRER01_SRA3H_22:00						st_3, di			59,12			
O2OLOM01_SRA3H_22:00						st_4, di				62,88		
O1CERV01_SRA3H_22:00						st_5, di					91,95	
B2BTUR01_SRA3H_22:00	**2005**	**4**	**25**	**1,90**	**0,39**		**0,40**	**0,10**	**0,20**	**0,00**	**0,10**	
B2BTUR01_SRA3H_22:00	**2005**	**6**	**25**	**20,00**	**7,68**		**7,70**		**8,00**	**8,60**	**8,00**	

Problematic detections (heavy rainfall), Radar information

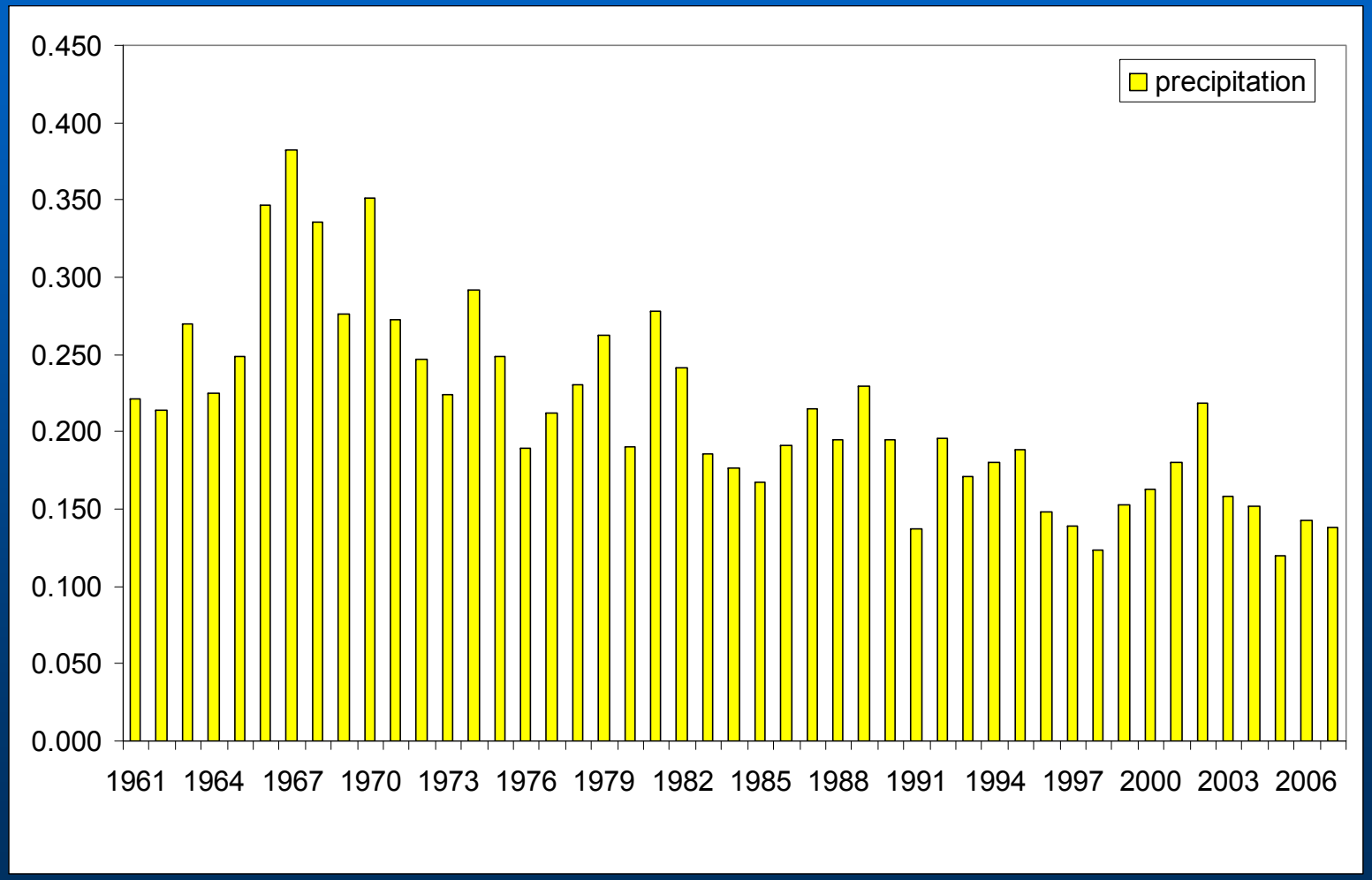


Precipitation, number of outliers 1961-2007, from 13.724.000 station-days

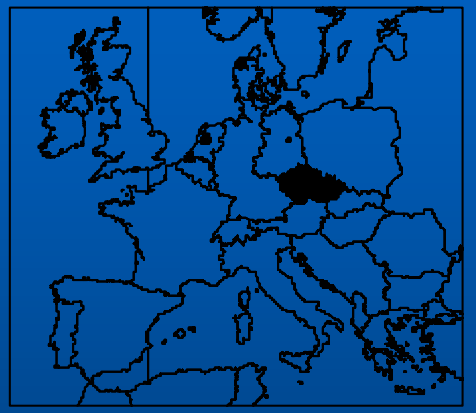


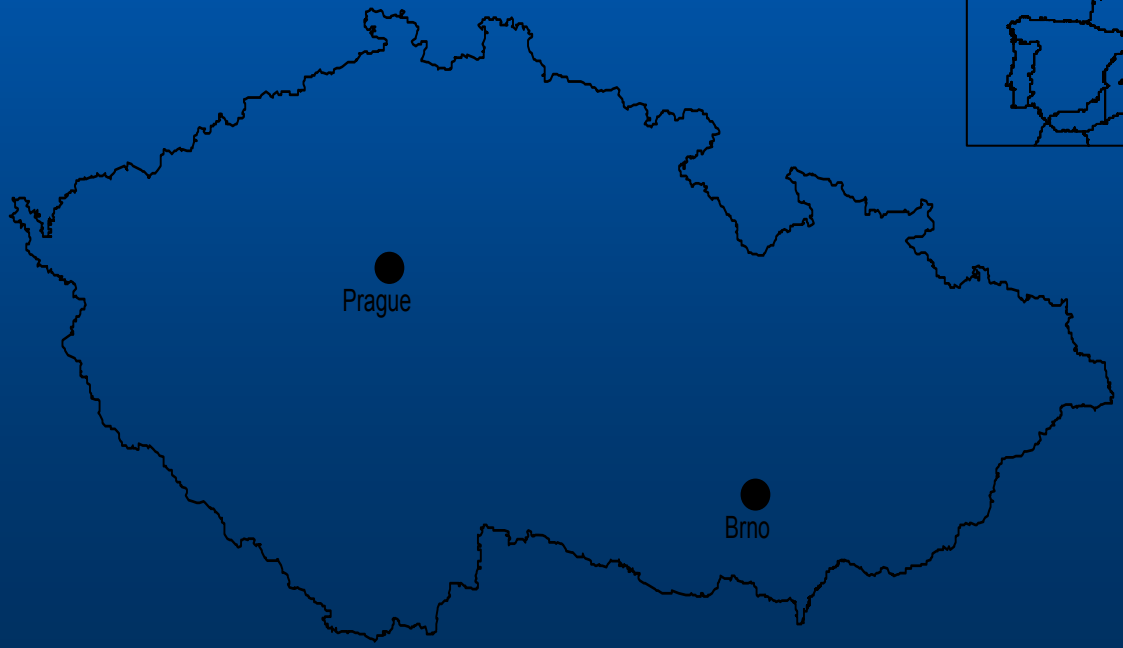
Precipitation, number of outliers 1961-2007,

Number of outliers per one station



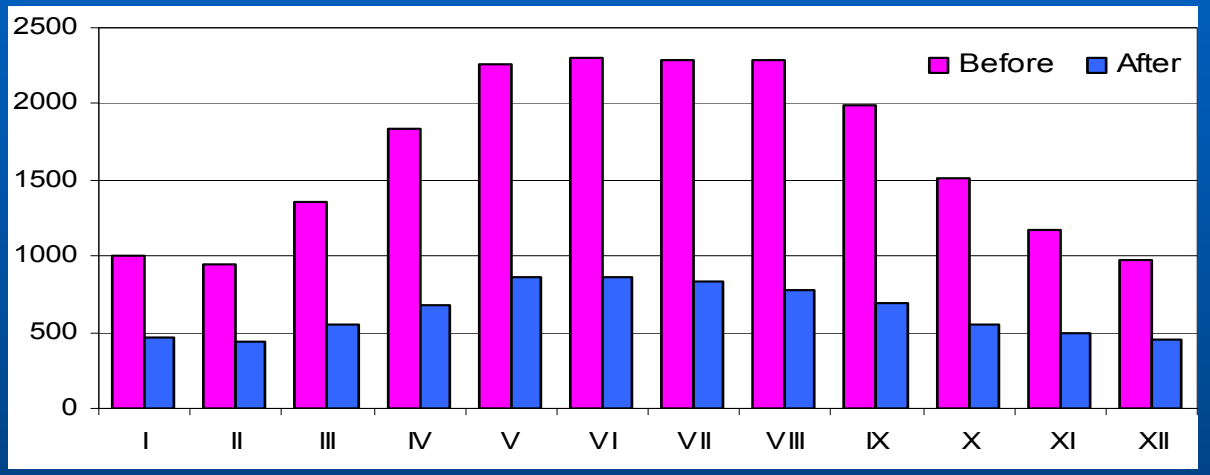
Homogenization
of the series in the Czech Republic

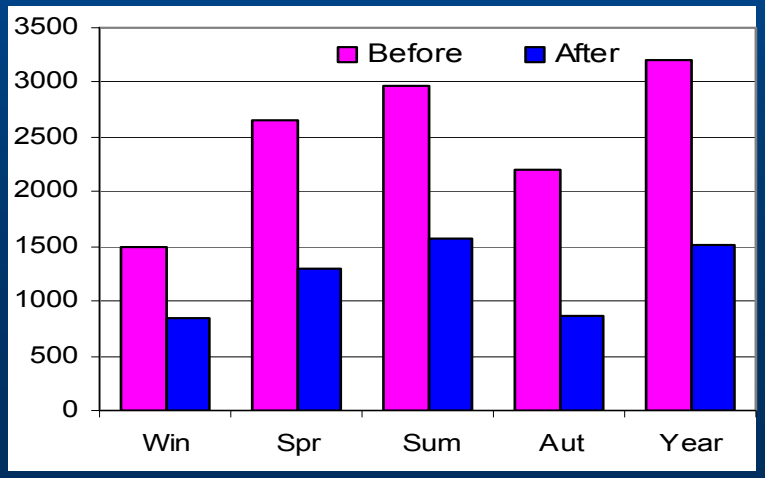




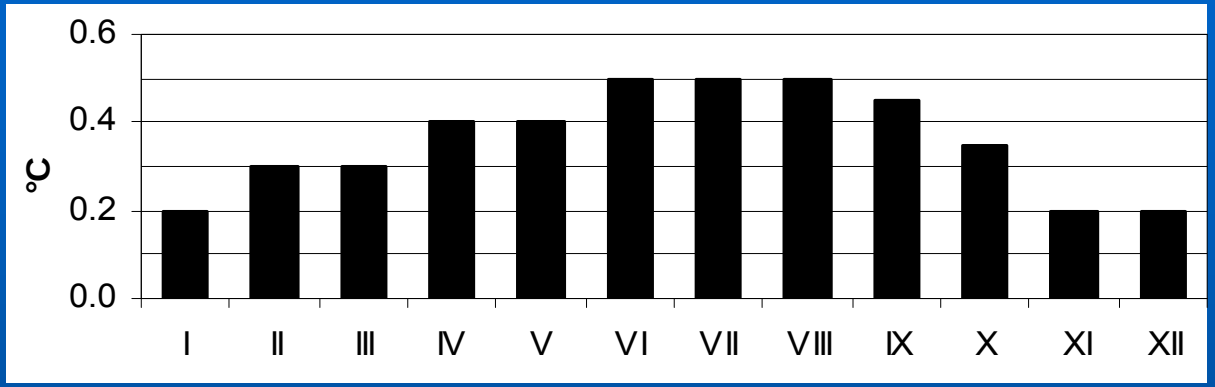
Example: CZ, air temperature (200 stations, 1848-2000)

Number of significant inhomogeneities before and after homogenization (p=0.05)

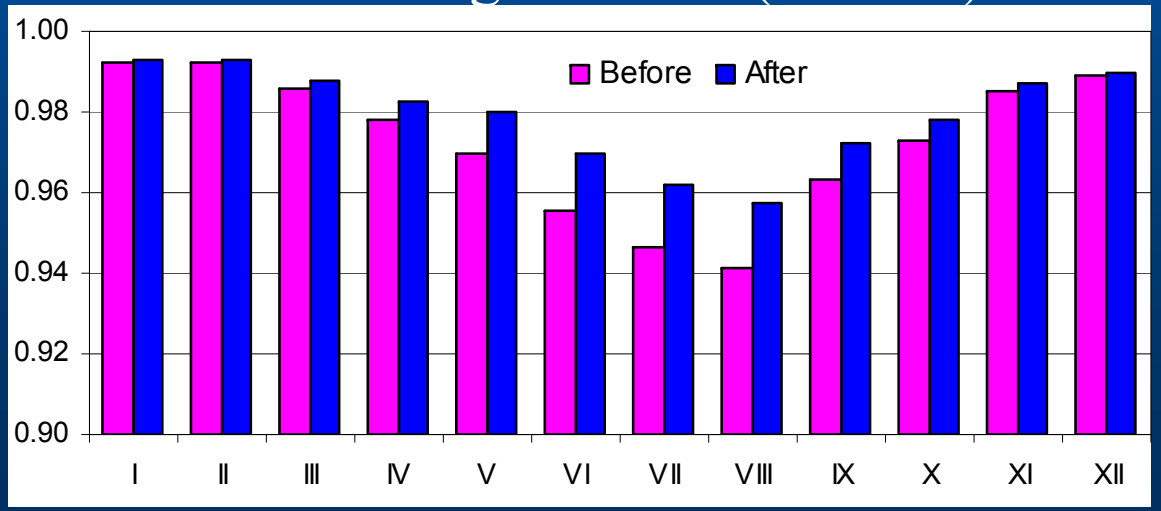




Amount of adjustments for homogenised series (absolute values) - median

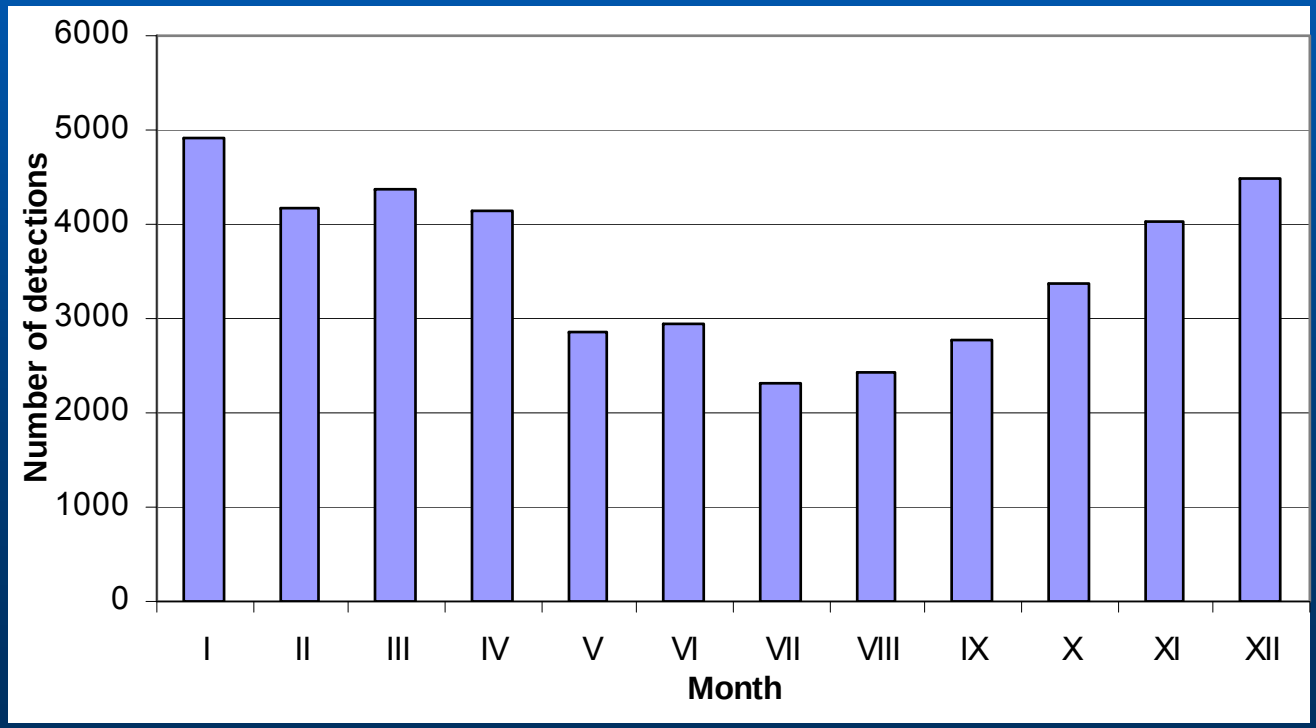


Correlation coefficients between candidate and reference series before and after homogenization (median)



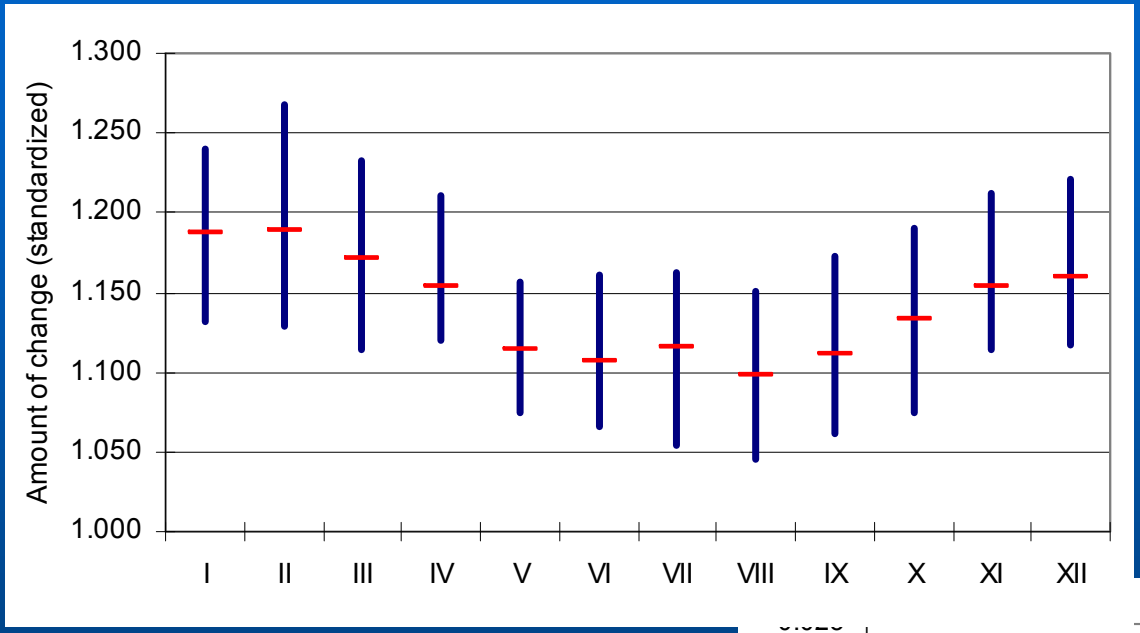
Example: CZ, precipitation (800 stations, 1961-2000)

- 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), precipitation

(reference series calculation based on correlations)

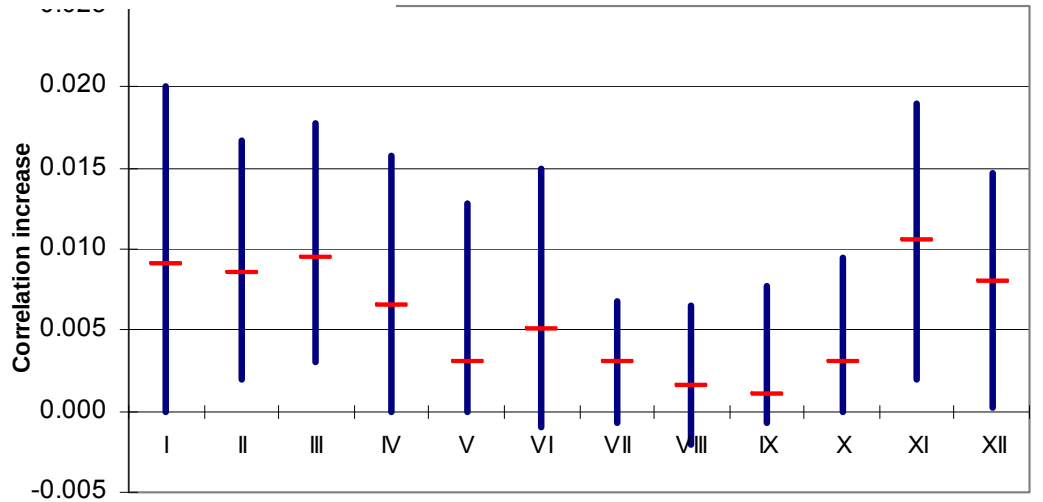


Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)

Correlation improvement



Inhomogeneities in summer versus in winter, Air **temperature**

- Change of measuring conditions at the station (relocation etc.) is manifested in the series mainly in **summer**
- in winter: active surface role is diminished, prevailng circulation factors, in summer: active surface role increases, prevailing radiation factors

Inhomogeneities in summer versus in winter, **Precipitation**

- Change of measuring conditions at the station (relocation etc.) is manifested in the series mainly in **winter**
- in winter: errors of measurement (solid precipitation - wind, …)

Homogenization
Final remarks, recommendations 1/2

- **data quality control before homogenization is of very importance** (if it is not part of it)
- **Using series of observation hours** (complementarily to daily AVG) **is highly recommended** (different manifestation of breaks)
- **be aware of annual cycle of inhomogeneities, adjustments, …**
- **to know behavior of spatial correlations** (of element being processed) **to be able to create reference series of sufficient quality …**

Homogenization
Final remarks, recommendations 2/2

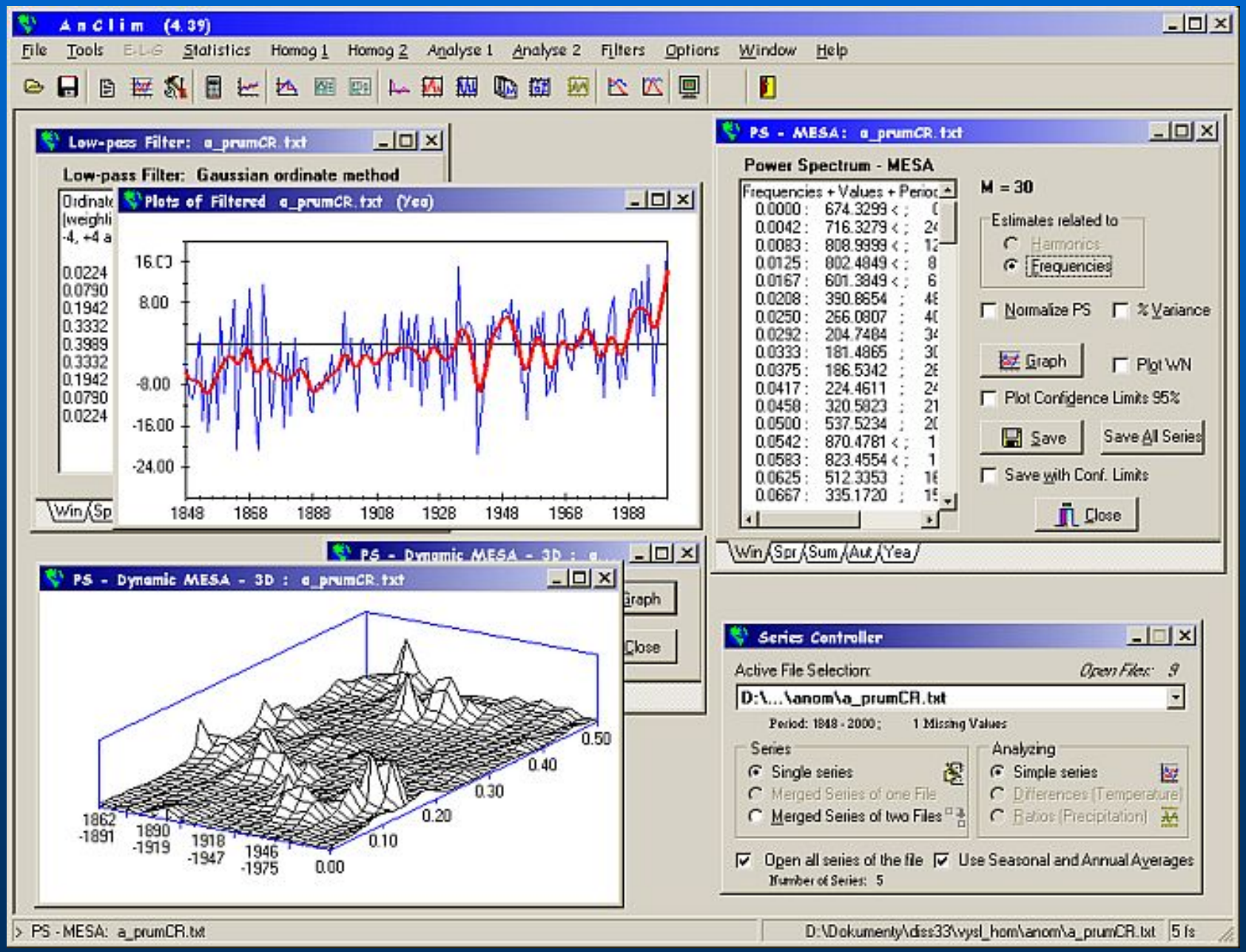
- **Because of Noise in the time series it makes sense:**
- **- „Ensemble“ approach to homogenization (combining information from different statistical tests, time frames, overlapping periods, reference series, meteorological elements, …)**
- **- more information for inhomogeneities assessment – higher quality of homogenization in case metadata are incomplete**

Software used for data processing

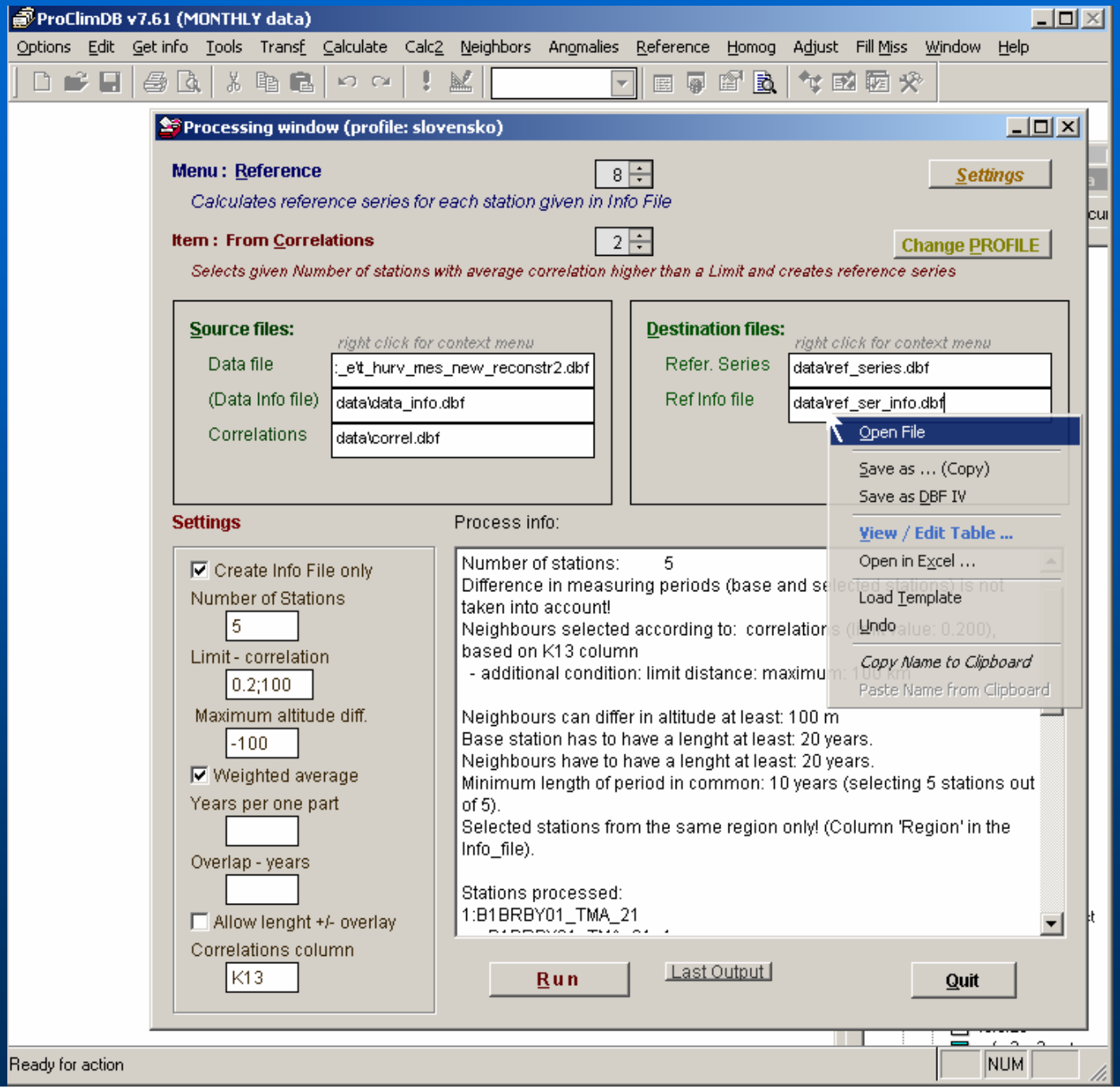
- **LoadData - application for downloading data from central database (e.g. Oracle)**
- **ProClimDB software for processing whole dataset (finding outliers, combining series, creating reference series, preparing data for homogeneity testing, extreme value analysis, RCM outputs validation, correction, …)**
- **AnClim software for homogeneity testing**

http://www.climahom.eu

AnClim software

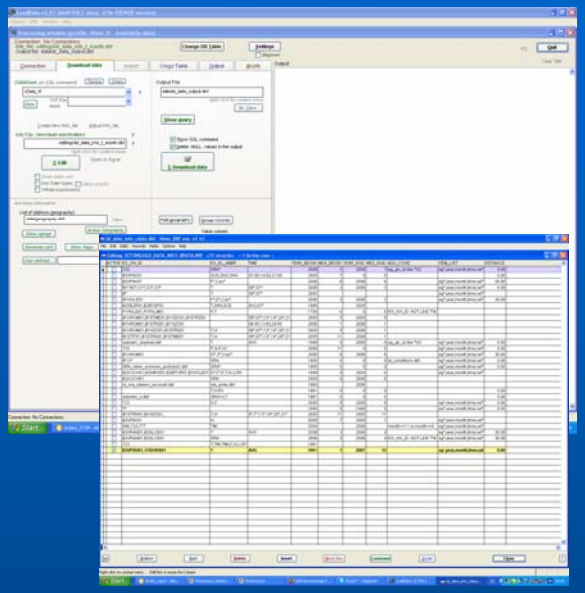


ProClimDB software



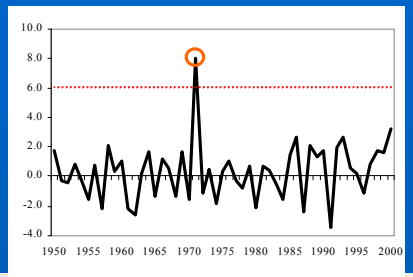
Download data from database (e.g. Oracle)

(LoadData)



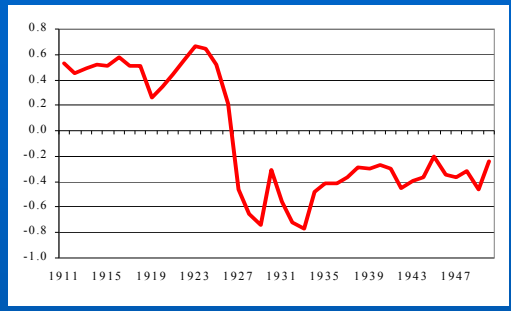
Quality control

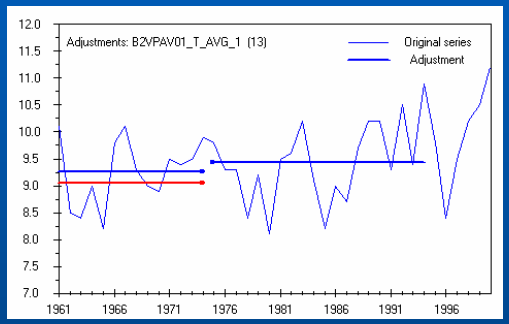
(ProClimDB)



Homogenization

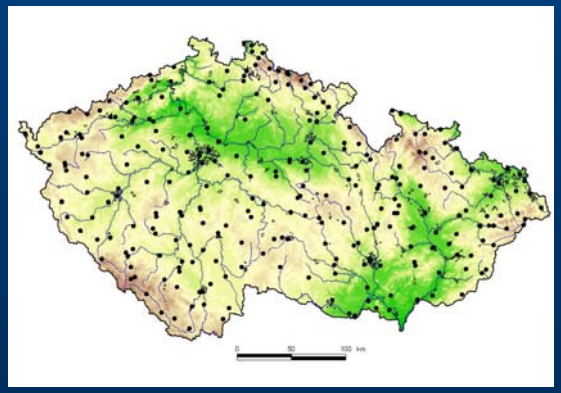
(ProClimDB/AnClim)

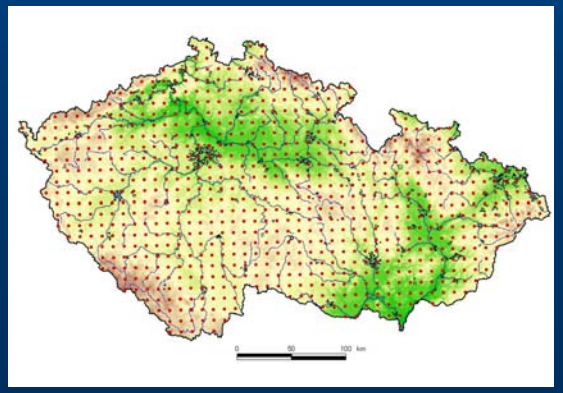




„Technical“ series and grid points calculation

(ProClimDB)





Statistical analysis

SPI, …

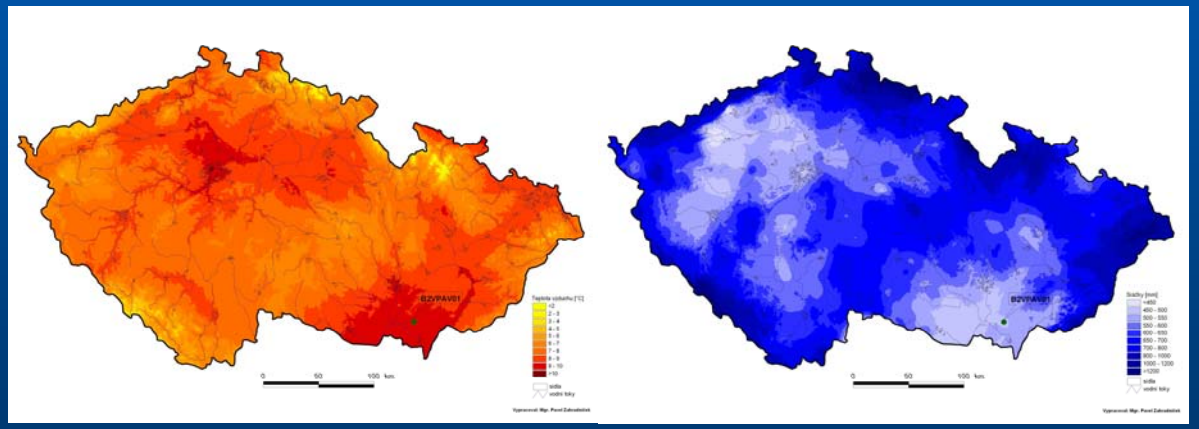
Validation of RCM outputs

Extreme value analysis

Correction of RCM outputs

Spatial analysis

(connection ProClimDB - ArcView)



Further tools:

(connection ProClimDB - R)