

Kontrola kvality dat

Homogenizace časových řad

Petr Štěpánek

Český hydrometeorologický ústav, p. Brno

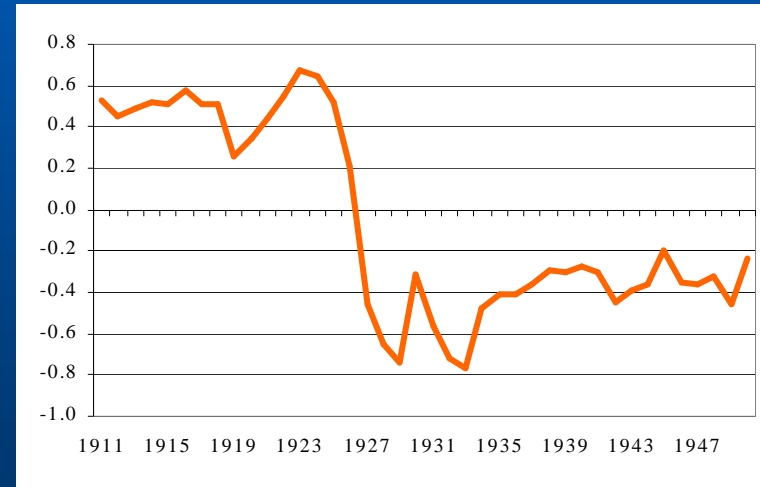
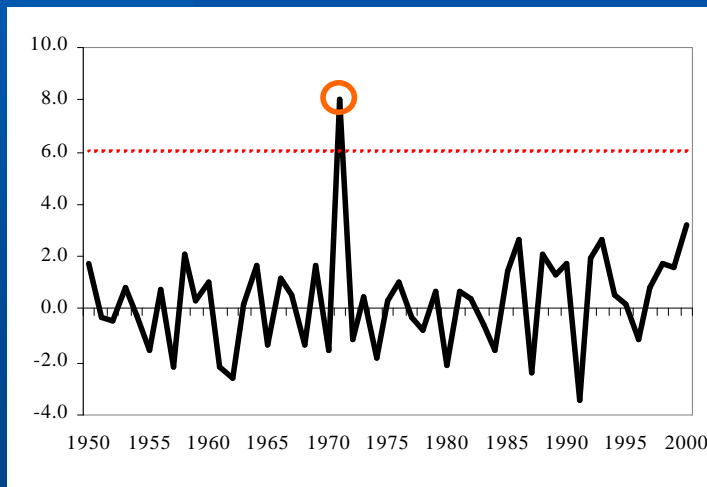
Klimatologické studie

- naměření a sběr dat



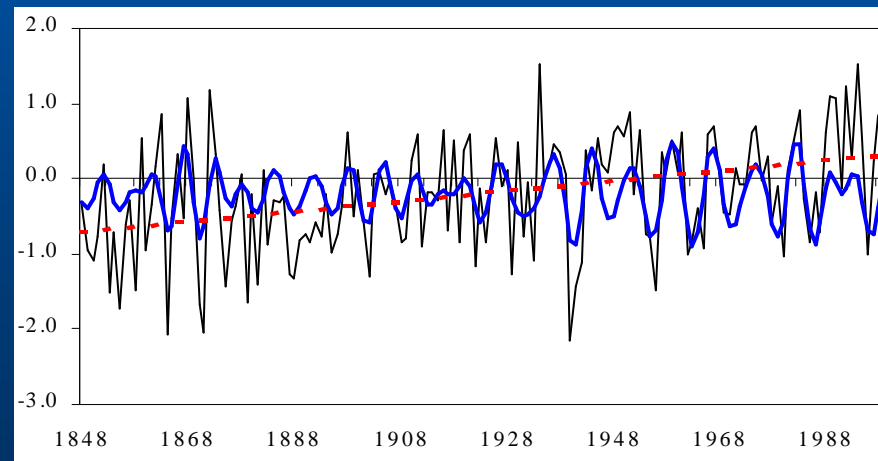
Klimatologické studie

- pořízení dat
- kontrola kvality dat a homogenizace



Klimatologické studie

- pořízení dat
- homogenizace
- analýza dat



Zpracování dat

Měsíční, sezonní a
roční průměry

Kontrola kvality -
vychýlené hodnoty

Mezikvart. odchylka

Porovnání se sousedy

Navazání řad
sousedních stanic

Testování homogenity

Alexanderssonův test

Bivariační test

t-test

Mann-Whitney-Pettit

Referenční řady

pomocí korelací

pomocí vzdáleností

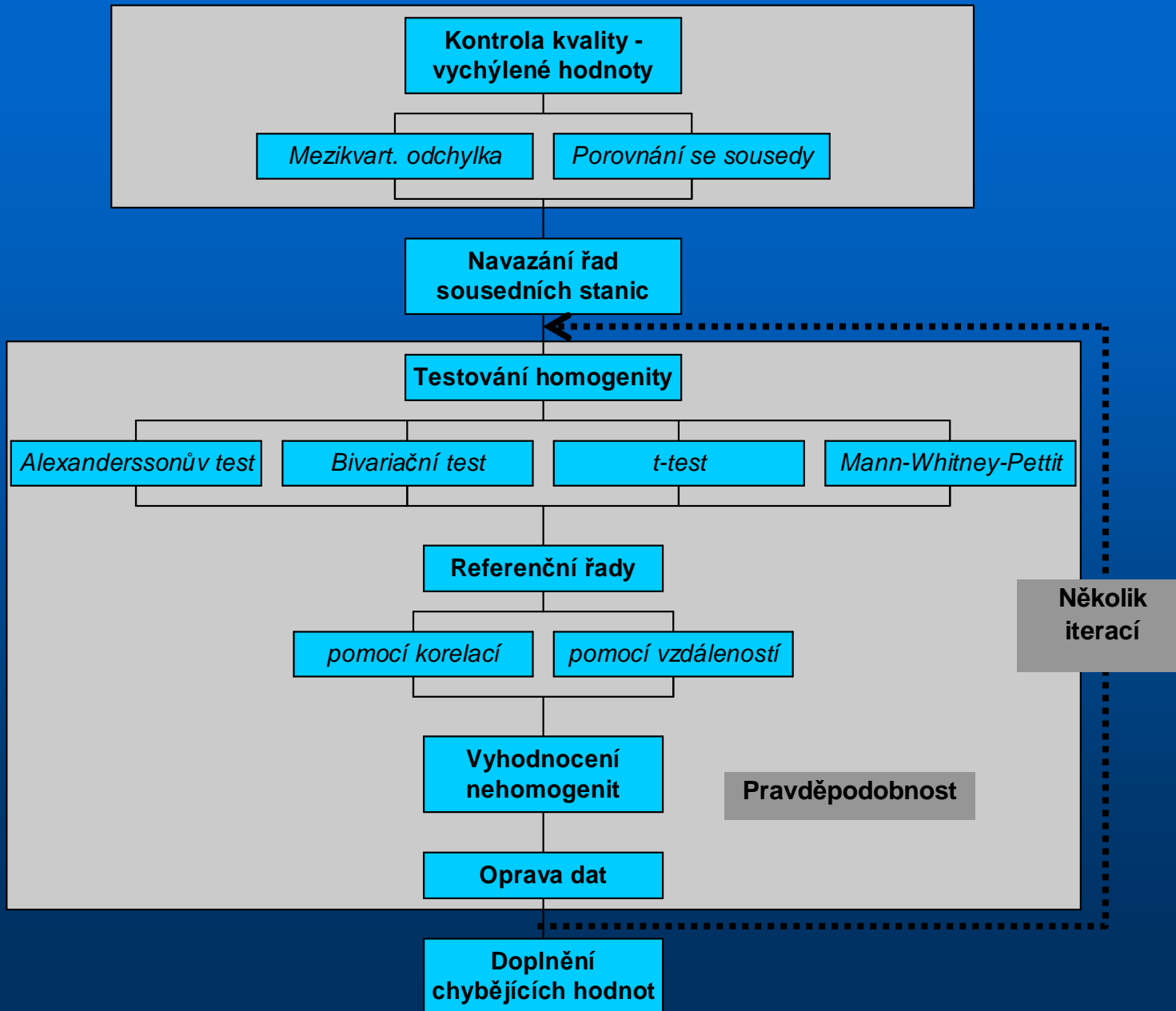
Vyhodnocení
nehomogenit

Pravděpodobnost

Oprava dat

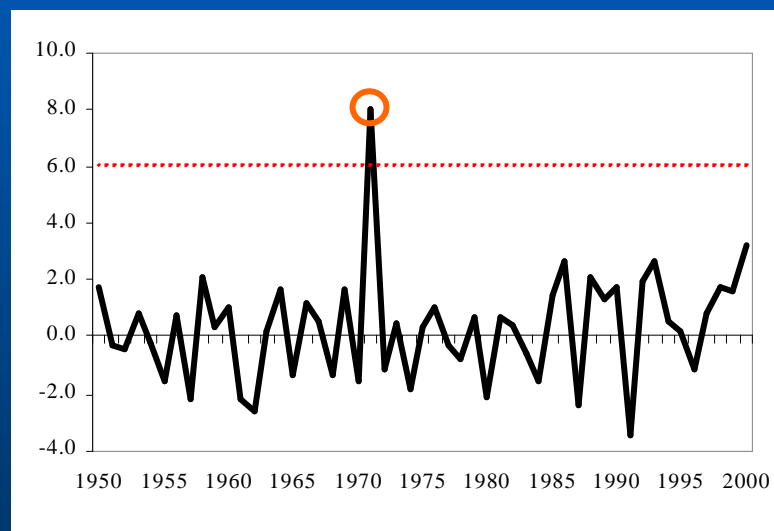
Doplnění
chybějících hodnot

Několik
iterací



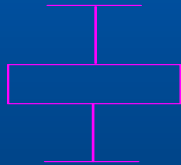
Kontrola kvality dat

Nalezení vychýlených hodnot



Kontrola kvality dat

Nalezení vychýlených hodnot

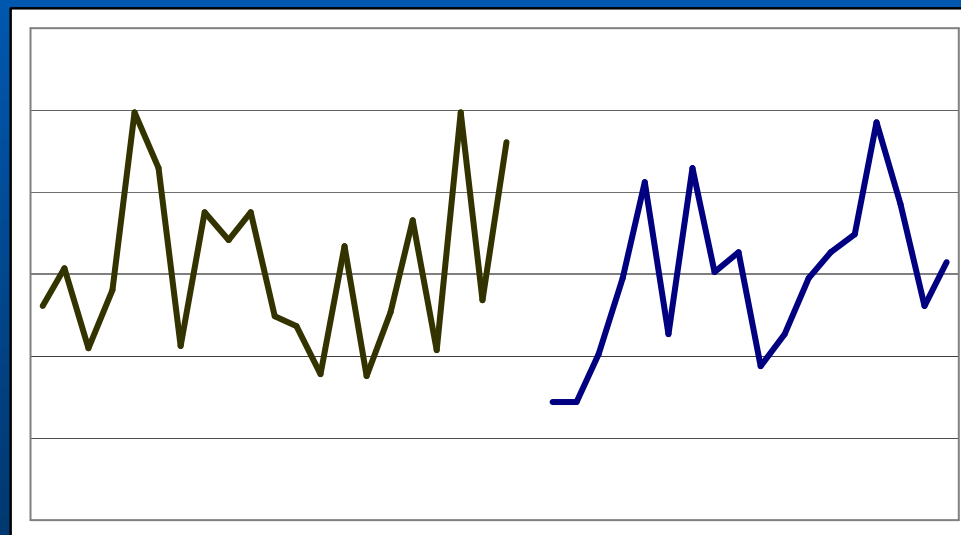
- **1. Limitní hodnoty získané z mezikvartilové odchylky**
 - řada poměrů (logaritmů) mezi testovanou a referenční řadou
 - referenční řada vytvořená jako průměr 5 nejvíce korelovaných stanic, max. vzdálenost 35 km
 - hranice: koeficient (násobek) = 3.0
- 
- **2. Porovnání hodnot s hodnotami sousedních stanic**
 - porovnání s ... nejbližšími stanicemi
 - řady standardizovaných poměrů (logaritmy poměrů)
 - počet případů překračujících 95% meze spolehlivosti
 - výpočet očekávané hodnoty ze standardizovaných řad

Příklad:

Porovnání hodnot s hodnotami sousedních stanic

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	

Navázání měření sousedních stanic



Navázání měření sousedních stanic (příklad)

- Výběr sousedů do **5 km**
- Mezera mezi dvěma řadami: maximálně **4 roky**
- Výsledná řada: dlouhá minimálně **30 roků**
- Záznam do metadat, zohlednění při homogenizace

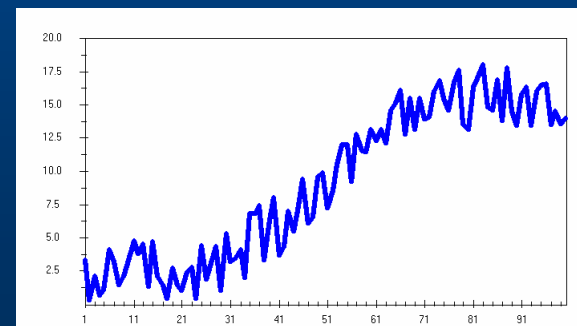
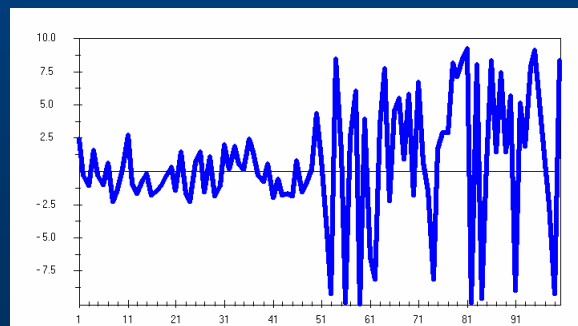
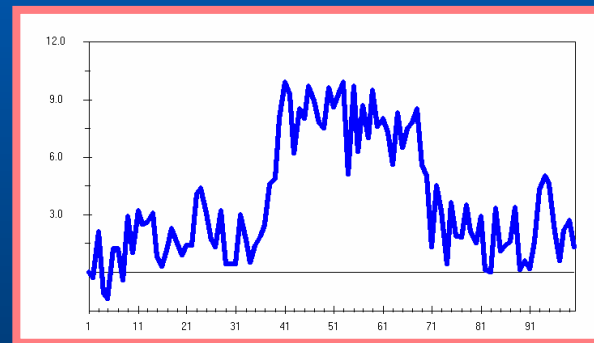
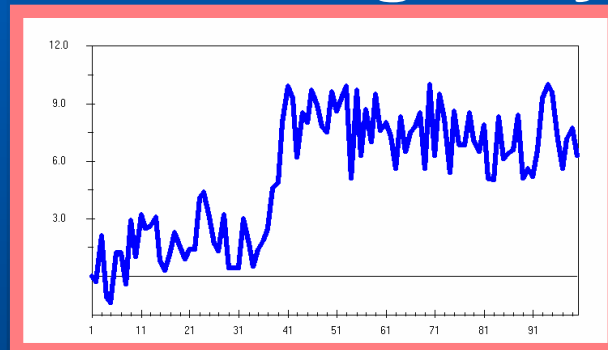
Homogenizace časových řad

- Quality control
- Homogenization
- Data Analysis

Homogenizace

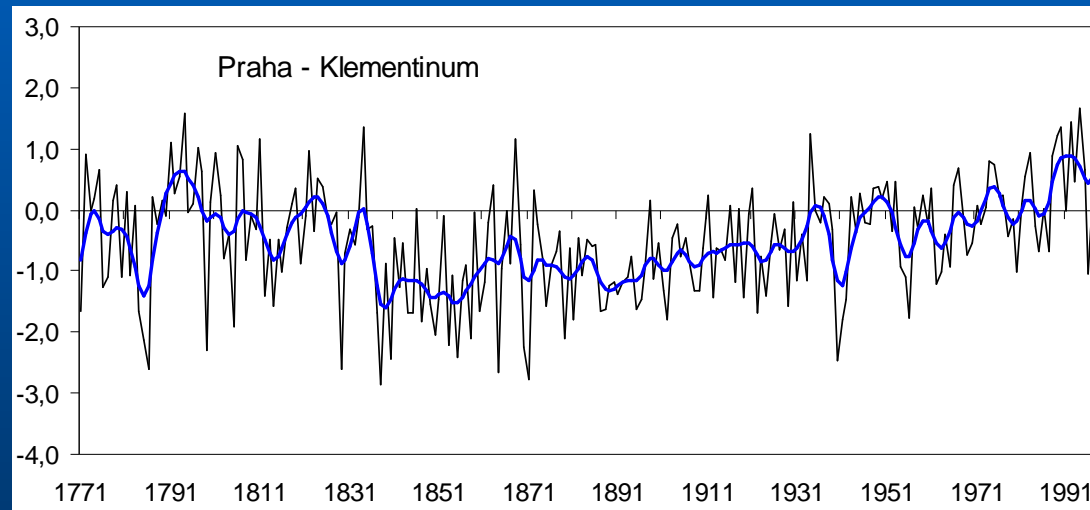
- změna podmínek na stanici

→ nehomogenity



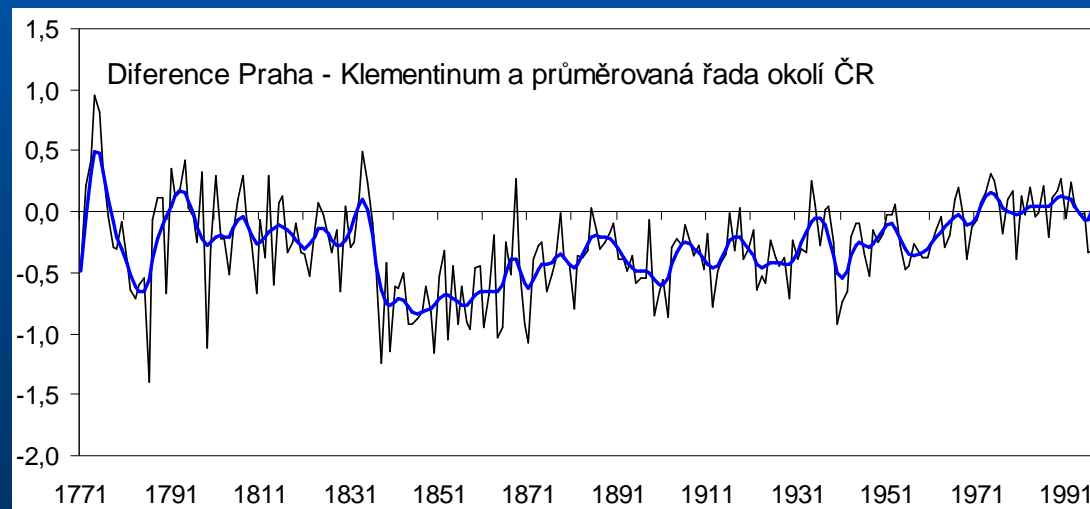
Detekce nehomogenit

- Testy absolutní homogenity



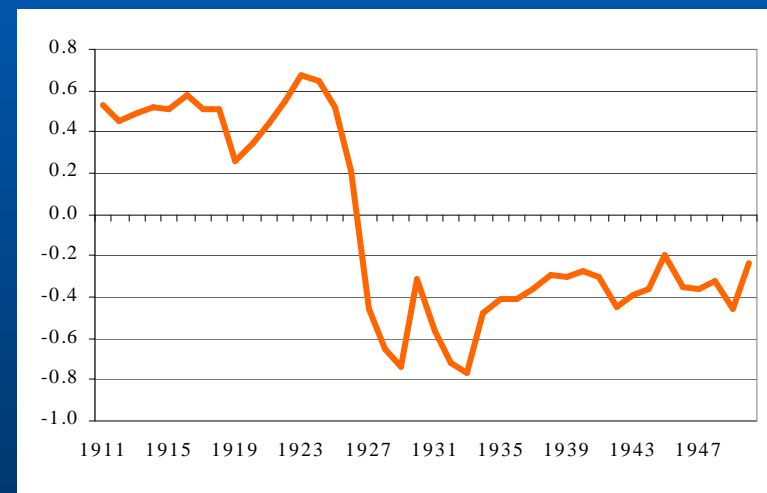
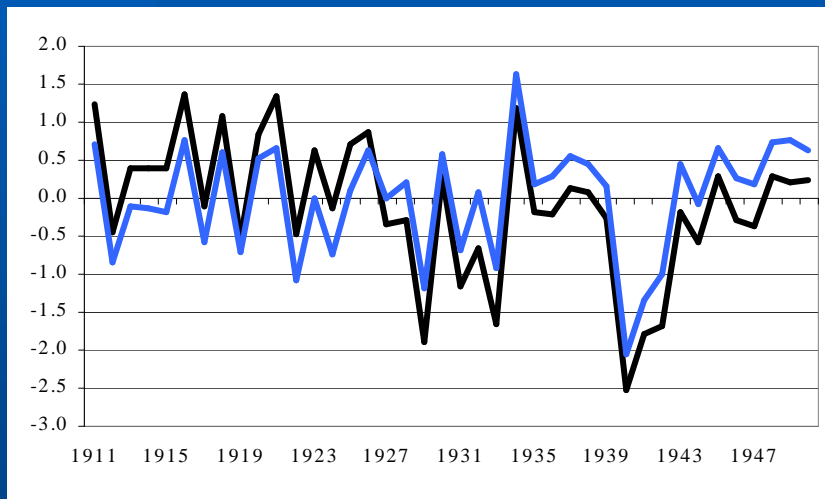
Detekce nehomogenit

- Testy absolutní homogenity
- Testy relativní homogenity



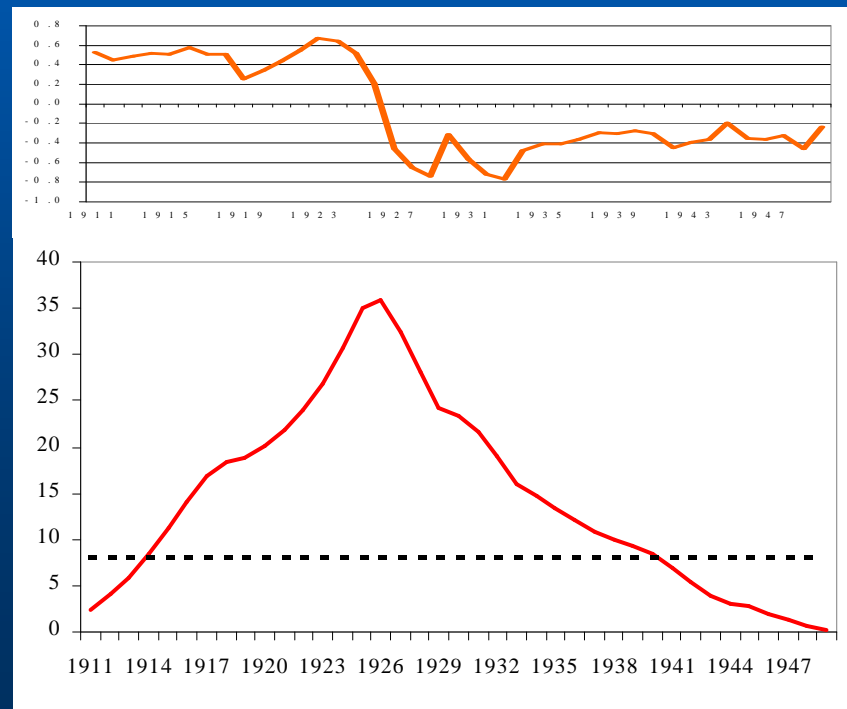
Homogenizace – postup (testování relativní homogenity)

- Tvorba referenčních řad



Homogenizace

- Tvorba referenčních řad
- Testování homogenity



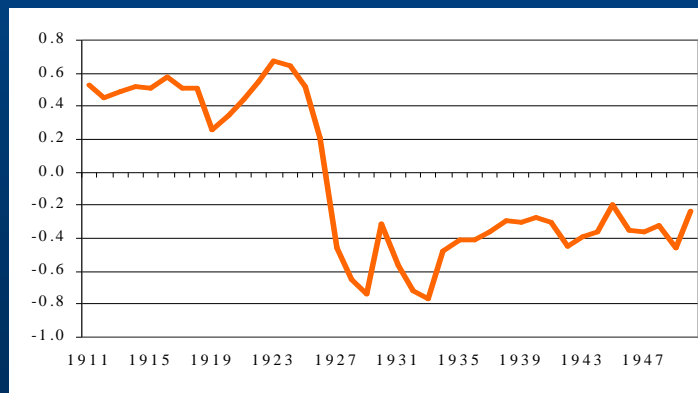
Homogenizace

- Tvorba referenčních řad
- Testování homogenity
- Posouzení nehomogenit v řadách

- Metadata



- fyzikálně zdůvodnitelné
("nezpochybnitelné") nehomogenity



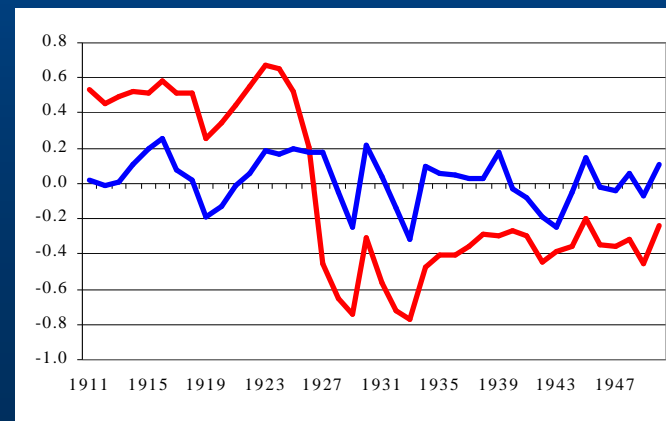
Homogenizace

- Tvorba referenčních řad
- Testování homogenity
- Posouzení homogenity řad

- Metadata

- fyzikálně zdůvodnitelné
("nezpochybnitelné") nehomogenity

- Oprava řad



Možné přístupy k homogenizaci

- nehomogenizovat vůbec
- použít statistické testy a věřit, že jejich výsledky jsou „svaté“
- ?

Spolehlivost detekce nehomogenit

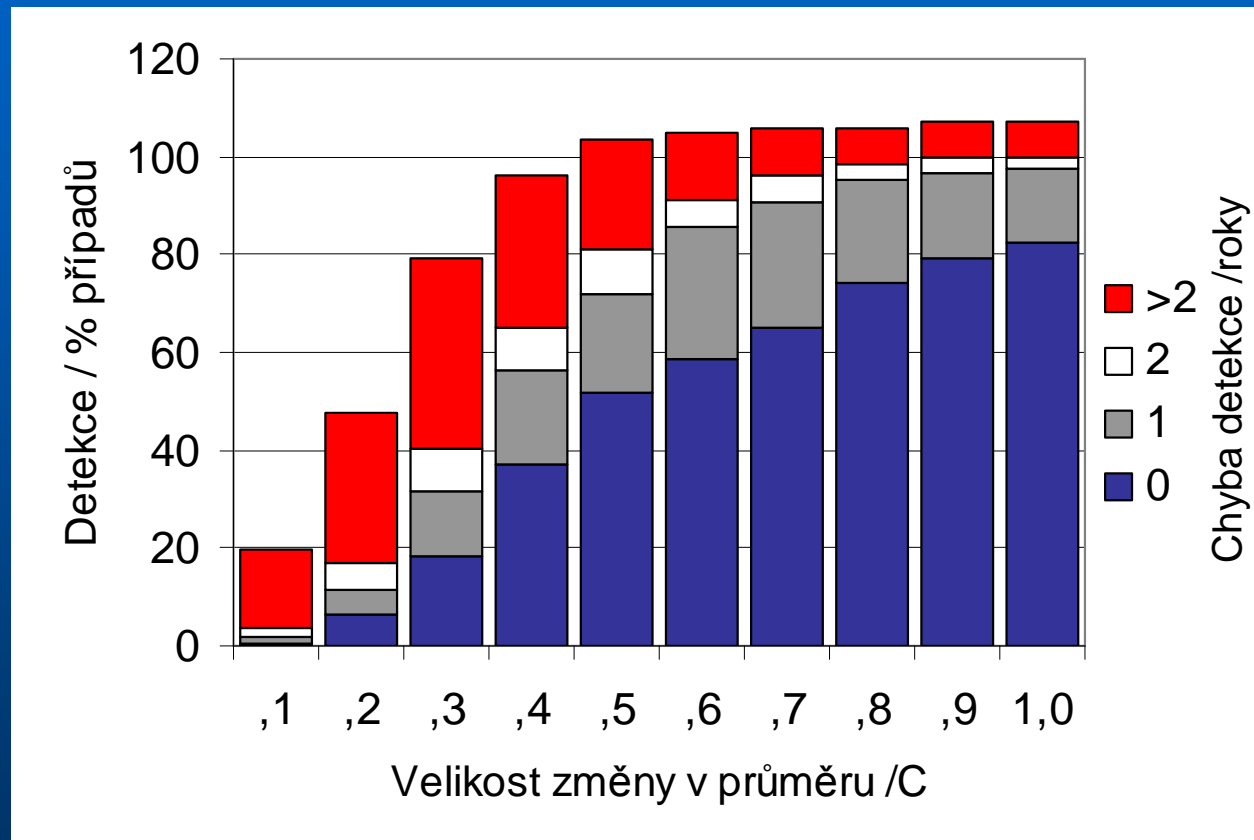


Inhomogeneities Detecting by SNHT ($p=0.05$, 950 series)

- generated series of random numbers (properties of air temperature series for year, summer and winter, CZ)
- introduced steps with various amount of change in level
- various position of the steps
- various lengths of the series

Schopnost detekce nehomogenit

SNHT ($p=0.05$, 950 series)



Změna:

- velikost nehomogenity
- délka řad
- poloha v řadě

Problémy při vyhodnocování nehomogenit

- většinou jsou metadata nekompletní

→ jsme tedy závislí pouze na výsledcích statistických testů

- ale správná detekce nehomogenit pomocí testů je často problematická

(splnění všech podmínek aplikace testů, problémy v řadách, ...)

Navržené řešení

Statistické zpracování velkého množství výsledků testování pro každou testovanou řadu (ensemble)

→ **pravděpodobnost dané nehomogenity - poměr počtu detekovaných nehomogenit na počtu všech teoreticky možných detekcí (pro každý rok, skupinu roků, celé období)**

Jak zvýšit spolehlivost testování homogenity

(bez metadat)

- měsíční, sezónní, roční průměry
- různé referenční řady
- různé testy homogenity
- 40-ti leté úseky
- několik kroků - iterací

Zpracování dat

Měsíční, sezonní a roční průměry

Kontrola kvality - vychýlené hodnoty

Mezikvart. odchylka

Porovnání se sousedy

Navazání řad sousedních stanic

Testování homogenity

Alexanderssonův test

Bivariační test

t-test

Mann-Whitney-Pettit

Referenční řady

pomocí korelací

pomocí vzdáleností

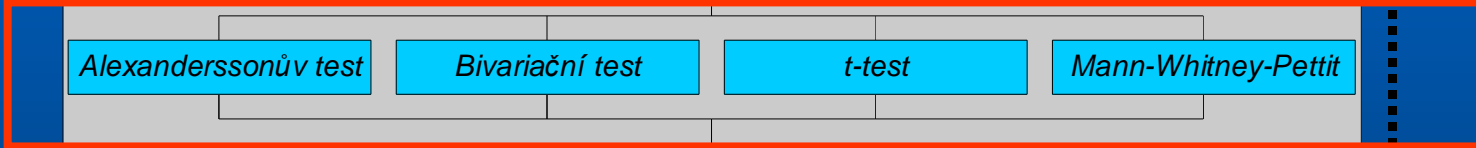
Vyhodnocení nehomogenit

Pravděpodobnost

Oprava dat

Doplnění chybějících hodnot

Několik iterací



Výhody statistického zpracování

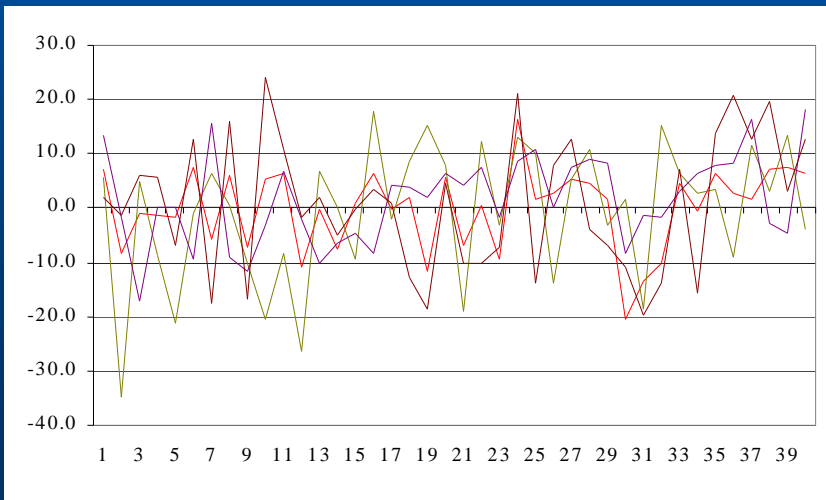
- **známe závažnost každé nehomogenity** (pravděpodobnost)
- **můžeme ohodnotit kvalitu měření dané stanice jako celku** (součet všech nehomogenit)

Referenční řady

- Quality control
- Homogenization
- Data Analysis

Referenční řady

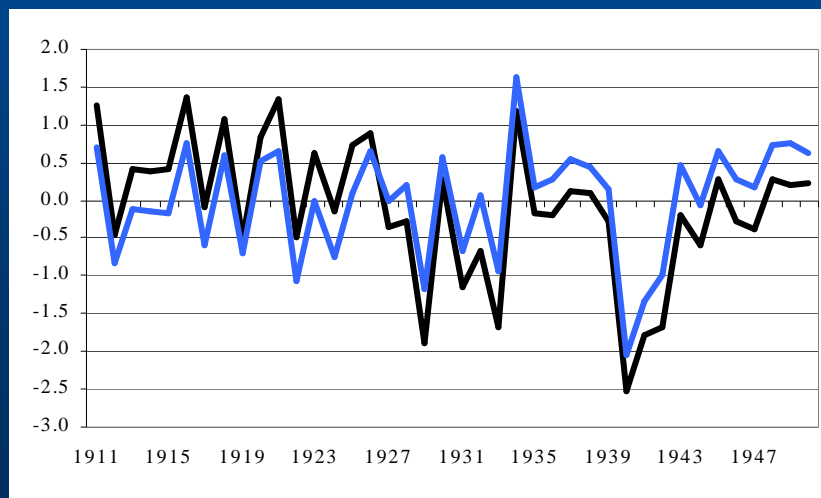
Regionální
průměr



Referenční řady

Regionální
průměr

Průměr
stanic s
nejvyššími
korelacemi

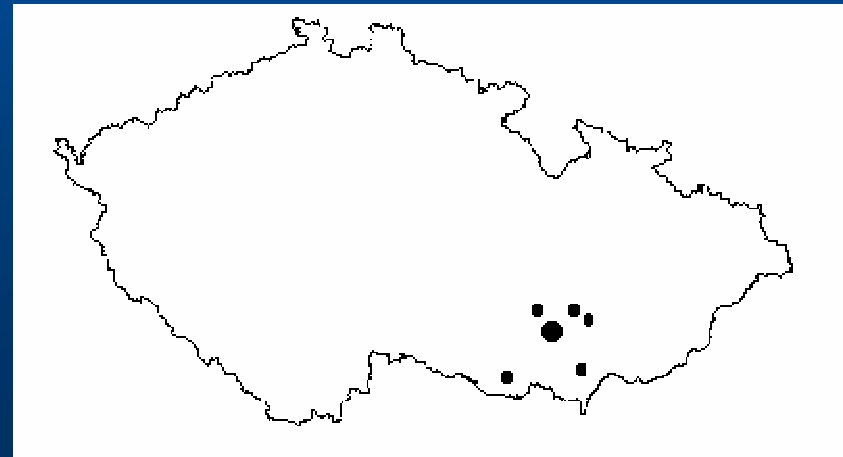


Referenční řady

Regionální
průměr

Průměr
stanic s
nejvyššími
korelacemi

Průměr
nejbližších
stanic



Referenční řady

Regionální
průměr

Průměr
stanic s
nejvyššími
korelacemi

Průměr
nejbližších
stanic

+ možné nehomogenity
jsou nejlépe potlačeny
- ref. řada je nejméně
korelovaná s test. řadou

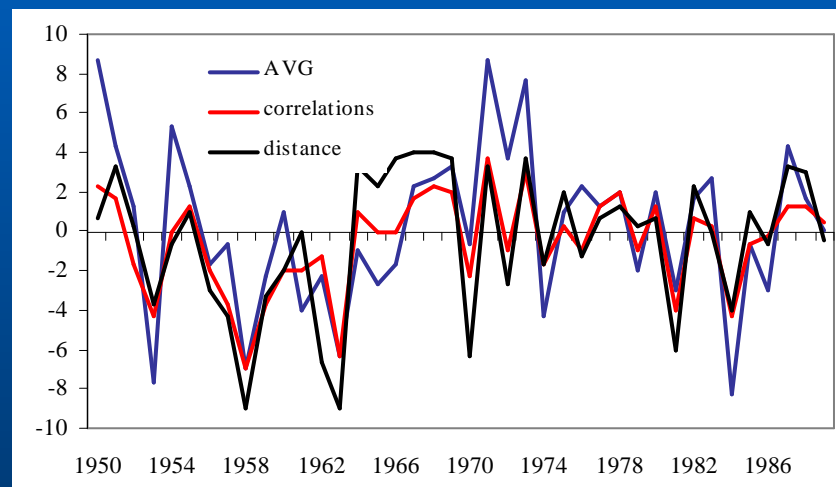
+ vytvořená ref. řada se
nejvíce podobá test. řadě
- podobné nehomogenity
s testovanou řadou

(řady prvních diferencí)

+ zachována
geografická blízkost
- různé klimatické
podmínky

Referenční řady se liší

diference testované a jejích referenčních řad



Example:

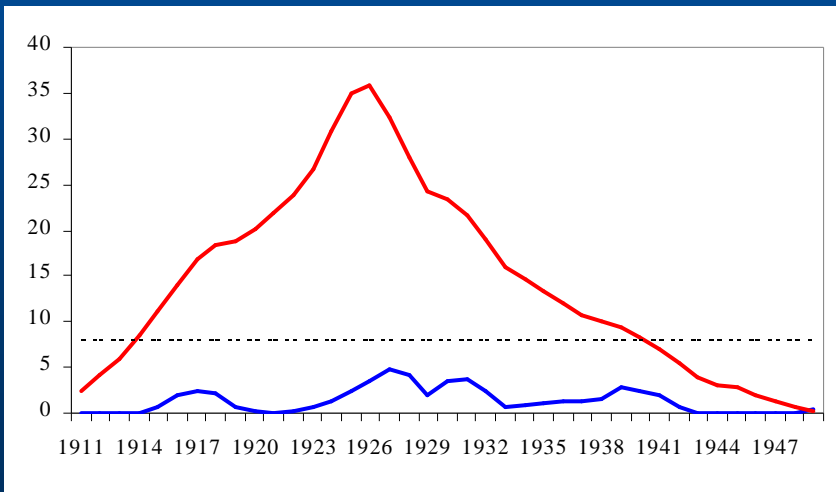
Proposed list of stations used for creating reference series

ID_1	ID_2	BEGIN	END	LENC	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

Testy homogeneity

Alexanderssonův test (SNHT)



kumulativní odchylky diferencí testované a referenční řady

Alexandersson Standard Normal Homogeneity Test (Single shift test)

Reference series:

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

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

Null and alternative hypothesis:

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

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

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

for $1 \leq a < n$ and $\mu_1 \neq \mu_2$.

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

Test statistic:

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

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

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

Testy homogeneity

Alexandersson
SNHT

Bivariate
Test

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 $d \neq 0$ -

$$N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho) \text{ pro } i < i_0$$

$$N(\mu_x, \mu_y + d, \sigma_x^2, \sigma_y^2, \rho) \text{ pro } i > i_0.$$

Test statistic:

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

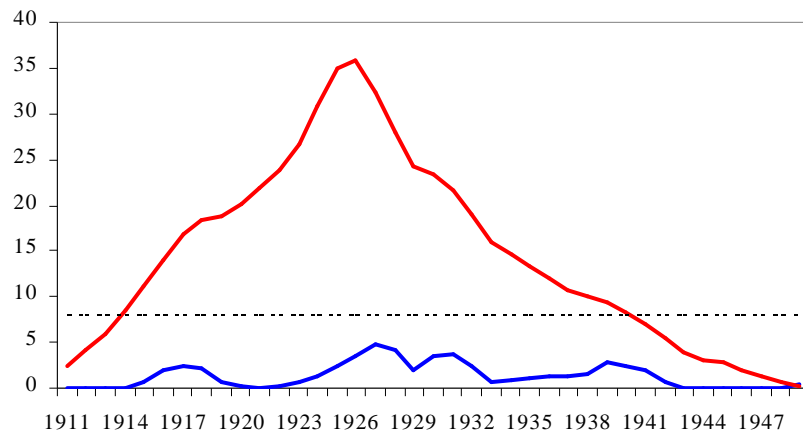
where: $X_i = 1/i \sum_{j=1}^i x_j$, $Y_i = 1/i \sum_{j=1}^i y_j$, $\bar{X} = X_n$, $\bar{Y} = Y_n$

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

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

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

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



Testy homogeneity

Alexandersson
SNHT

Bivariate
Test

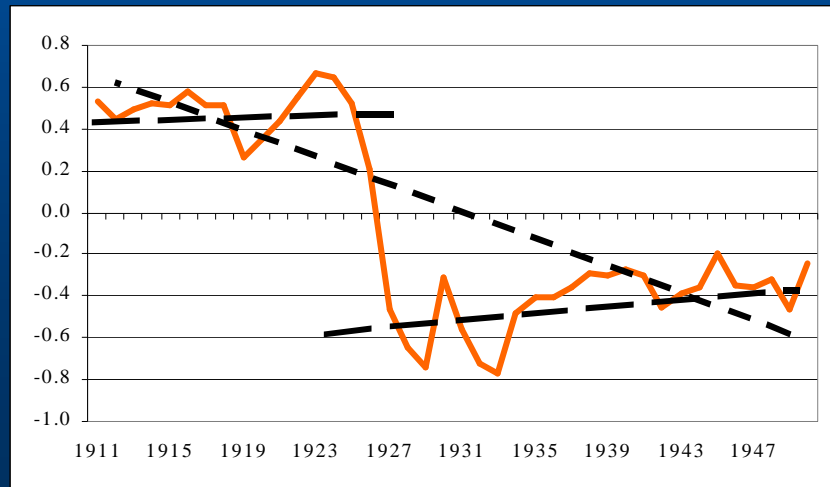
Vincent
Technique

Easterling and Peterson

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

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

dvoufázová lineární regrese



Testy homogeneity

```
graph TD; A[Testy homogeneity] --- B[Alexandersson SNHT]; A --- C[Bivariate Test]; A --- D[Vincent Technique];
```

Alexandersson
SNHT

Bivariate
Test

Vincent
Technique

40-ti leté úseky řad

(30-40 let na jednu nehomogenitu, Auer et al., 2001)

Posouzení nehomogenit

(výstupy ze softwarů)

Čáslav, 3. část, 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

- Quality control
- Homogenization
- Data Analysis

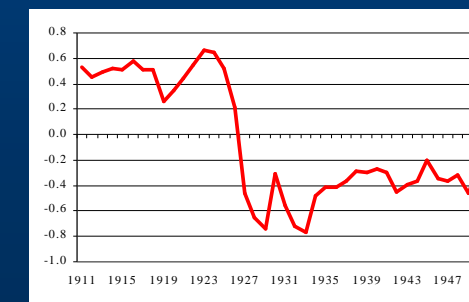
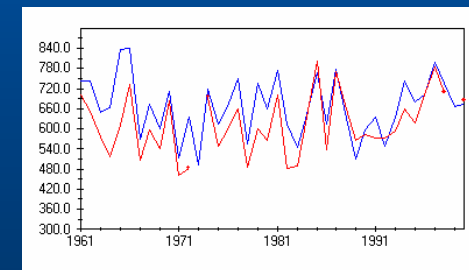
Posouzení nehomogenit

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		

Vyhodnocení nehomogenit

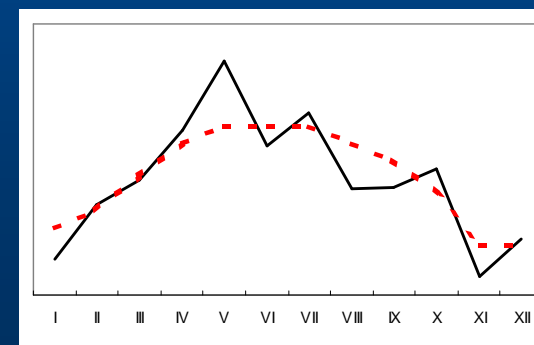
- Použití několika výstupů (sumace počtu detekcí v daném roce, použití metadat, grafy poměrů, ...)

ID	EL	YEAR	BEGIN	END	YEAR	COUNT	Y_POSSIBL	YEA	MIS	X_BEGIN	DX	END_DATE	X	X	LL	LAB	REMARK	CC
x B1BOJK01	x	1985				41	14.24		12	23.3.1984		31.3.2003	#	#			Echange	
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	#	#			Rchange	
B1RADE01	x	1994				44	15.28		2	1.1.1994		31.12.9999	#	#			obs	Jc B
x B1RYCH01	x	1973				49	17.01			1.5.1973		28.2.1991	#	#			Vchange	
B1RYCH01	x	1973				49	17.01			1.9.1972		28.2.1991	#	#			obs	MB
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	#	#			Uchange	
B1UHBR01	x	1983				31	10.76			18.2.1984		12.5.1993	#	#			obs	Jc B
x B1UHBR01	x	1984				77	26.74			18.2.1984		31.1.1999	#	#			Uchange	
B1UHBR01	x	1984				77	26.74			18.2.1984		12.5.1993	#	#			obs	Jc 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	#	#			Vchange	
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	#	#			Echange	
B2BRUM01	x	1989				51	17.71			1.2.1989		31.3.1994	#	#			obs	MB



Oprava nehomogenit

- z diferencí (poměrů) mezi opravovanou a referenční řadou
- pomocí referenční řady vypočtené jako průměr z nejlépe korelovaných stanic k dané stanici
- shlazení měsíčních hodnot oprav (1-2-1) (eliminace náhodných chyb)



Příklad:

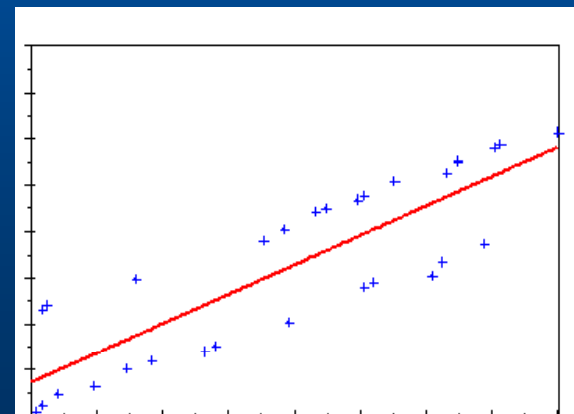
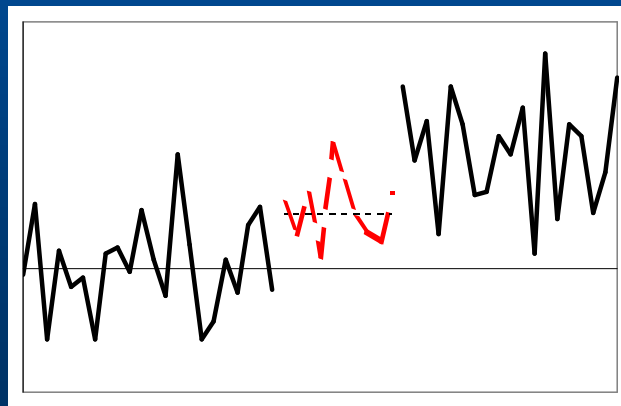
Oprava nehomogenit - vyhodnocení

ID_1	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

Doplnění chybějících hodnot

- Před homogenicí: vliv na správnou detekci nehomogenit
- Po homogenizaci: přesnější - neovlivněné případnými posuny v řadách

Závislost testované
na referenční řadě



Doplnění chybějících hodnot

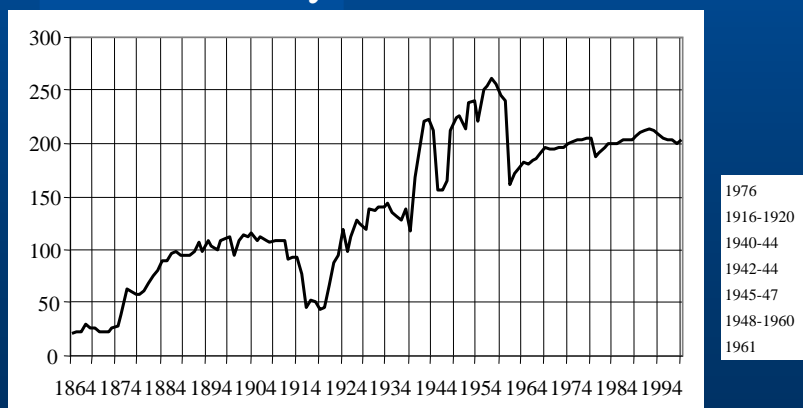
- pomocí lineární regrese
 - referenční řada jako průměr nejlépe korelovaných stanic
- výpočet „očekávané hodnoty“ při porovnání se sousedy

Homogenizace časových řad v České republice

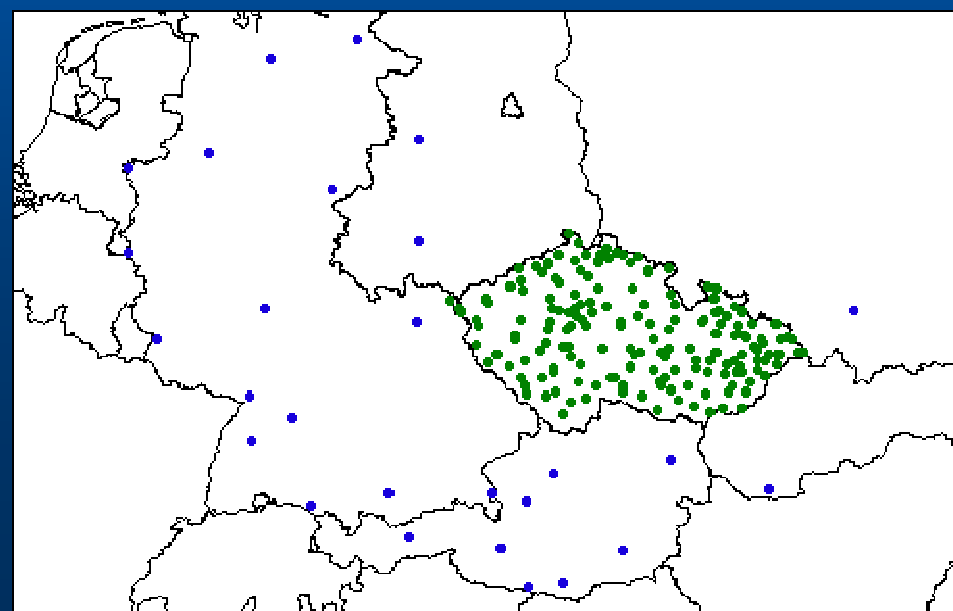
Homogenizace teplotních řad

- měsíční průměry teploty vzduchu
- téměř 200 stanic měřící v různém období v průběhu přístrojových měření
- testování změny v úrovni (průměru)

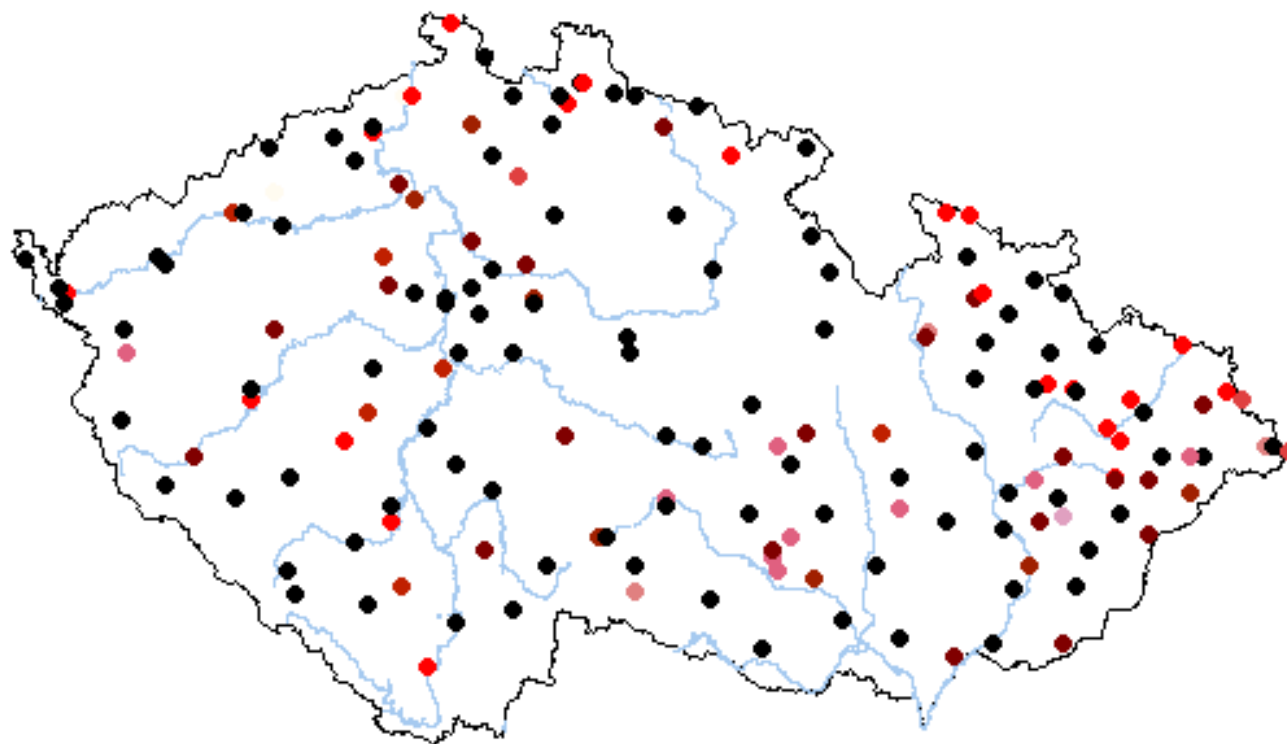
Počet klimatických stanic v ČR od roku 1864



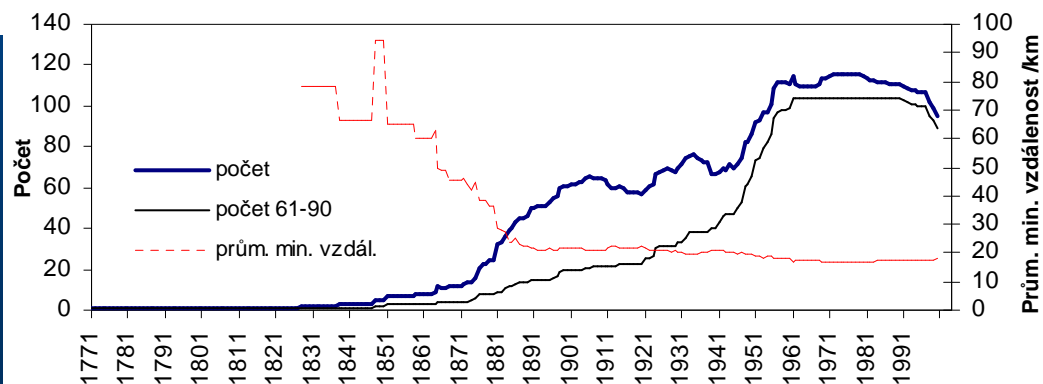
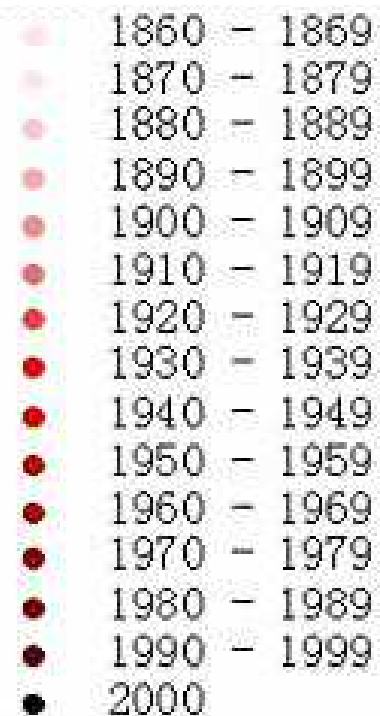
Jahrbücher der k. k. Zentral-Anstalt für Meteorologie und Erdmagnetismus 1848-1915. Wien.
Bericht der meteorologischen Commission des naturforschenden Vereines in Brünn 1881-1911. Brünn 1882-1917.
Ročenka povětrnostních pozorování meteorologických stanic 1916-1960. Praha 1934-1966.



Prostorové rozložení stanic ČR



Konec měření:



Homogenizace - přehled

Přehled počtu zpracovávaných řad při homogenizaci (ČR)

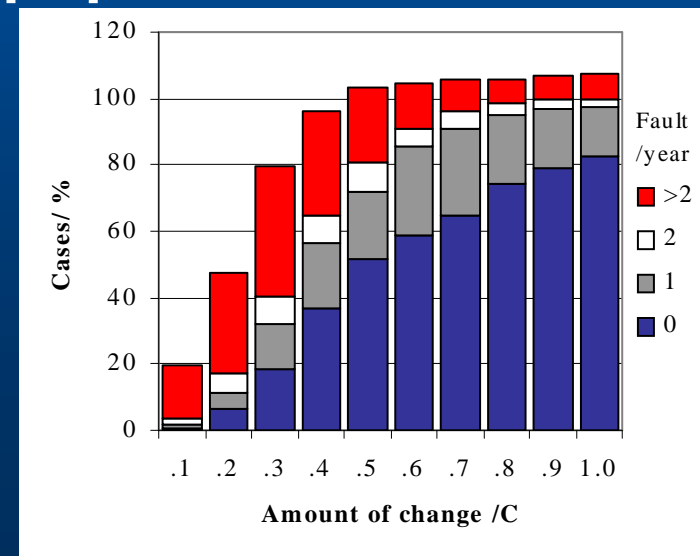
Charakteristika	Data	
	původní	opravená
Počet stanic	192	174
Počet 40-ti letých úseku řad	348	307
Počet oprav		231
Počet testovaných řad celkem	40716	35919
Počet významných nehomogenit ($p=0.05$)	32445	13802
<i>Podíl význ. nehomogenit na počtu řad</i>	79,7%	38,4%

Počet testovaných řad – původní data

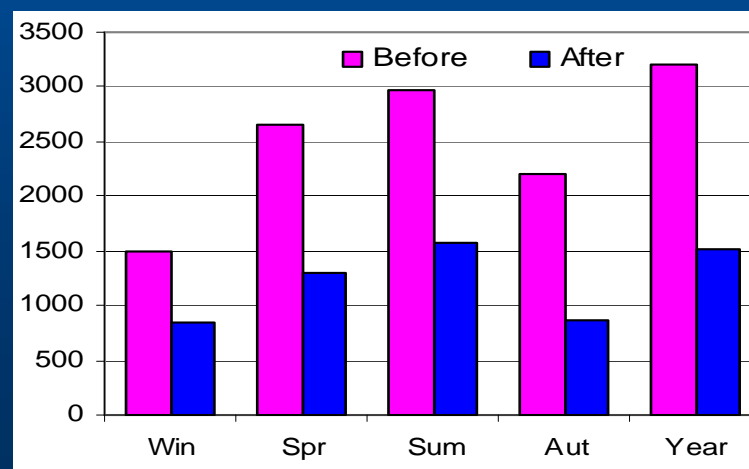
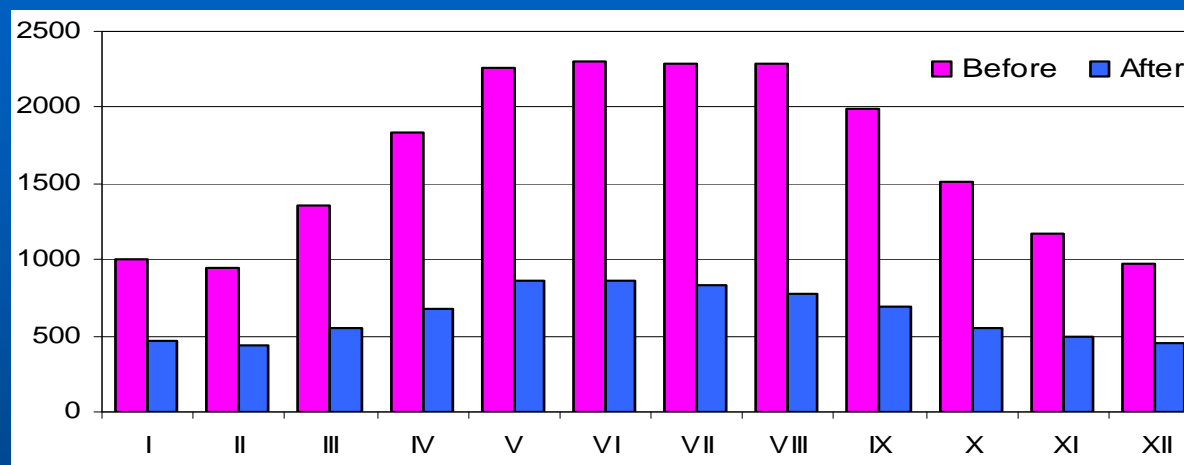
Test	Měsíce	Sezóny a rok	Typy ref. řad	40-ti leté úseky řad	Počet řad pro test
A	12	5	3	348	17748
B	12	5	3	348	17748
V		5	3	348	5220
Celkem					40716

Velké množství nehomogenit po opravě?

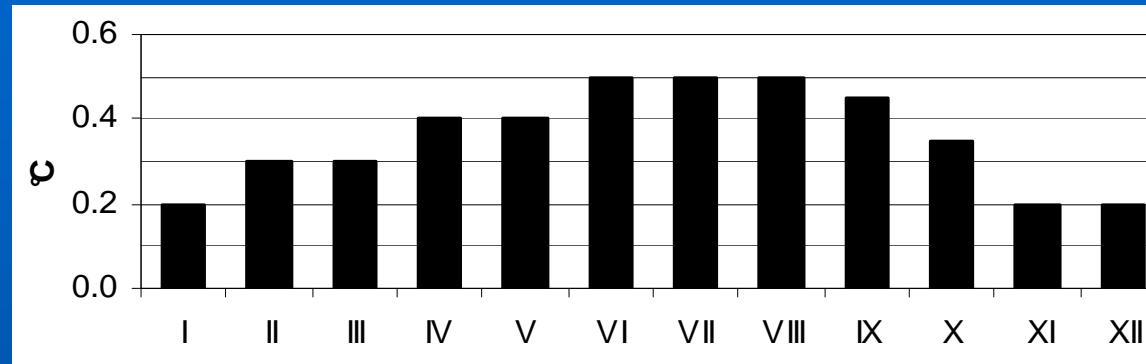
- 40% nehomogenních řad po homogenizaci (80% před)
- Šumová složka řad: nehomogeneity pro skok menší než 0.5 °C jsou správně detekovány v méně než 50% případů



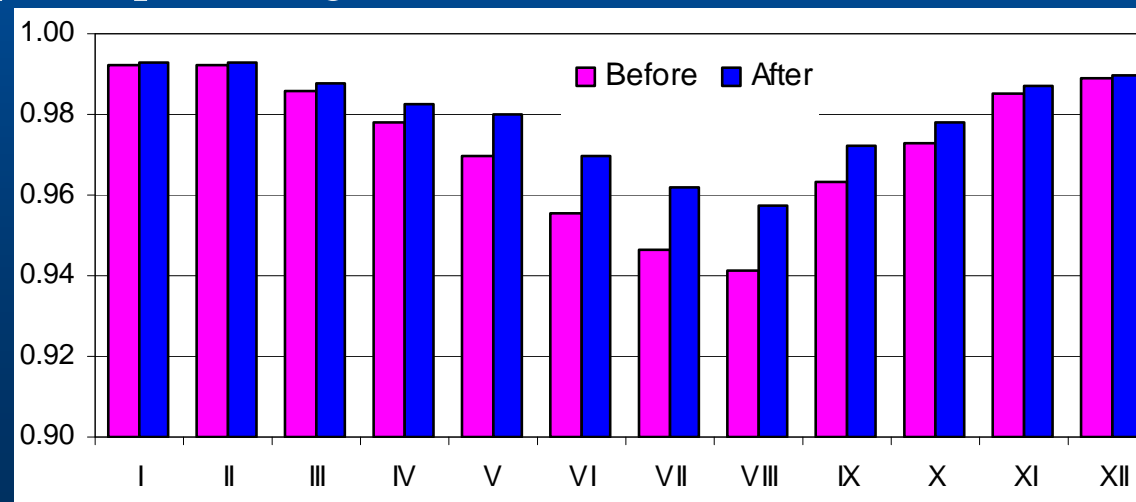
Počet významných nehomogenit detekovaných před a po homogenizaci ($p=0.05$)



Velikost opravy homogenizovaných řad (absolutní hodnoty) - median



Korelační koeficienty mezi testovanými a referenčními řadami před a po homogenizaci (median)



Nehomogenity v létě versus v zimě

- Změna měřících podmínek na stanici (přemístění atd.) se projevuje především v létě
- v zimě: menší role aktivního povrchu, převládají cirkulační faktory, v létě: větší role aktivního povrchu, převládají radiální faktory

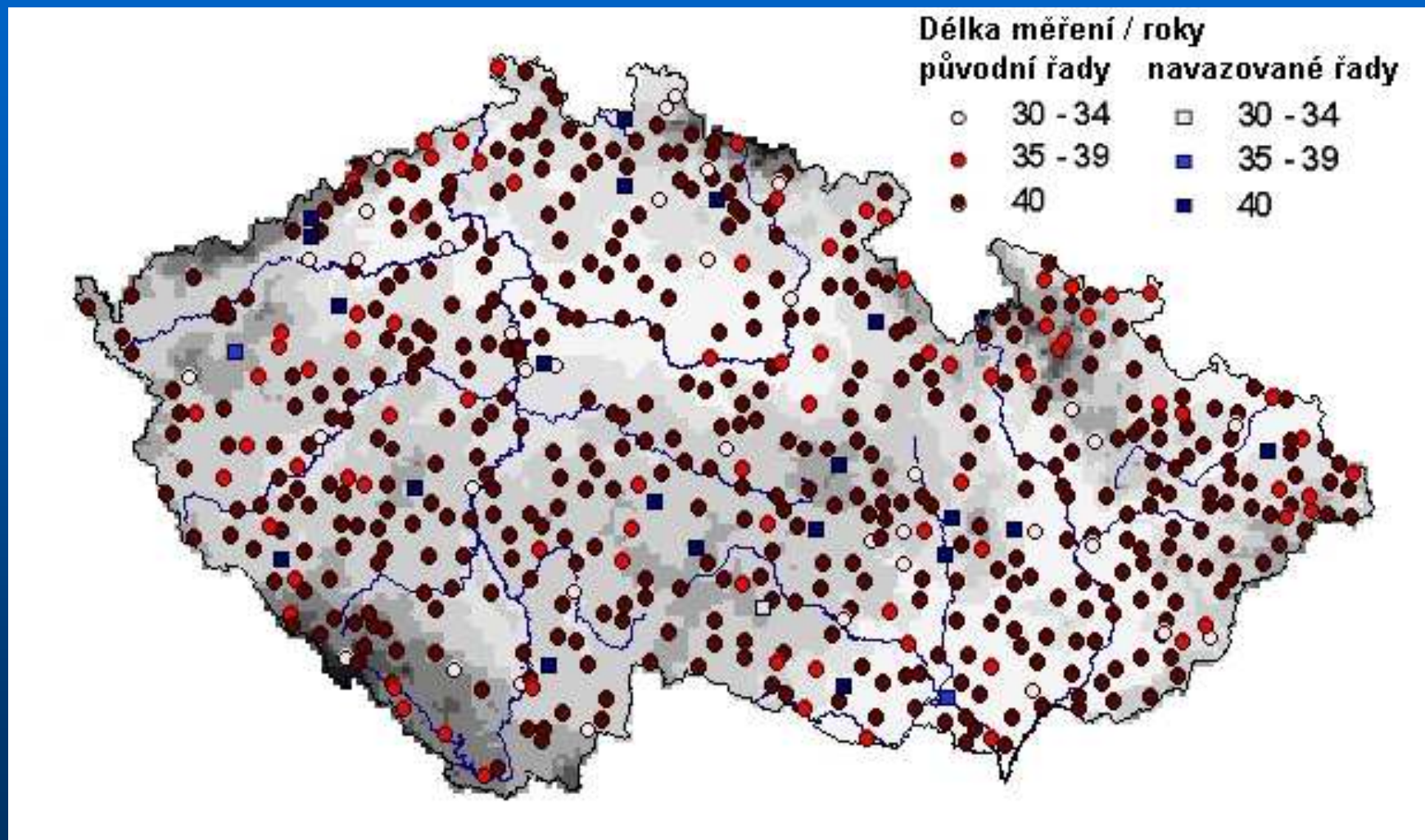
Přechod na automatická měření - konsekvence

- **Zatím je příliš brzy na opravu řad - je k dispozici málo hodnot (ale nehomogenity způsobené přechodem na AMS jsou již detekovatelné)**
- **Nemohou se připojit nová měření**

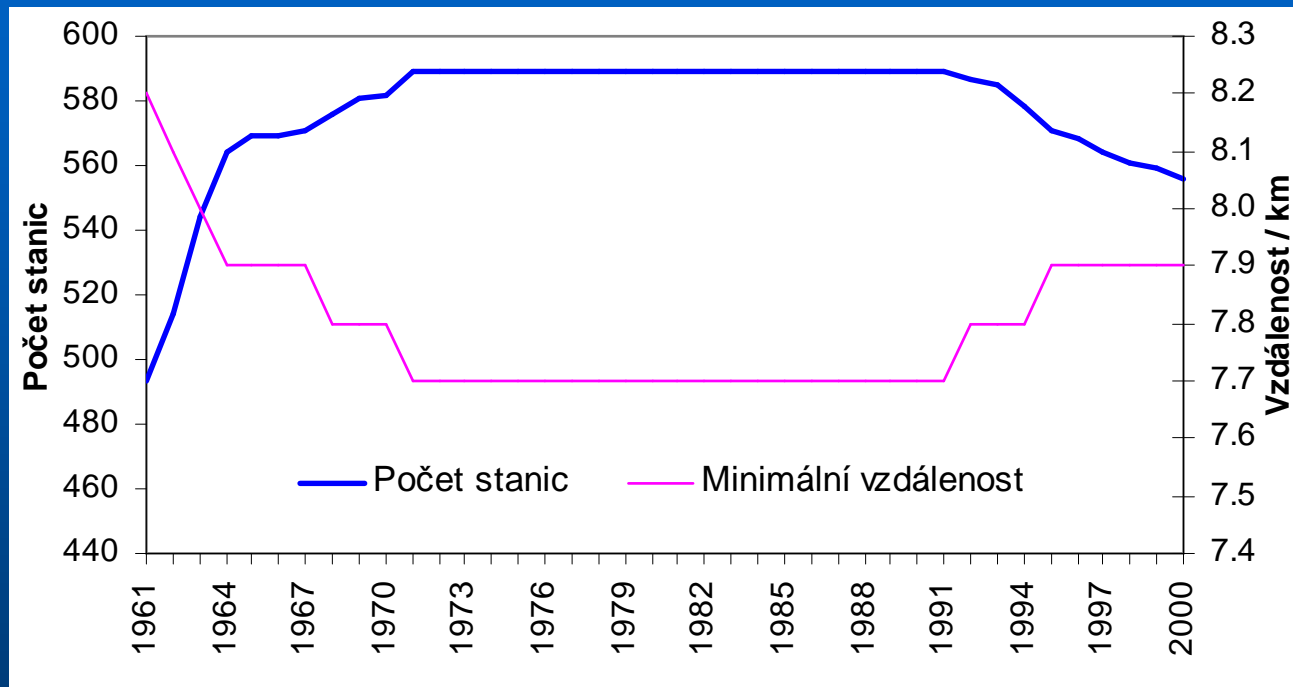
Homogenizace srážkových řad

- měsíční sumy (+sezónní a roční sumy), sezónní a roční počty srážkových dnů (s denními úhrny srážek $\geq 0,1, 1,0, 5,0$ a $20,0$ mm)
- období 1961-2003
- 589 stanic
 - 566 řad s měřeními delšími více než 30 let (přerušeni měření nesmělo být delší než 4 roky)
 - 23 řad vytvořené kombinací sousedních stanic
 - 457 stanic má měření delší více než 40 let
- průměrná minimální vzdálenost: 7.5 km

Prostorové uspořádání použitých srážkoměrných stanic ČR měřících v období 1961-2000



Počet použitých stanic (s minimální délkou pozorování 30 let) a průměrná minimální vzdálenost těchto stanic v jednotlivých letech v období 1961-2000.

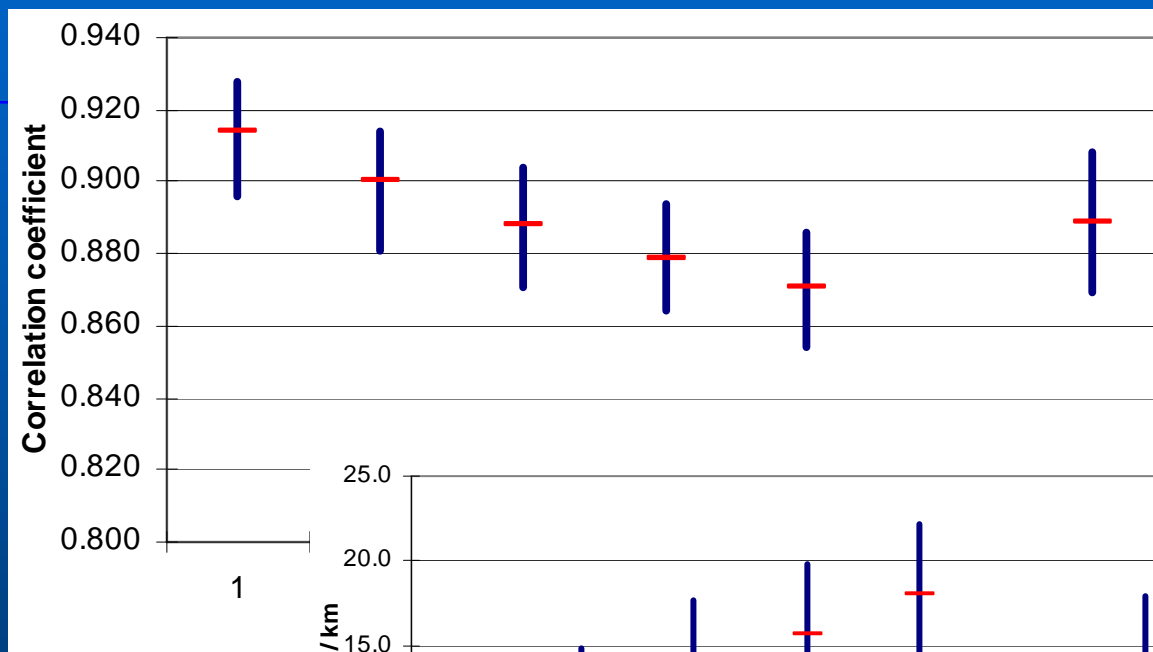


Průměrná minimální vzdálenost použitých stanic: 7.5 km

Tvorba referenčních řad:

Statistika pro výběry nejlépe korelovaných stanic

(pro všechny testované řady)



Boxplots:

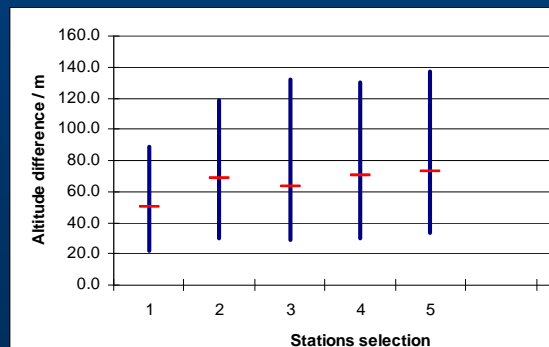
- Median
 - Upper and lower quartiles
- (for 589 testes series)

Vzdálenosti:

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

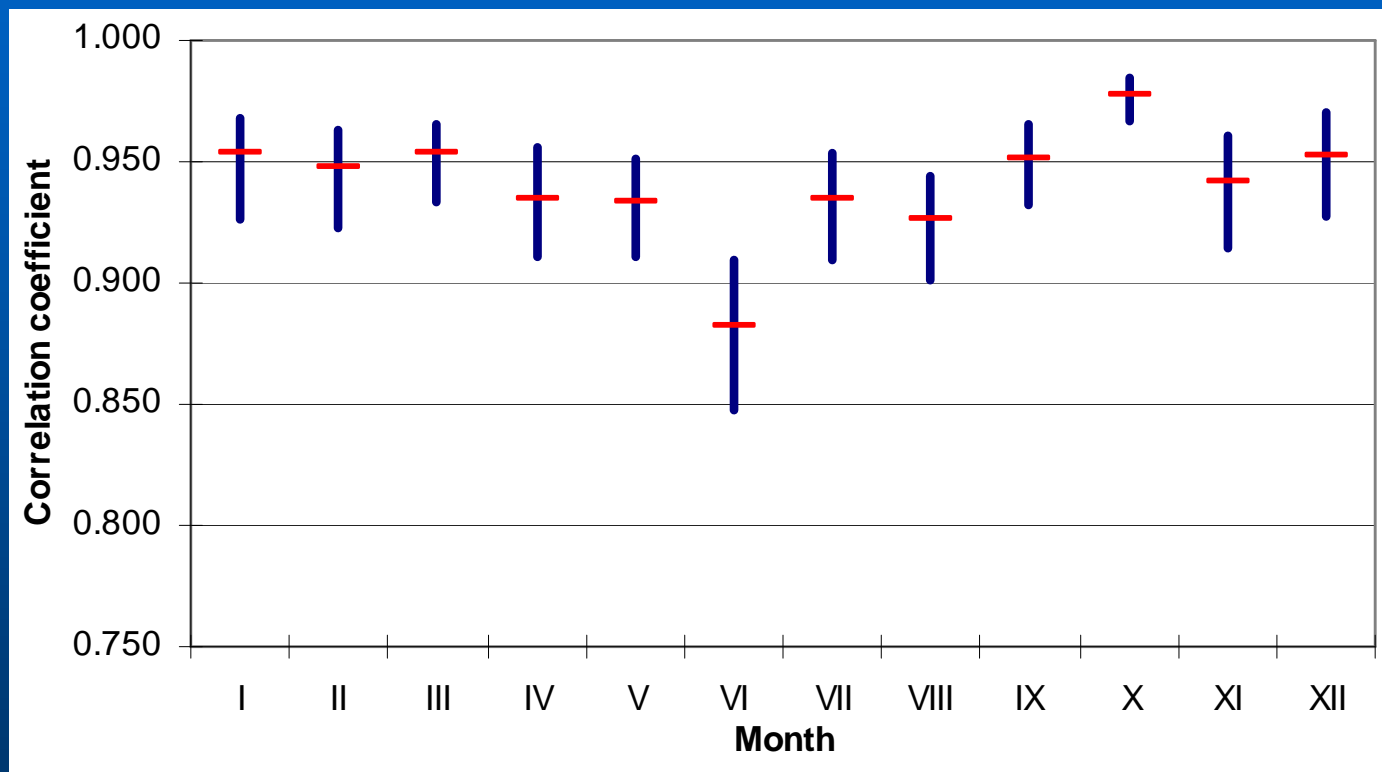
Rozdíl v nadm. výšce

(absolutní hodnoty)



Korelační koeficienty mezi testovanými a referenčními řadami

(referenční řada počítána podle korelací)

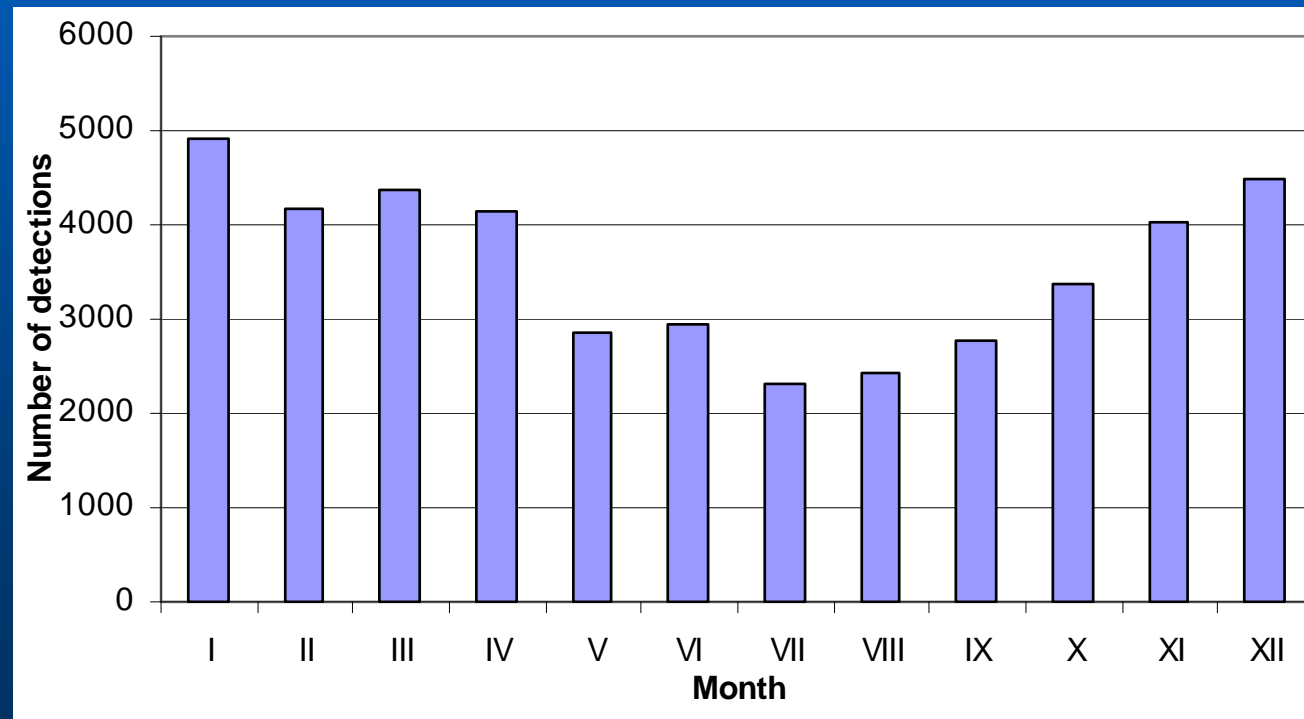


Boxplots:

- Median
 - Upper and lower quartiles
- (for 589 testes series)

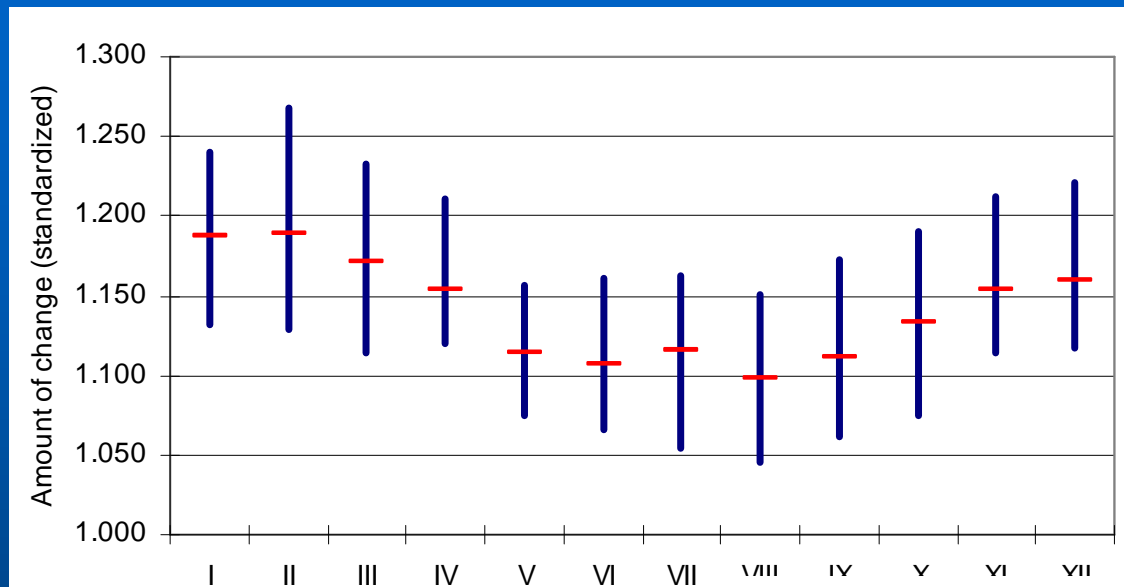
Výsledky testování homogenity

- Počet detekovaných nehomogenit (stat. významných)



Velikost opravy (poměry - standardizace na >1.0)

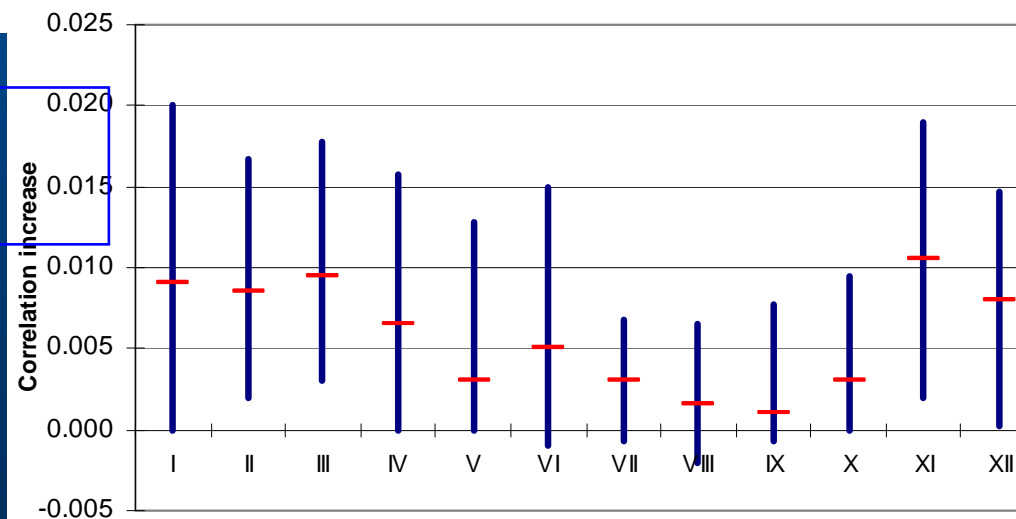
(referenční řada počítána podle korelací)



Boxplots:

- Median
 - Upper and lower quartiles
- (for 589 testes series)

Vzrůst korelačních koeficientů



Nehomogenity v létě versus v zimě

- Změna měřících podmínek na stanici (přemístění atd.) se projevuje především v zimě
- v zimě: větší chyby při měření (tuhé srážky – vítr, ...)

Závěr, poznámky

- kontrola kvality dat před samotnou homogenizací je velmi důležitá
- použití stanic z okolí ČR (měřící především v začátcích přístrojových pozorování)
- testování v několika iteracích
 - několik iterací testování homogenity a opravy řad (3 iterace byly v tomto případě dostačující)
 - problém homogenity referenčních řad je tímto vyřešen:
 - jednak jsou možné nehomogenity eliminovány použitím průměrů několika sousedních stanic
 - pokud toto není splněno: při dalším kroku (iteraci) by sousedé měly být již homogenní
- doporučeno testovat jednotlivé klimatické termíny
- problémy spojené s automatizací měření v posledních letech

Závěr

- roční chod počtu statisticky významných detekcí nehomogenit a velikostí oprav nehomogenit
- ensembles: pomocí pravděpodobnosti dané nehomogenity (závažnosti) – kvalitnější vyhodnocení, ohodnocení měření dané stanice jako celku (případně různých období měření)
- ...

Software použitý pro zpracování dat

- **LoadData** - aplikace pro stažení dat z centrální databáze (např. Oraclu)
- **ProClimDB software** pro zpracování celé databáze (nalezení vychýlených hodnot, navázání sousedních řad, tvorba referečních řad, příprava dat pro testování homogeneity, ...)
- **AnClim software** pro testování homogeneity

<http://www.klimahom.com/software>

AnClim software

AnClim (4.39)

File Tools Statistics Homog 1 Homog 2 Analyse 1 Analyse 2 Filters Options Window Help

Low-pass Filter: a_prumCR.txt

Low-pass Filter: Gaussian ordinate method

Ordinate (weight) -4, +4 a

0.0224
0.0790
0.1942
0.3332
0.3989
0.3332
0.1942
0.0790
0.0224

Plots of Filtered a_prumCR.txt (Yea)

Win/Spr

PS - MESA: a_prumCR.txt

Power Spectrum - MESA

Frequencies + Values + Period	M = 30
0.0000 : 674.3299 < : 1	
0.0042 : 716.3279 < : 24	
0.0083 : 808.9999 < : 12	
0.0125 : 802.4849 < : 8	
0.0167 : 601.3849 < : 6	
0.0208 : 390.8654 < : 48	
0.0250 : 266.0807 < : 40	
0.0292 : 204.7484 < : 34	
0.0333 : 181.4865 < : 30	
0.0375 : 186.5342 < : 28	
0.0417 : 224.4611 < : 24	
0.0458 : 320.5823 < : 21	
0.0500 : 537.5234 < : 20	
0.0542 : 870.4781 < : 1	
0.0583 : 823.4554 < : 1	
0.0625 : 512.3353 < : 18	
0.0667 : 335.1720 < : 18	

Estimates related to

Harmonics

Frequencies

Normalize PS % Variance

Plot WN

Plot Confidence Limits 95%

Save with Conf. Limits

Graph Save Save All Series Close

Win/Spr/Sum/Aut/Yea/

PS - Dynamic MESA - 3D : a_prumCR.txt

Graph Close

Series Controller

Active File Selection: *Open Files: 9*

D:\...\anom\va_prumCR.txt

Period: 1848 - 2000; 1 Missing Values

Series

Single series

Merged Series of one File

Merged Series of two Files

Analyzing

Simple series

Differences (Temperature)

Ratios (Precipitation)

Open all series of the file Use Seasonal and Annual Averages

Number of Series: 5

PS - MESA: a_prumCR.txt | D:\Dokumenty\diss33\vysl_hom\anom\va_prumCR.txt | 5 / 8

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 length +/- overlay

Correlations column: K13

Process info:

Number of stations: 5
Difference in measuring periods (base and selected) taken into account!
Neighbours selected according to: correlation 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 (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

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

Ready for action

NUM

