An Approach to Homogenization of Air Temperature Series in the Czech Republic during Period of Instrumental Measurements

Petr Štěpánek

Department of Geography, Masaryk University, Brno, CZ

Czech Hydrometeorological Institute, local office Brno, CZ

Homogenization

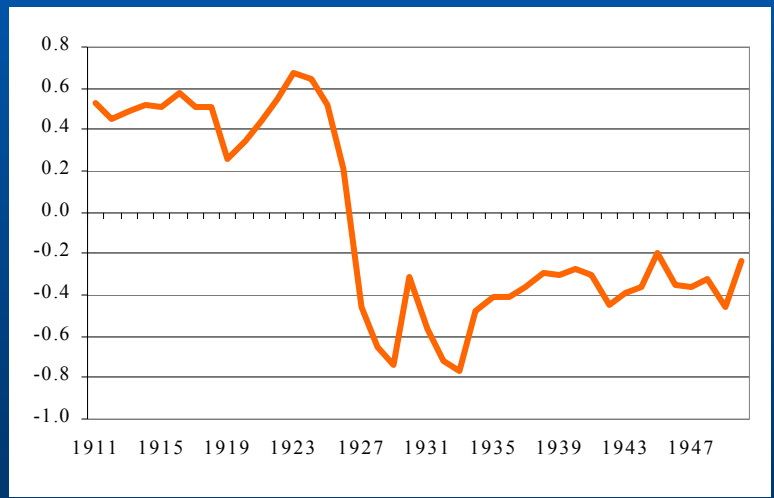
- monthly averages of air temperature measurements
- almost 200 stations measuring in some period during instrumental measurements (1771-2000)
- change in level (mean)

Climatological studies

- **Measuring and collecting data**

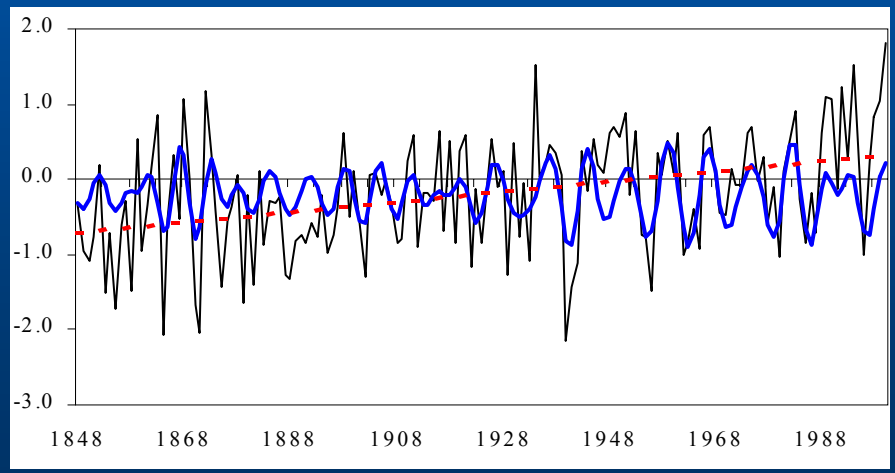
Climatological studies

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



Climatological studies

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



Data Quality Control

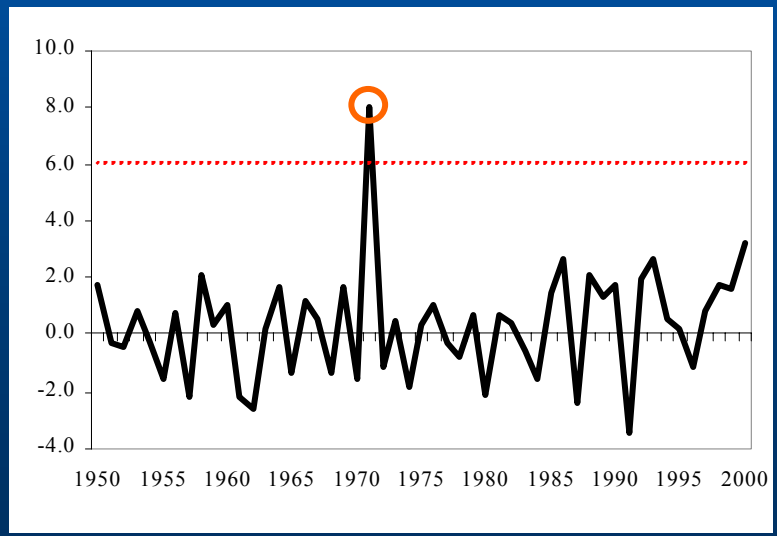
- Quality control
- Homogenization
- Data Analysis

Data Quality Control

Metadata

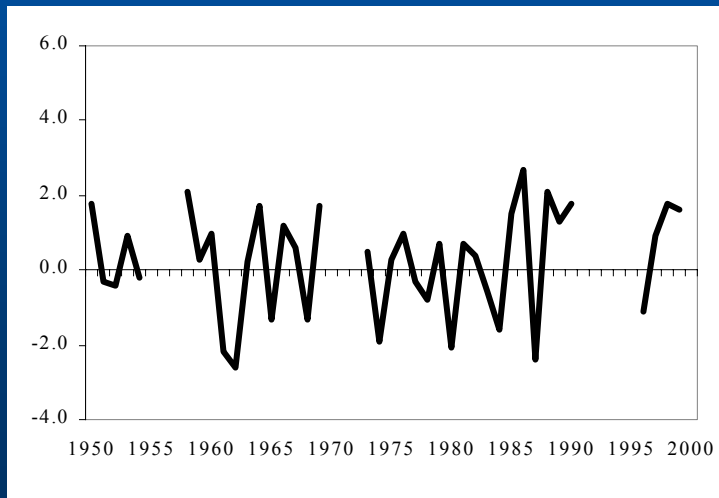
Data Quality Control

- Metadata
- Outliers



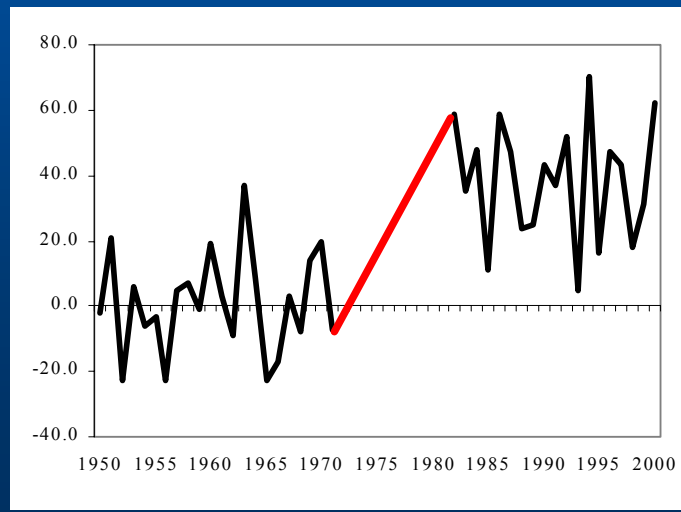
Data Quality Control

- Metadata
- Outliers
- Missing Data



Filling of missing values

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



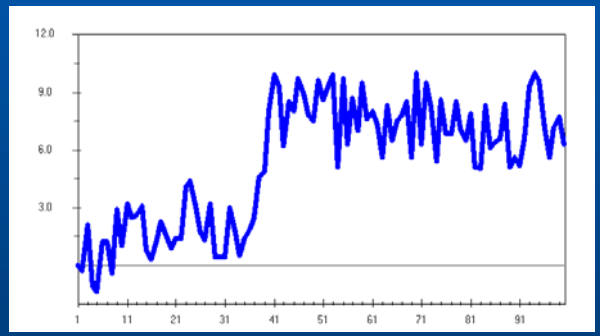
Homogenization

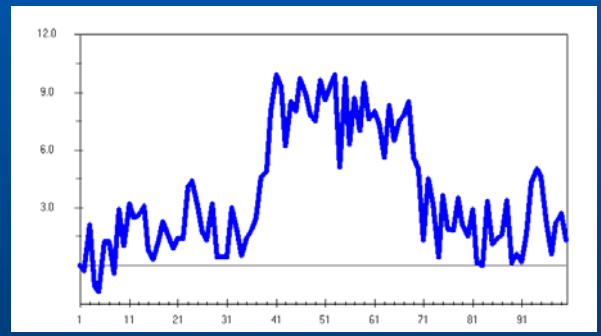
- Quality control
- Homogenization
- Data Analysis

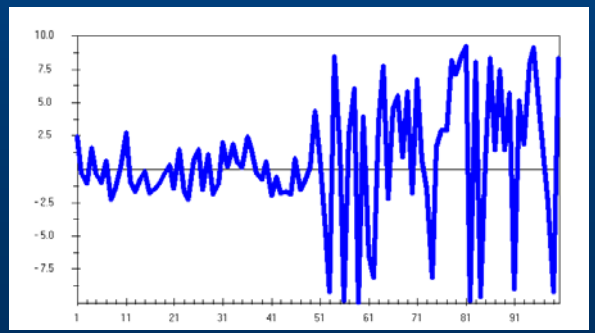
Homogenization

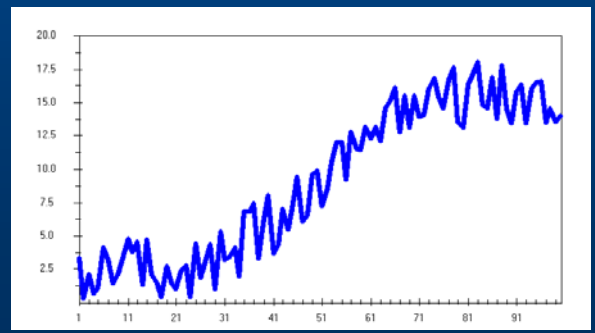
- **Change of measuring conditions**

  → **inhomogeneities**



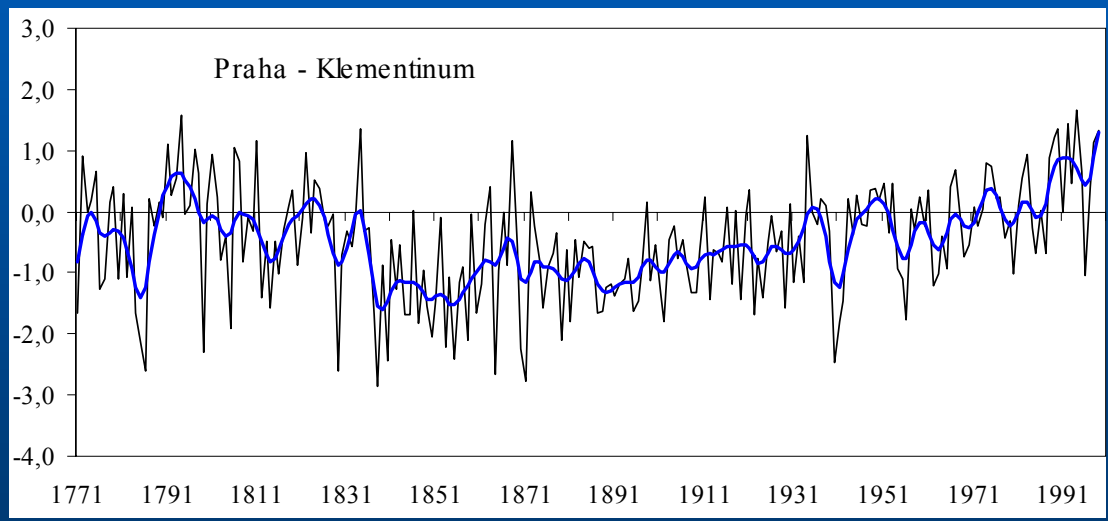






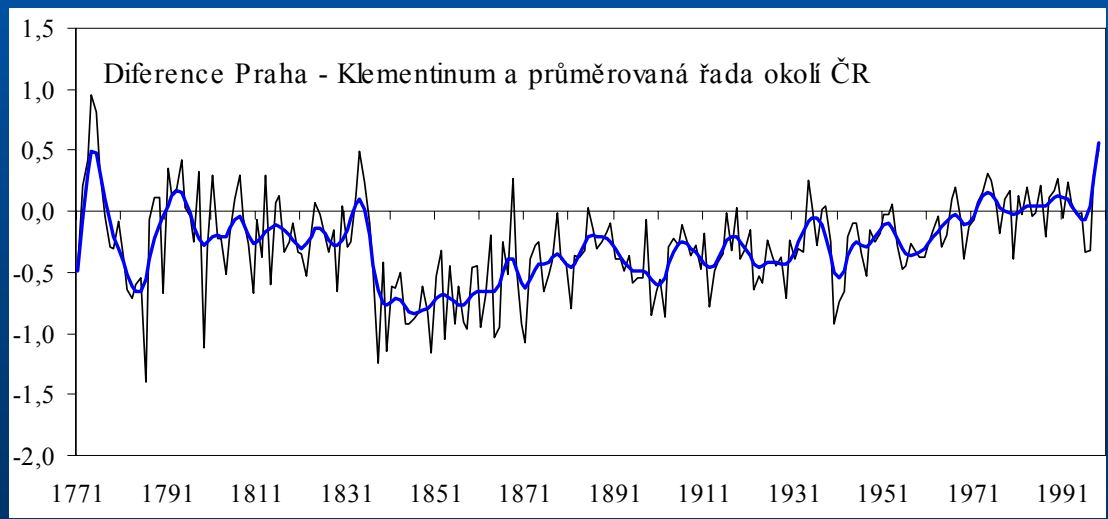
Inhomogeneity Detection

- **Absolute Homogeneity Testing**



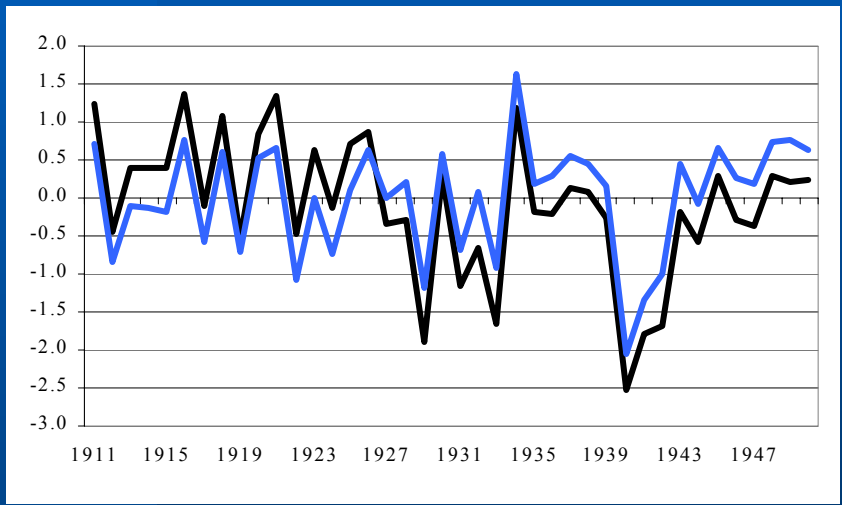
Inhomogeneity Detection

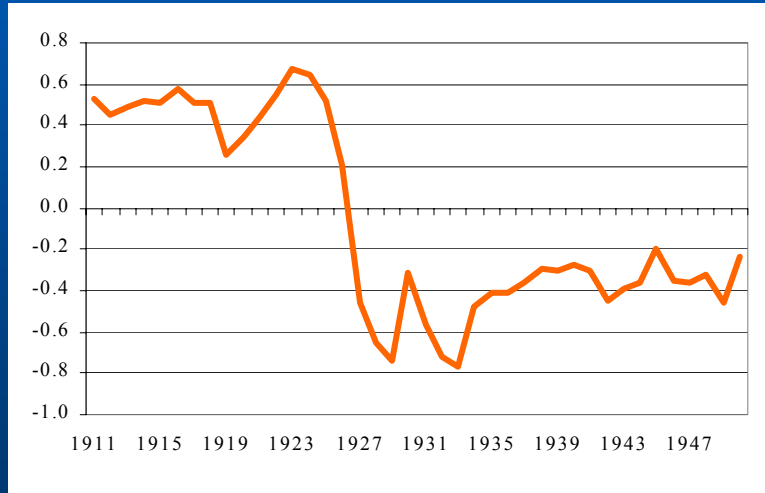
- **Absolute Homogeneity Testing**
- **Relative Homogeneity Testing**



Relative Homogeneity Testing

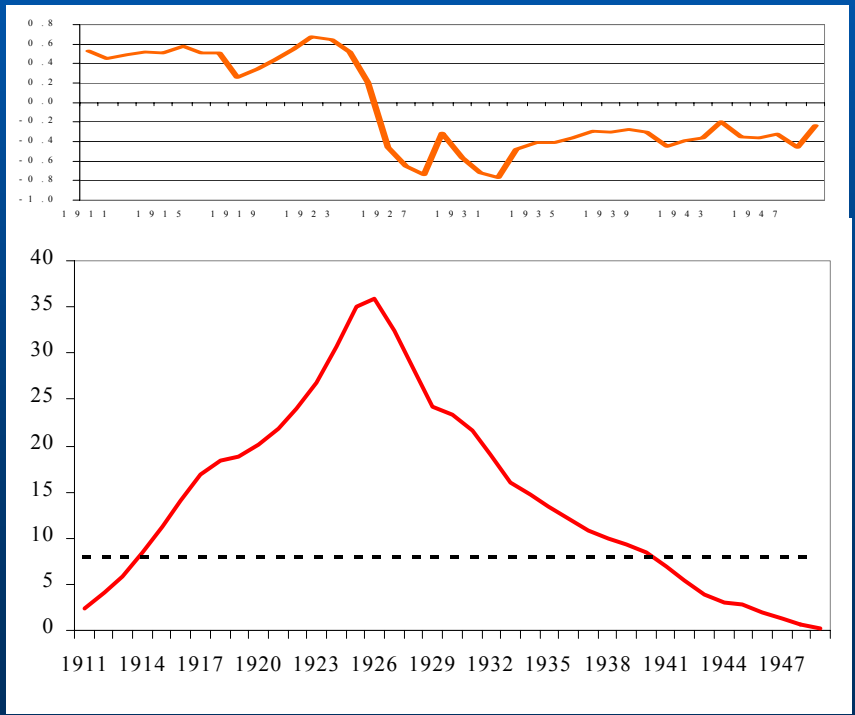
- **Creating reference Series**





Relative Homogeneity Testing

- Creating reference Series
- Tests of homogeneity



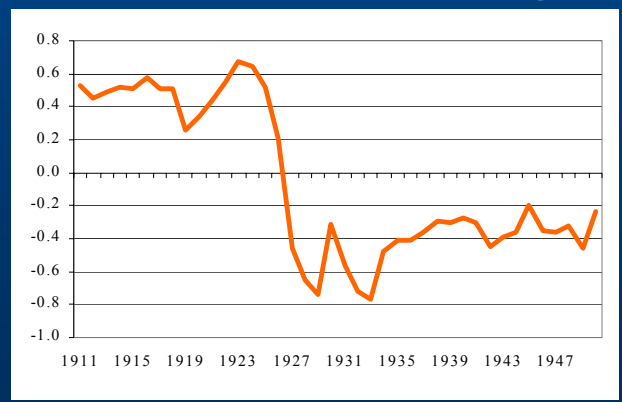
Relative Homogeneity Testing

- Creating reference Series
- Tests of homogeneity
- Assessing homogeneity

- Metadata

- physically justified

("undoubted" inhomogeneity)



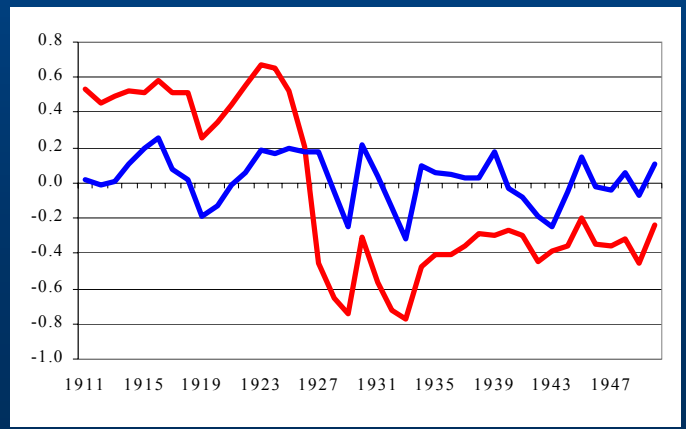
Relative Homogeneity Testing

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

  - Metadata

  - physically justified ?
  (“undoubted” inhomogeneity)

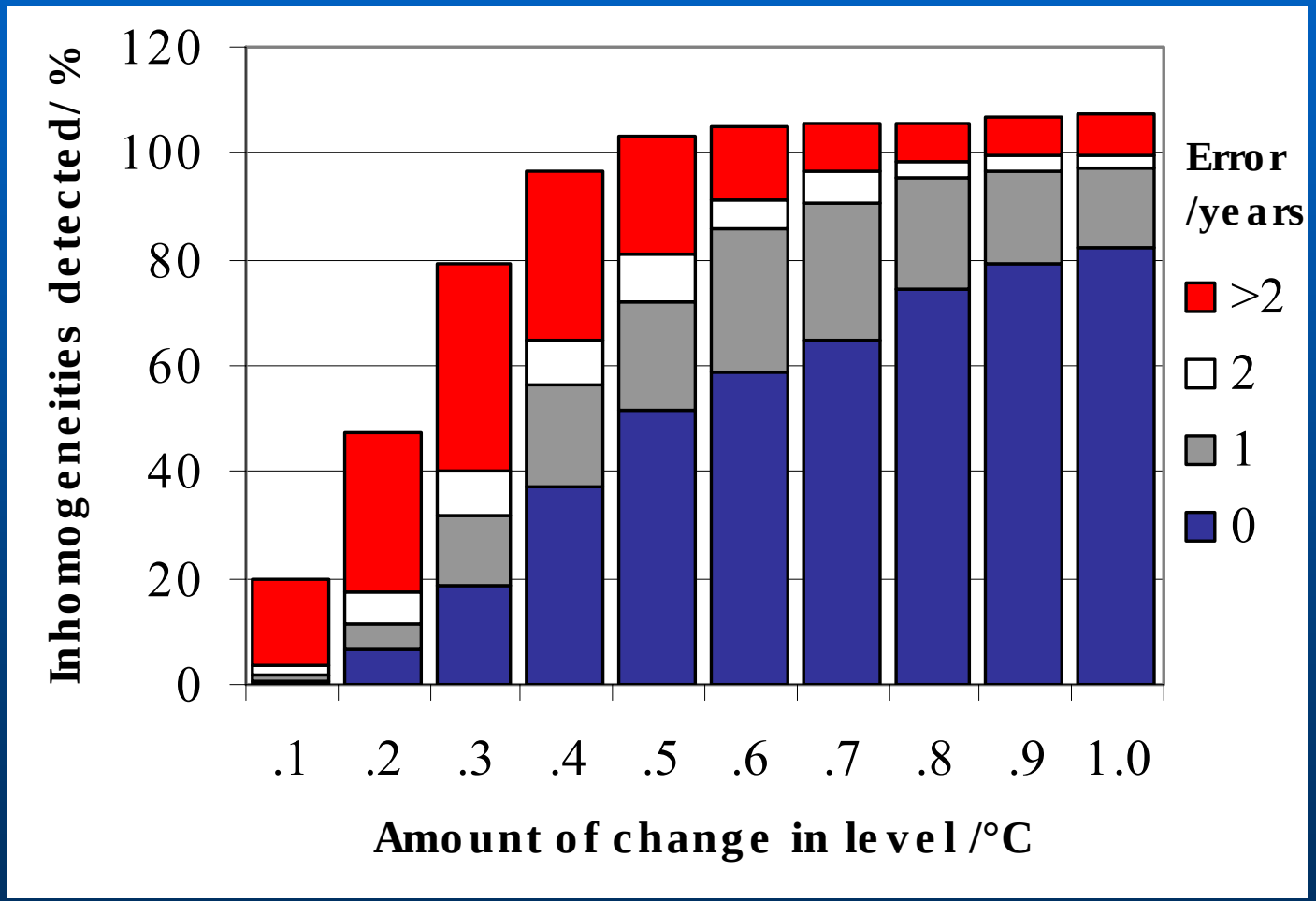
- **Adjusting Series**



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

- **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**

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

Proposed solution

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

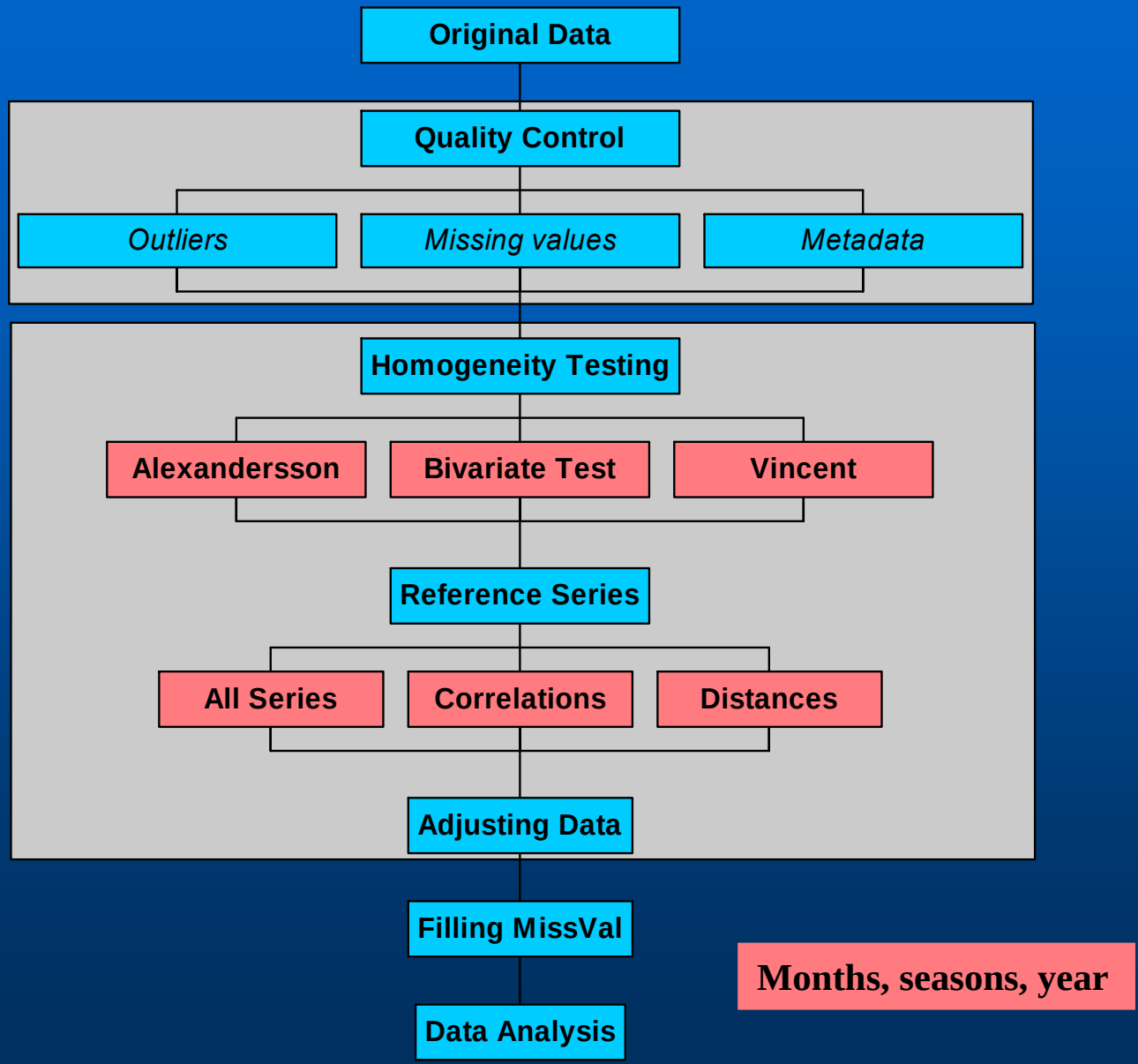
→ **Statistical processing of big amount of test results for each individual series**

For each year, group of years and whole series:
portion of number of detected inhomogenities in number of all possible (theoretical) detections

Adventages of statistical processing

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

How to increase number of test results

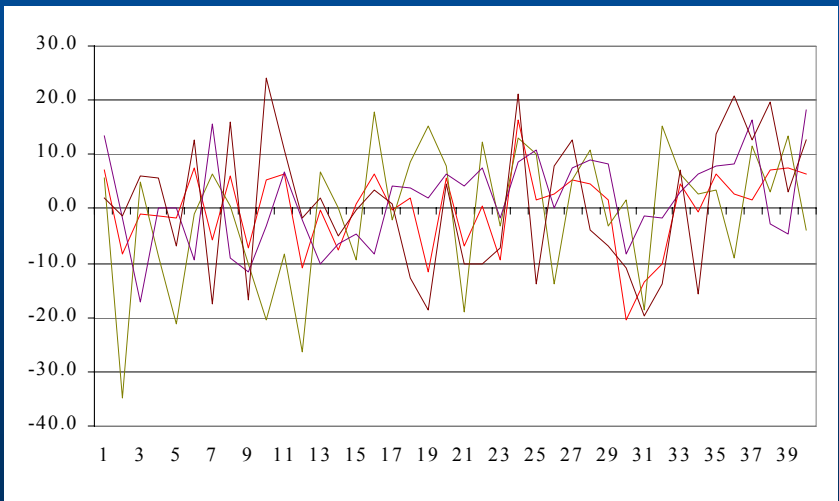


Reference Series

- Quality control
- Homogenization
- Data Analysis

Reference Series

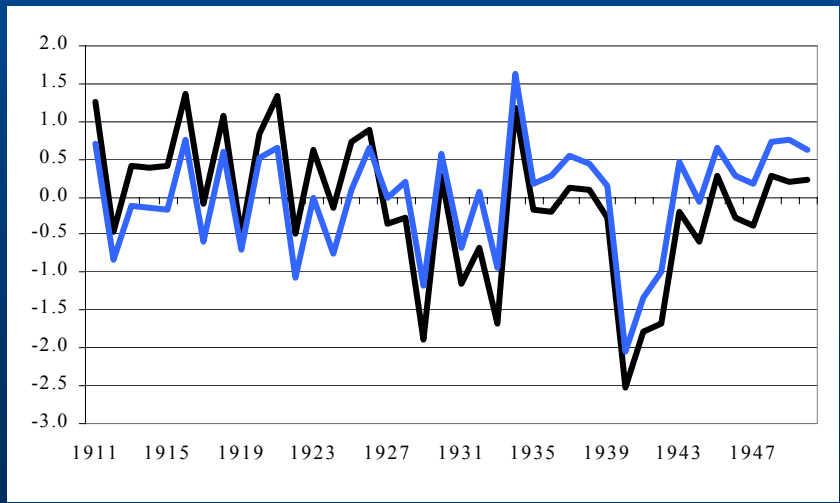
Average of all series **available**



Reference Series

Average of all series **available**

Average from mostly **correlated** series

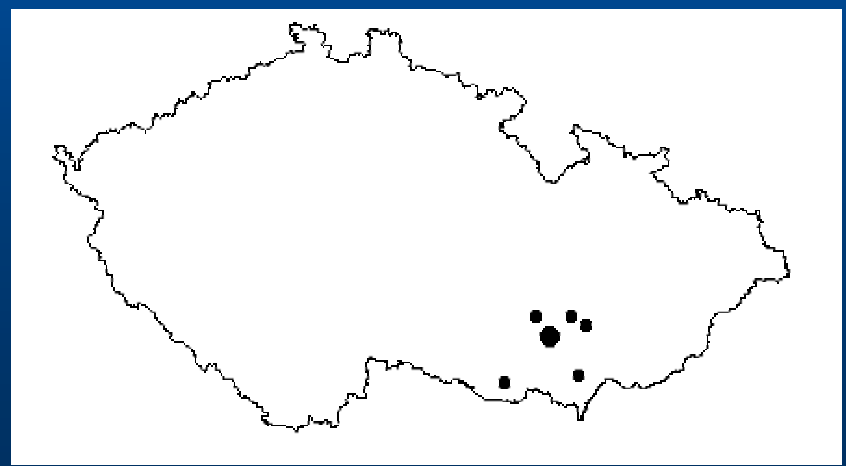


Reference Series

Average of all series **available**

Average from mostly **correlated** series

Average from the **nearest** stations



Reference Series

Average of all series **available**

Average from mostly **correlated** series

Average from the **nearest** stations

+ possible inhomogeneities are suppressed the most

- the least correlated with tested series

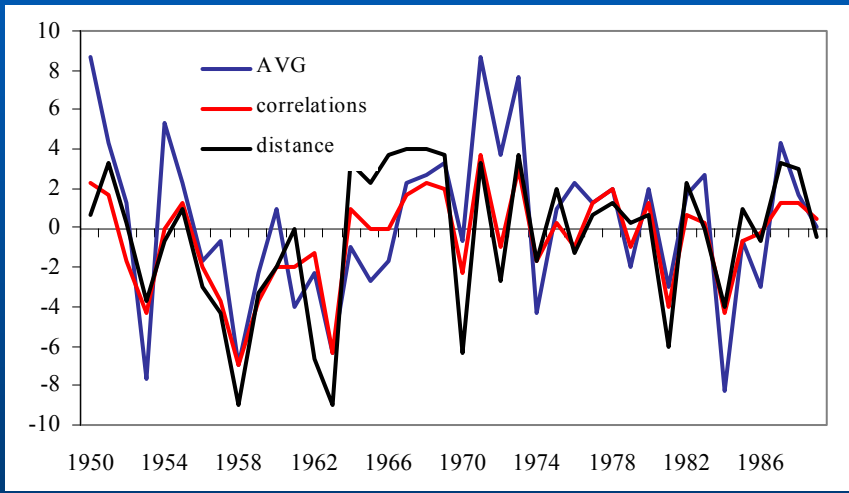
+ created reference series most similar to tested one

- similar inhomogeneities with tested series

+ geographical vicinity preserved

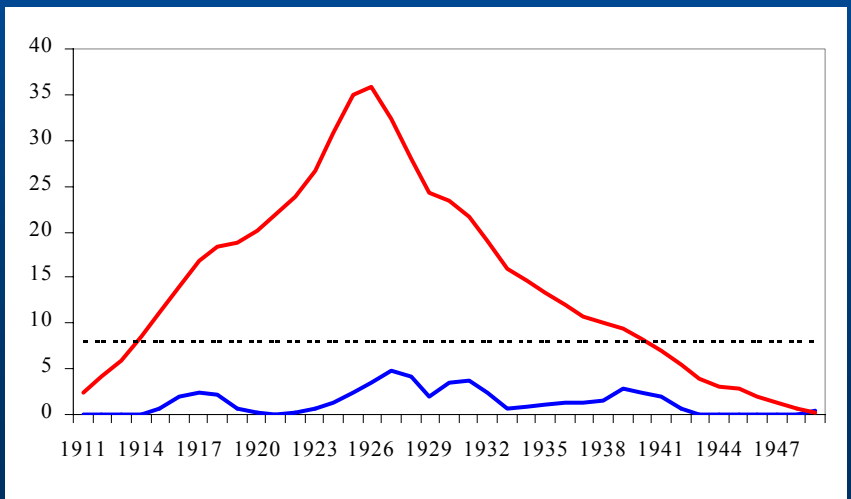
- different climatic conditions

Reference Series - differences



Homogeneity Tests

Alexandersson SNHT



Alexandersson Standart Normal Homogeneity Test

(Single shift test)

Reference series:

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

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

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 μ_1 ? μ_2 .

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

Test statistic:

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

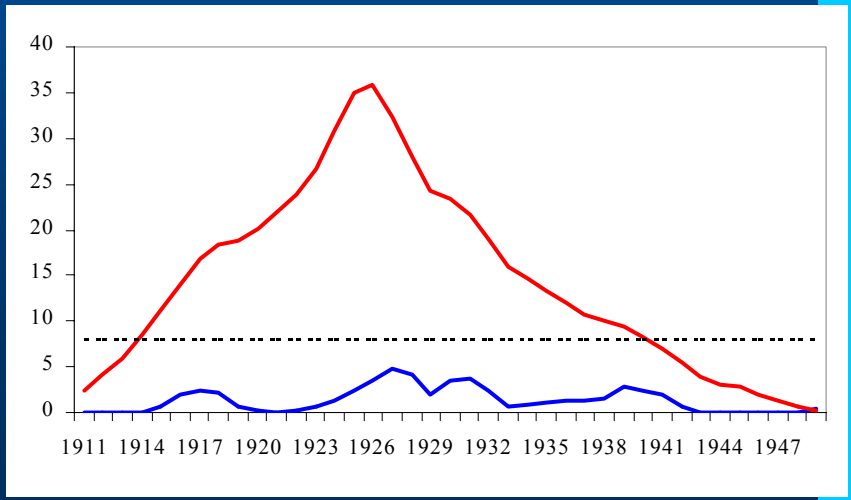
where $\overline{z}_1 = \frac{1}{a}\sum_{i=1}^{a} z_i$, ($\overline{z}_1$? μ_1) ,

$\overline{z}_2 = \frac{1}{(n-a)}\sum_{i=a+1}^{n} z_i$, ($\overline{z}_2$? μ_2) .

Homogeneity Tests

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

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

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

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

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

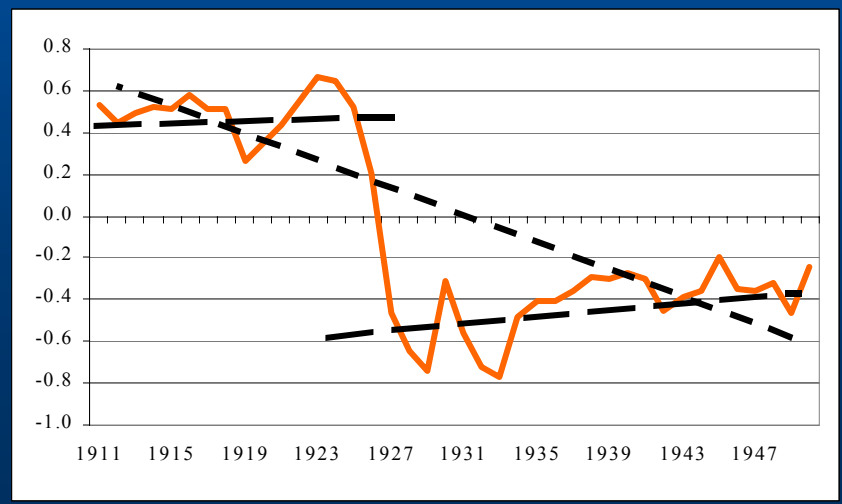
Homogeneity Tests

- Alexandersson SNHT
- Bivariate Test
- Vincent Technique

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 Tests

40 year parts of the series

(one inhomogeneity per 30-40 years)

Homogeneity assessment

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

- Quality control
- Homogenization
- Data Analysis

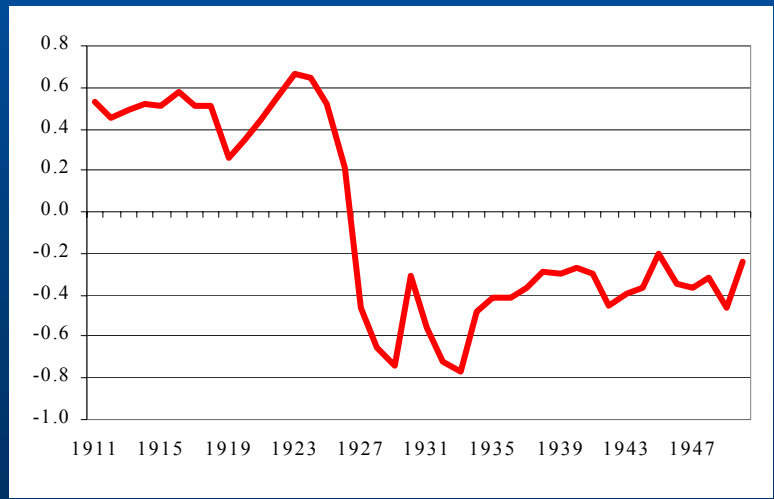
Homogeneity assessment

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		

Homogeneity assessment

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

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		



Adjusting the series

- differences: ± 20 values around inhomogeneity (each month)

- Reference series as
  an average of the best correlated stations

Filling missing values

- linear regression (± 20 values)
- Reference series as
  an average of the best correlated stations

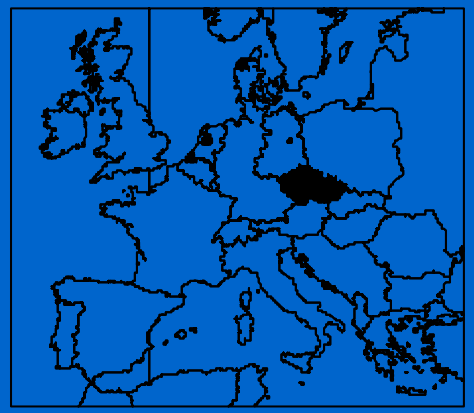
Further remarks

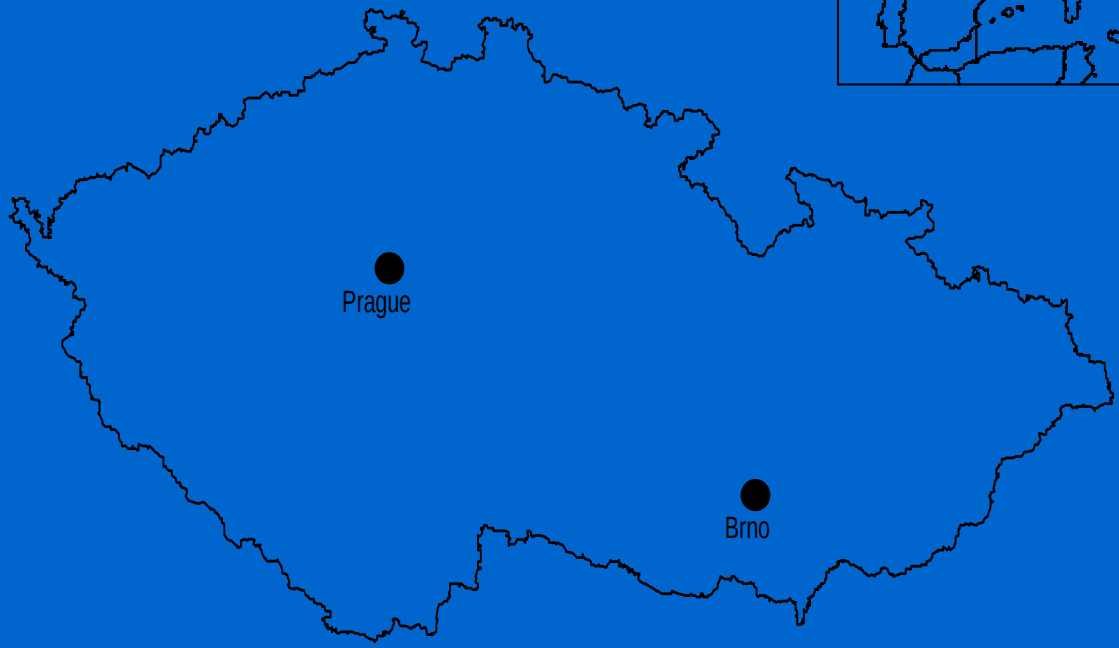
- for creating reference series - used stations outside the Czech Republic
  (measuring in the beginnings of instrumental measurements)

- Several steps – iterations of homogenization(3)

Homogenization
of the series in the Czech Republic

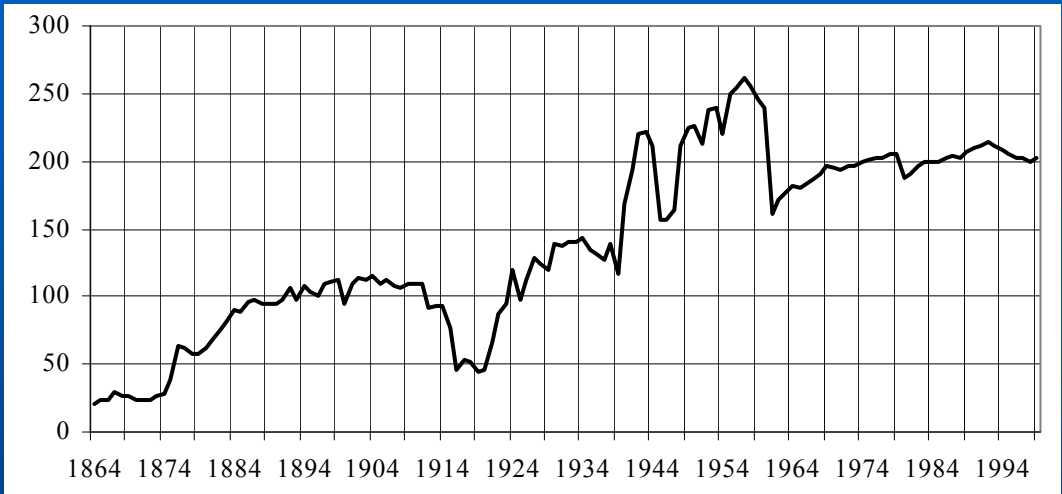
Czech Republic





Number of available stations

Number of climatological stations



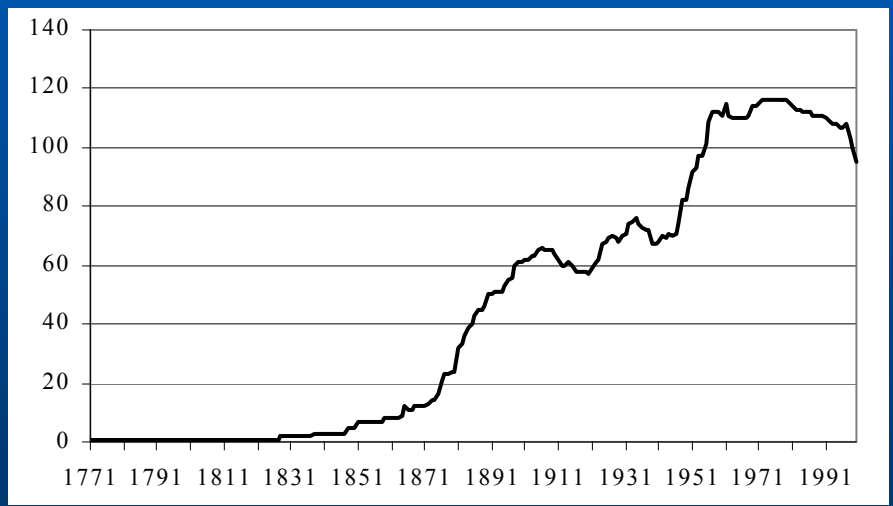
1976
1916-1920
1940-44
1942-44
1945-47
1948-1960
1961

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.

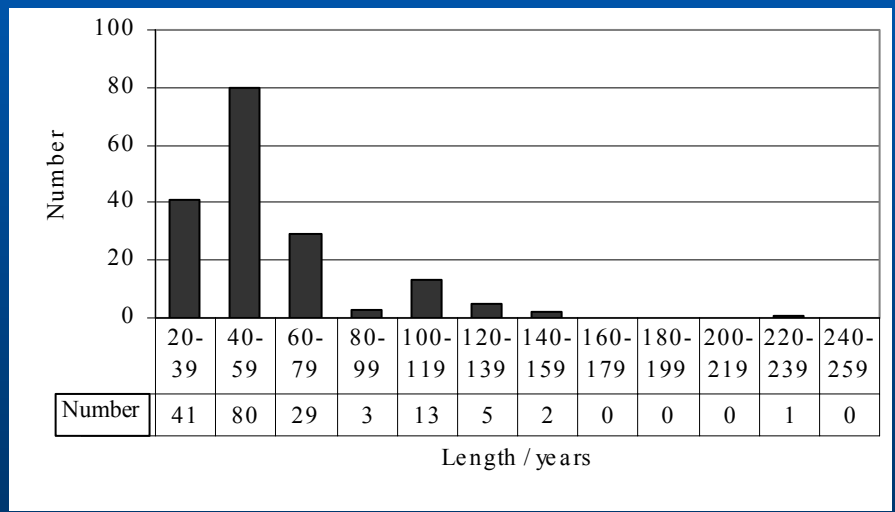
Number of homogenized stations

Number of homogenized stations	174
Average series length	59.1 years
Average minimum distance	13.3 km

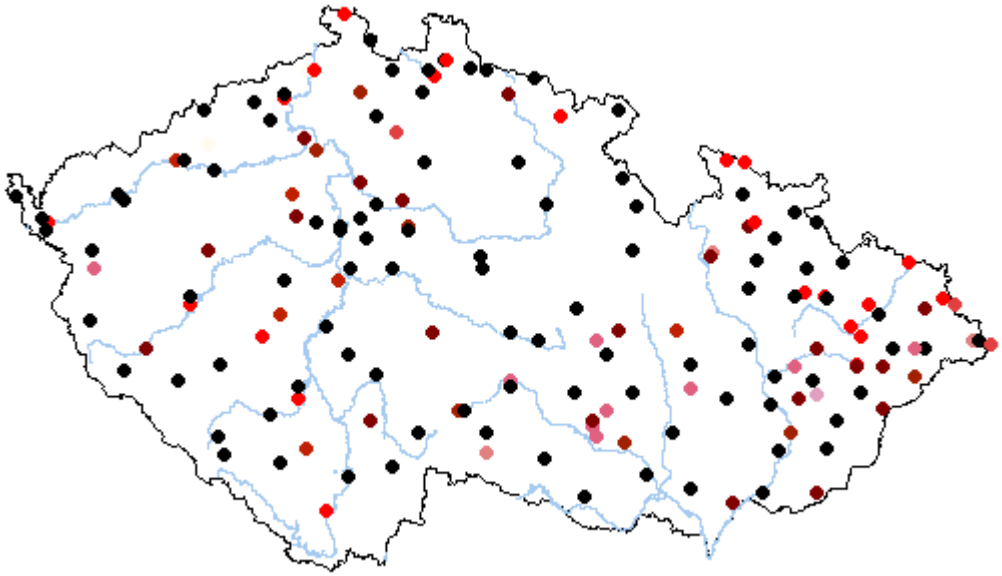
Number of stations for a given year



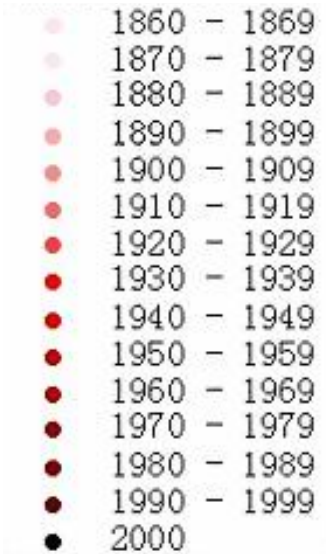
Number of stations for a given length



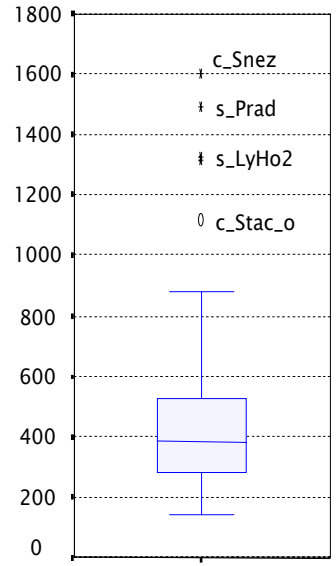
Spatial distribution of the stations



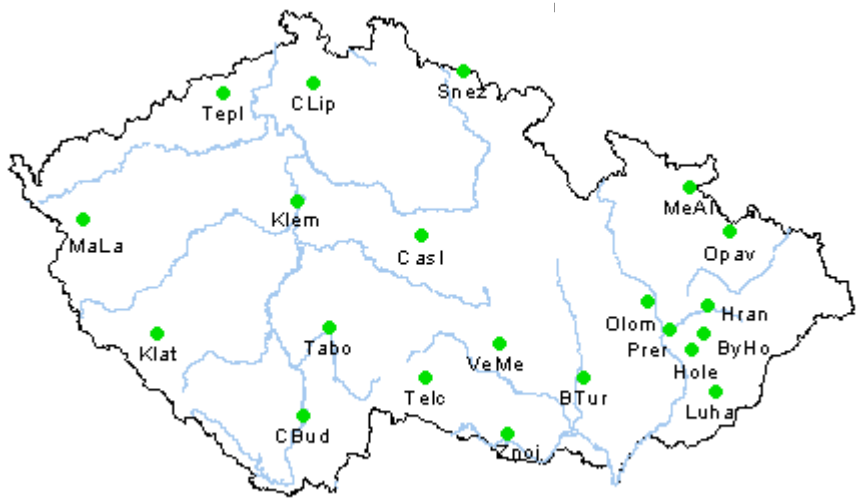
End of measurement



Altitude, metres above see



100 years long series



Homogenization overview

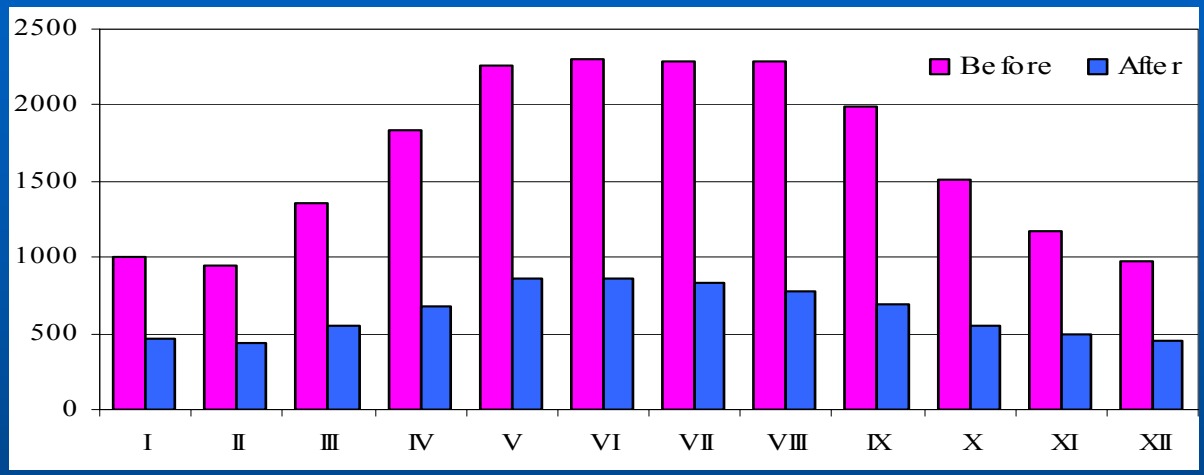
Charakteristic	Data	
	Original	Adjusted
Number of stations	192	174
Number of stations - 40 year parts	348	307
Number of adjustments		231
Number of tested series	40716	35919
Number of signif. inhomogeneities p=0.05	32445	13802
Number of sign. ***inhomogenities*** *per No. of series*	***79.7%***	***38.4%***

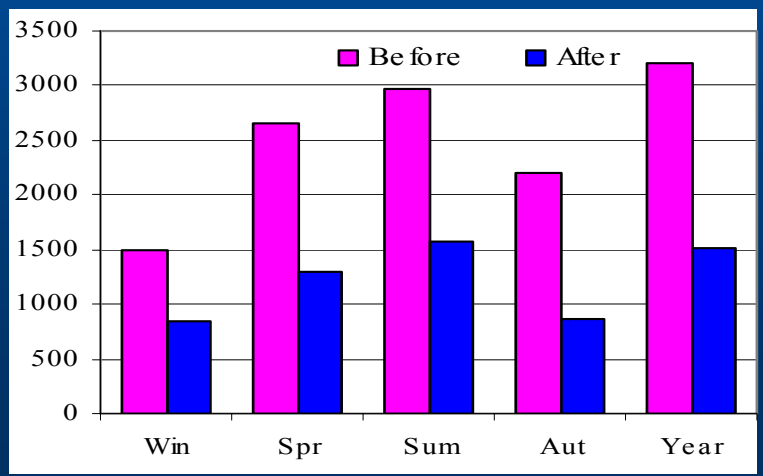
Number of tested series - original data

Tests	Months	Seasons + Year	Reference series	Stations - parts	Number of tests
A	12	5	3	348	17748
B	12	5	3	348	17748
V		5	3	348	5220
Sum					40716

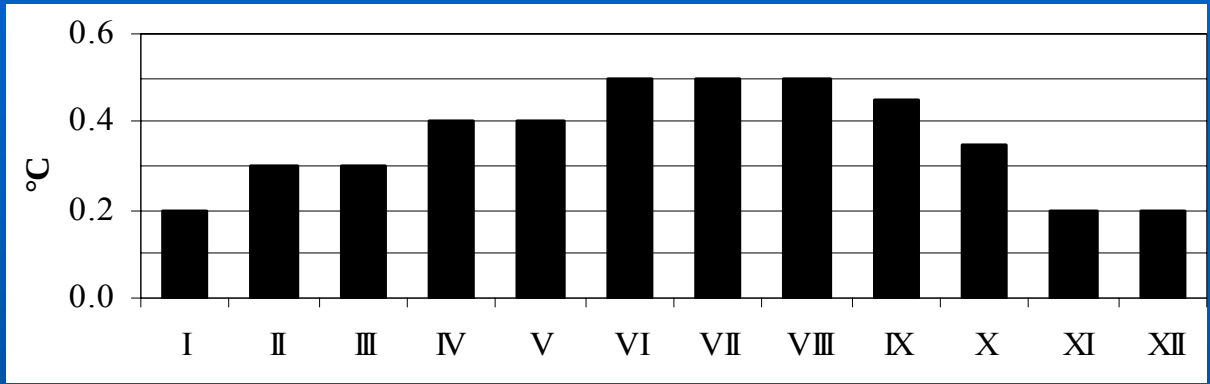
A = Alexandersson SNHT, B = Bivariate Test, V - Vincent method

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

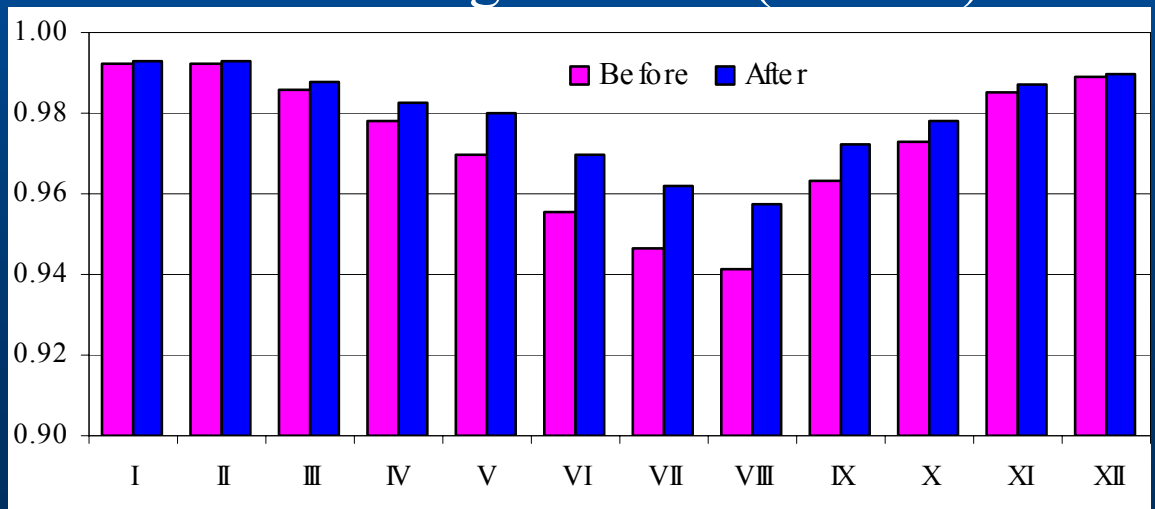




Amount of adjustments for homogenised series (absolute values) - median



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

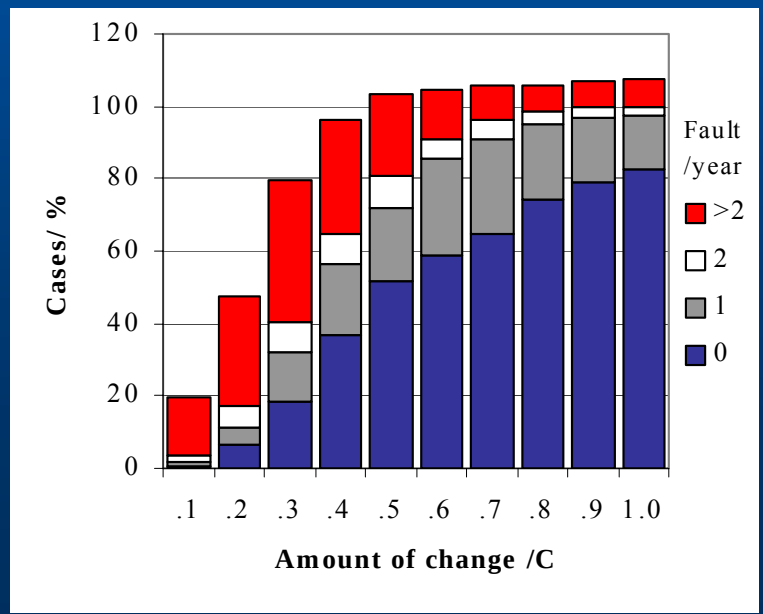


Summer vs. winter inhomogeneities

- changed measuring conditions (relocation, etc.) manifested mainly in summer
- role of active surface: in winter diminished (prevailing circular factors), in summer increased (radiative factors)

Homogenization - conclusions

- **40% inhomogeneous series after homogenization (80%before)**
- **uncertainty in correct inhomogeneities detection (random component of the series; correct inhomogeneity detections for a step lower than 0.5 °C in less than 50% of cases)**



Without complete metadata
- how to increase confidence of the tests

- **Monthly, seasonal, annual averages**
- **Various reference series**
- **Various statistical tests (40 year periods)**
- **Statistical processing of all the test results**
- **Several steps - iterations**

Transition to Automatic Measurements

- Introduced since 1997
- the replacement accomplished within short period (several years)
- only few station with comparative measurements (manual and automatic)

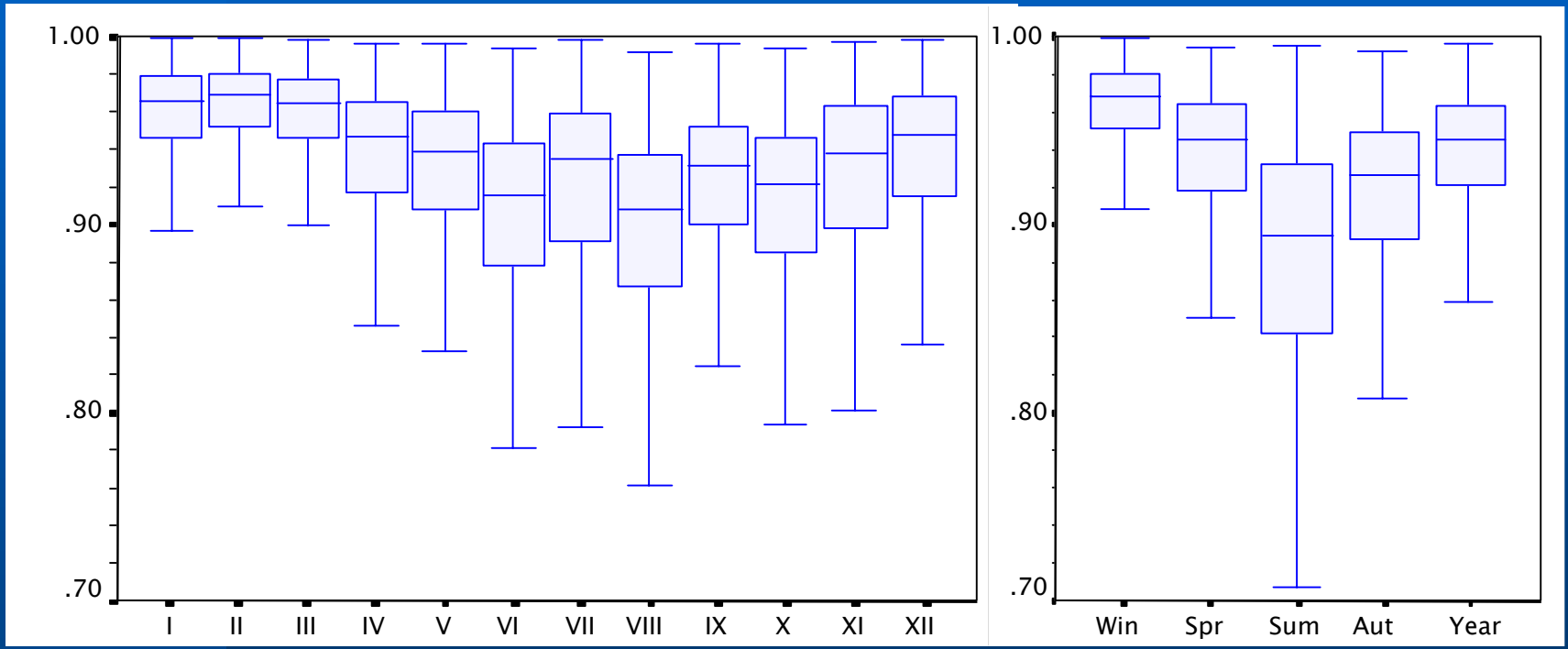
Transition to Automatic Measurements - Consequences

- too early for adjusting data **(but inhomogeneities caused by the transition already detectable)**
- after transition of all stations: no stations available for creating *homogeneous* reference series

→ no way how to assess or adjust inhomogeneities!!!

Data Analysis

Box plot for correlation coefficients



Correlations between averaged series of the Czech Republic and averaged series of stations outside the Czech Republic, 1848-1999

Season	Win	Spr	Sum	Aut	Year
Corr.	**0.97**	**0.93**	**0.94**	**0.93**	**0.96**

Averaged series of the Czech Republic

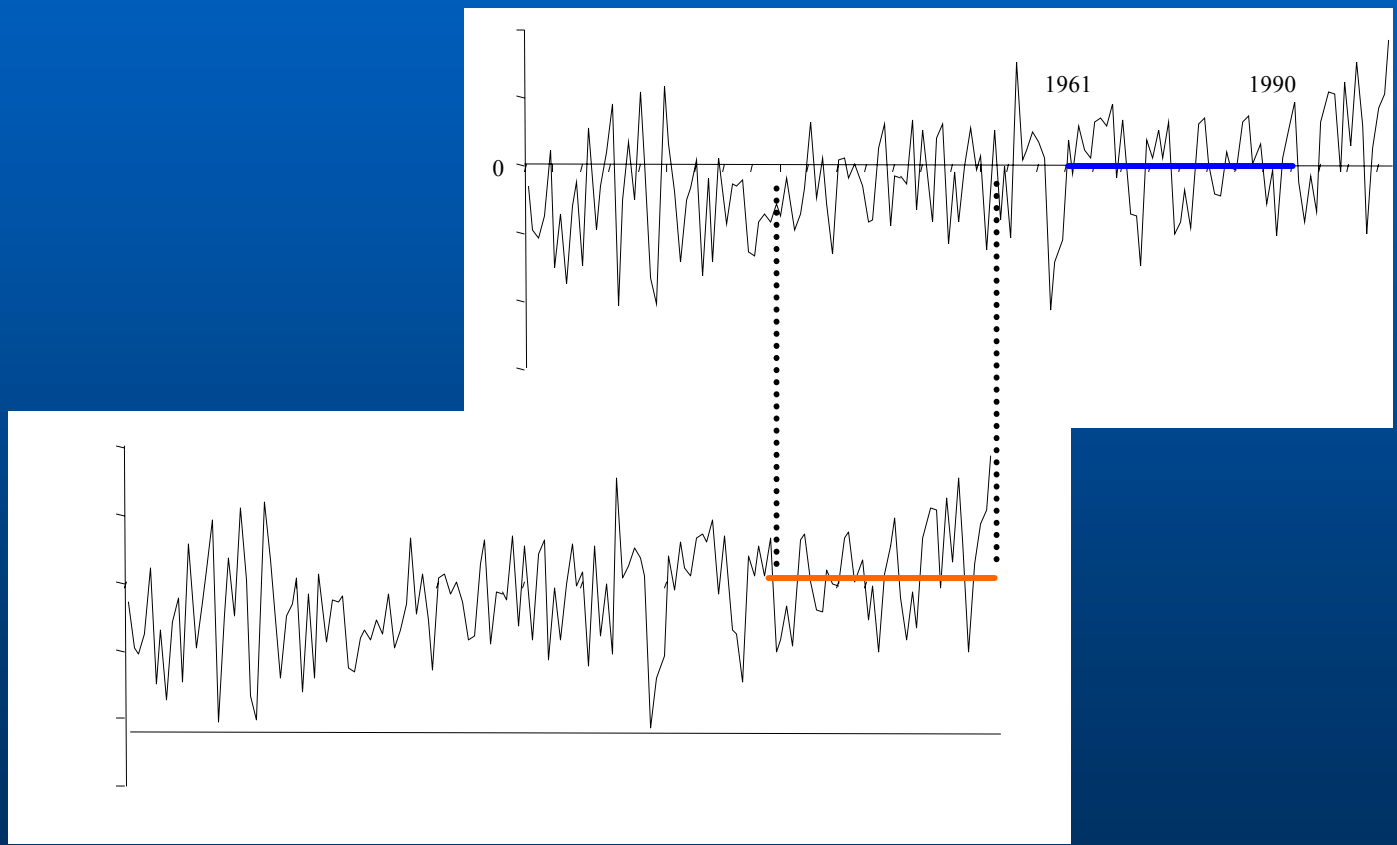
An average from all available stations

- after converting series into anomalies 1961-1990

(so that series measuring in different periods are comparable)

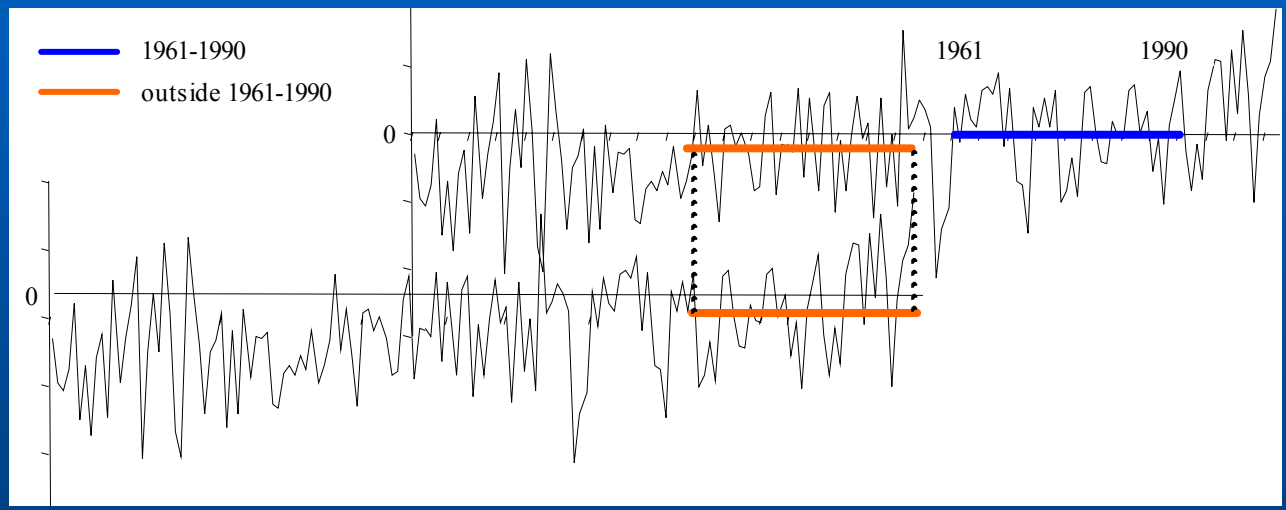
An average from all available stations

- converting series into anomalies



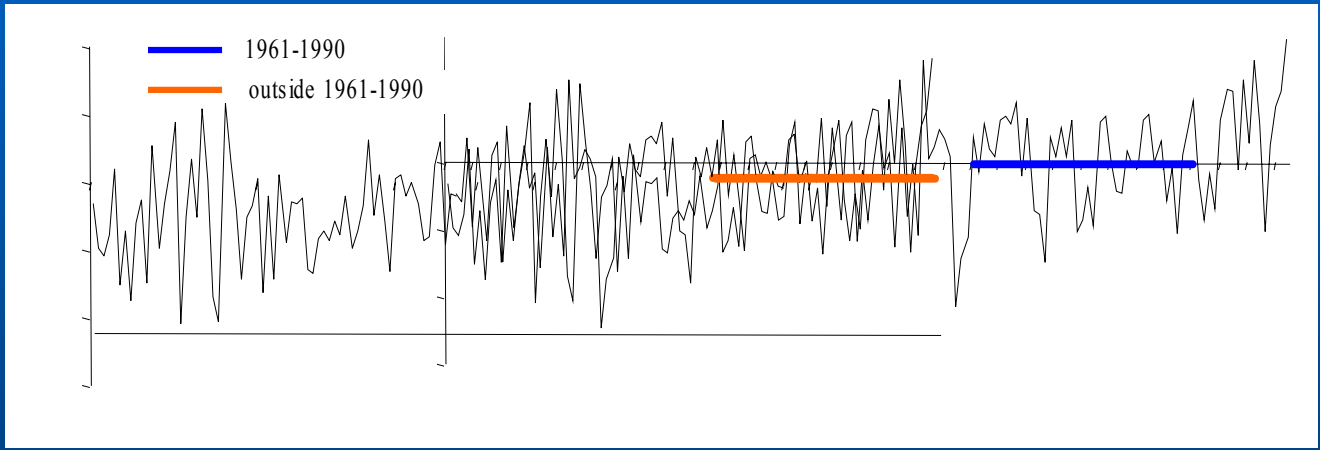
An average from all available stations

- converting series into anomalies



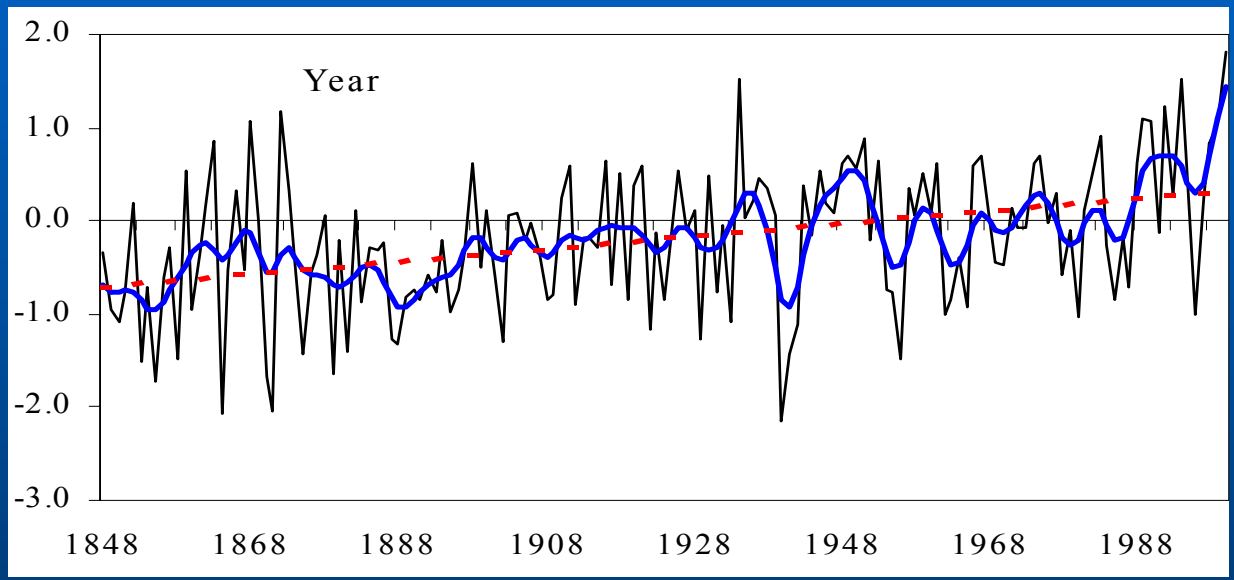
An average from all available stations

- converting series into anomalies



Averaged series of the Czech Republic

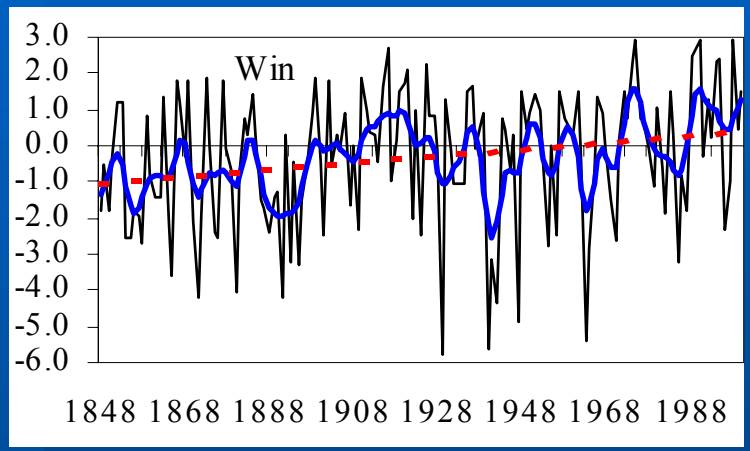
smoothed by 10-year Gaussian low-pass filter.

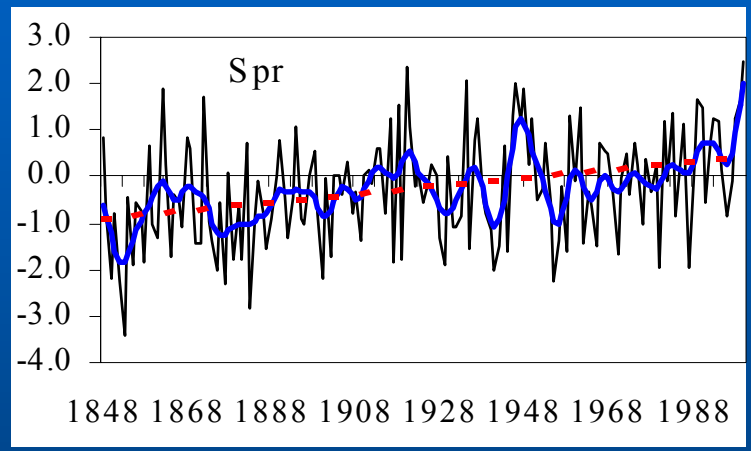


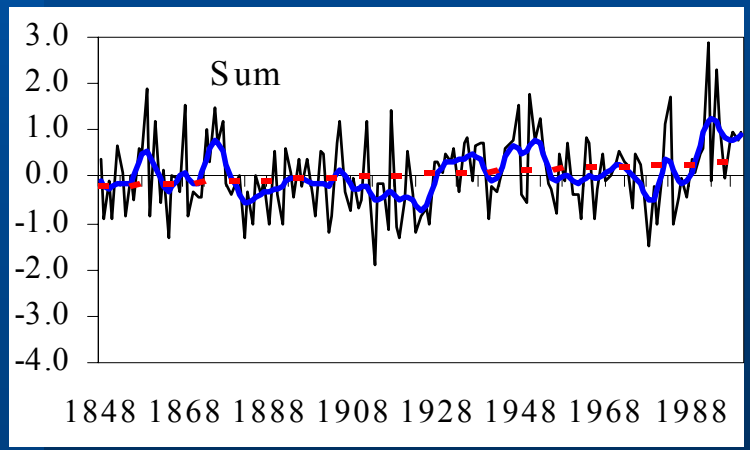
Trend /100 years (°C) 1848-2000.

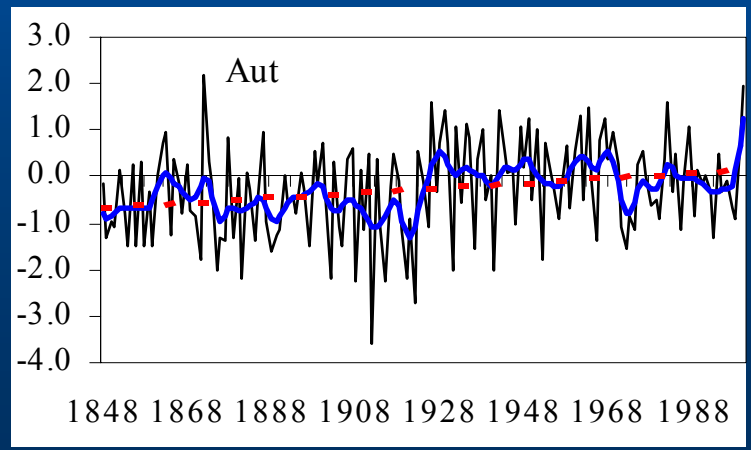
Month	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Trend	**1.17**	0.47	**1.22**	**0.64**	**0.79**	0.13	0.39	**0.56**	0.30	0.22	**1.06**	**1.29**
Season	Win	Spr	Sum	Aut	Year							
Trend	**0.96**	**0.88**	**0.36**	**0.52**	**0.69**							

Averaged series of the Czech Republic

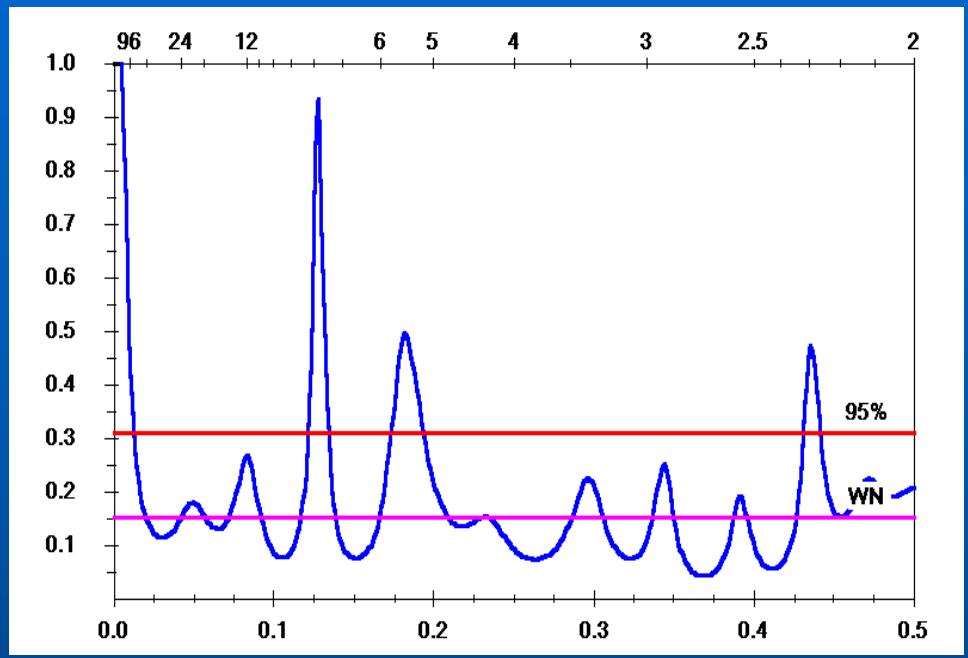




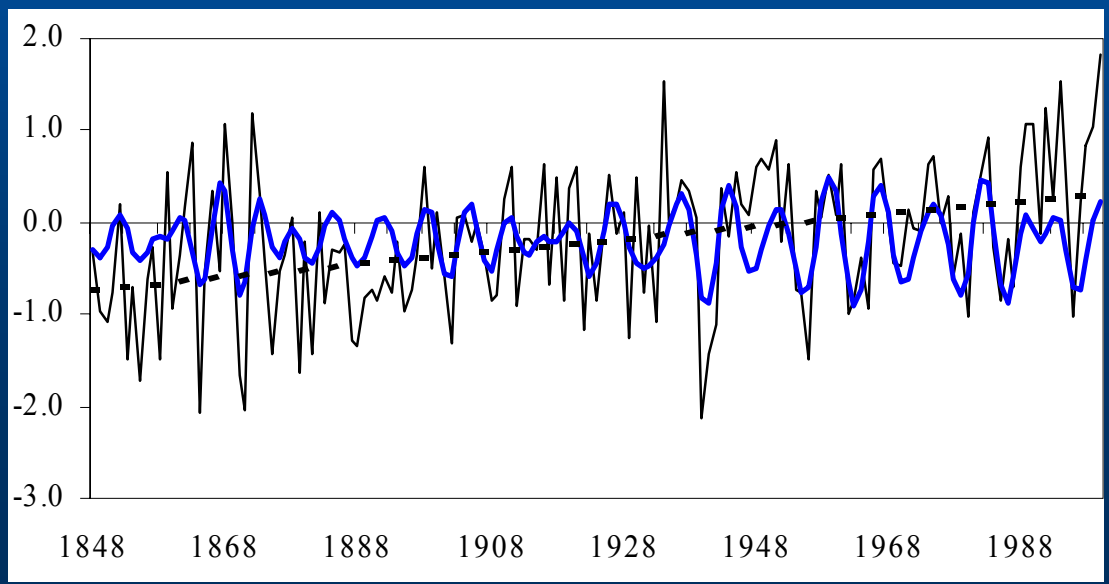




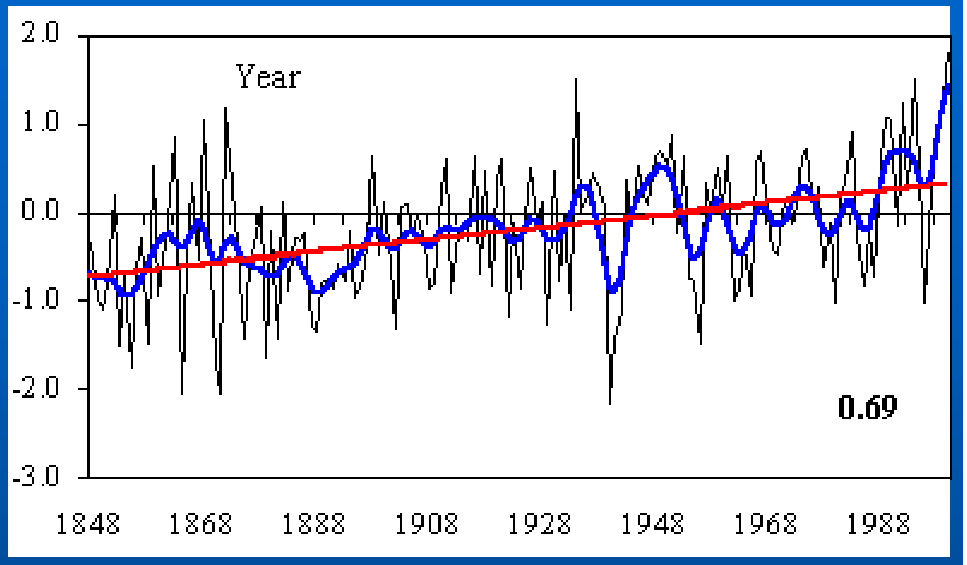
Maximum Entropy Spectral Analysis

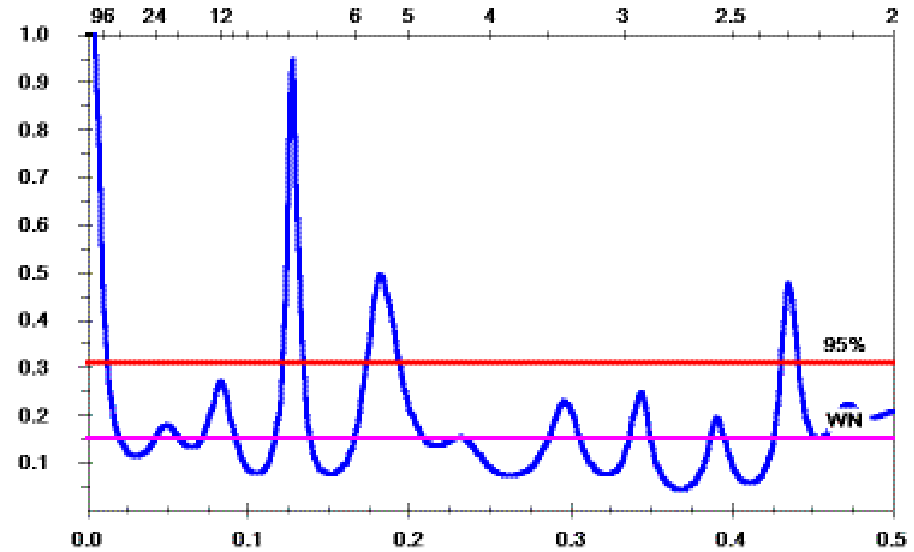


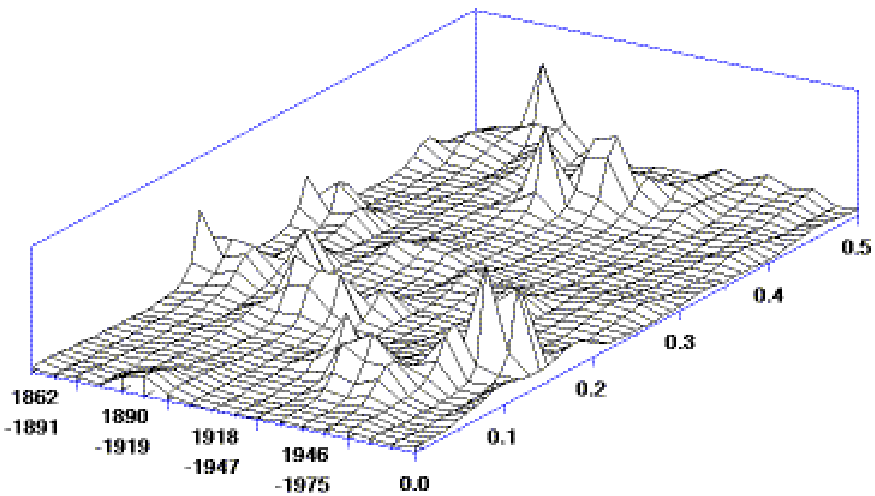
Period
7.8 y.
5.5y.
2.3y.

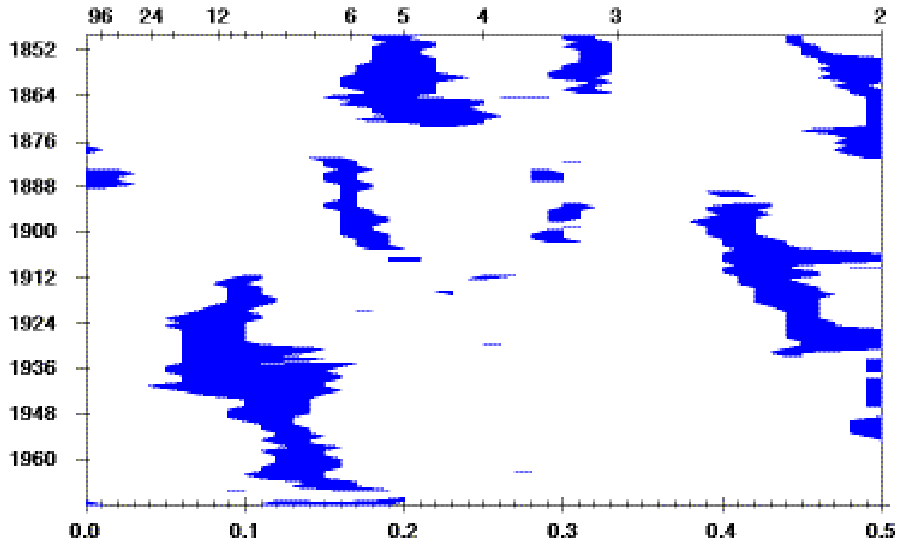


Maximum Entropy Spectral Analysis







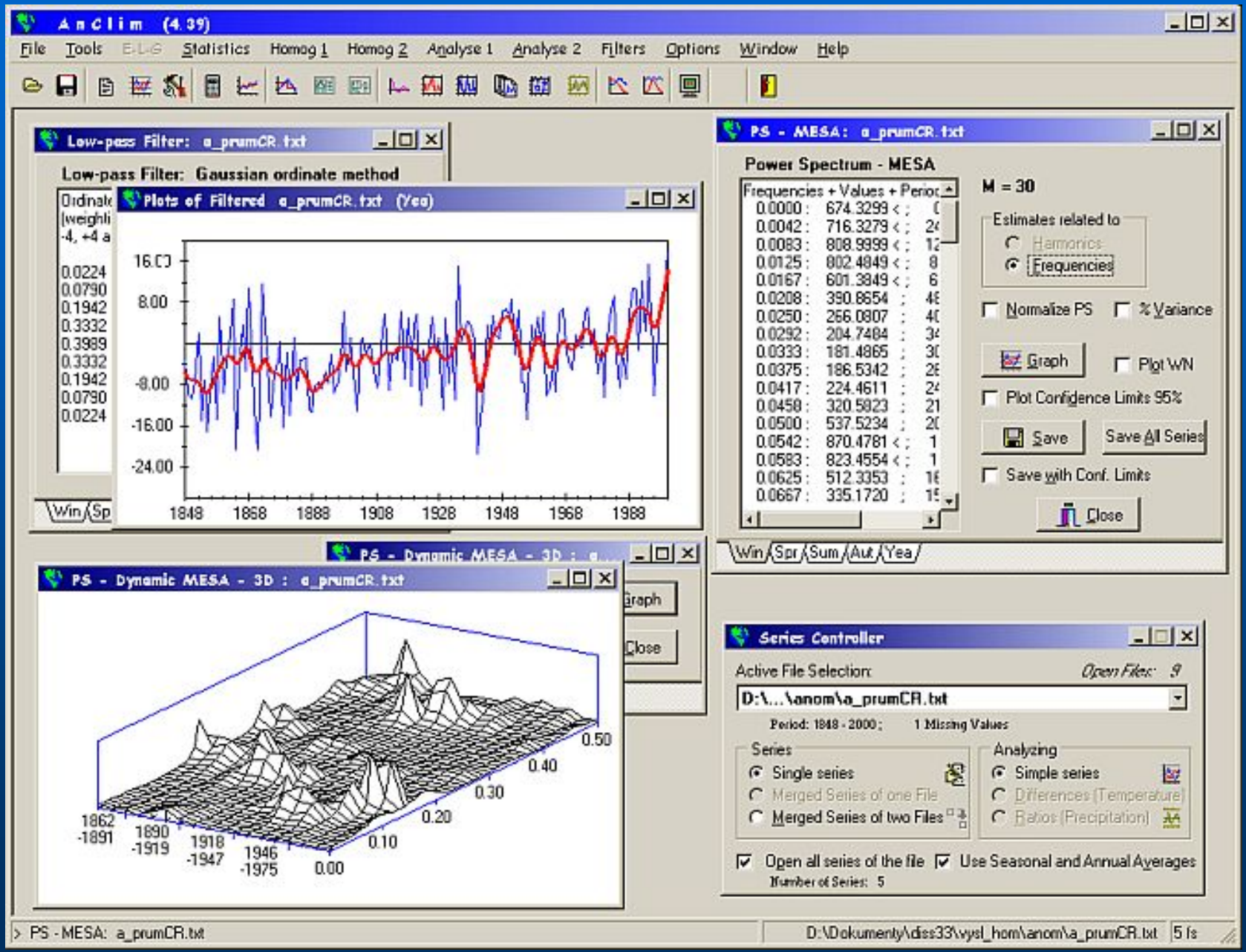


AnClim software

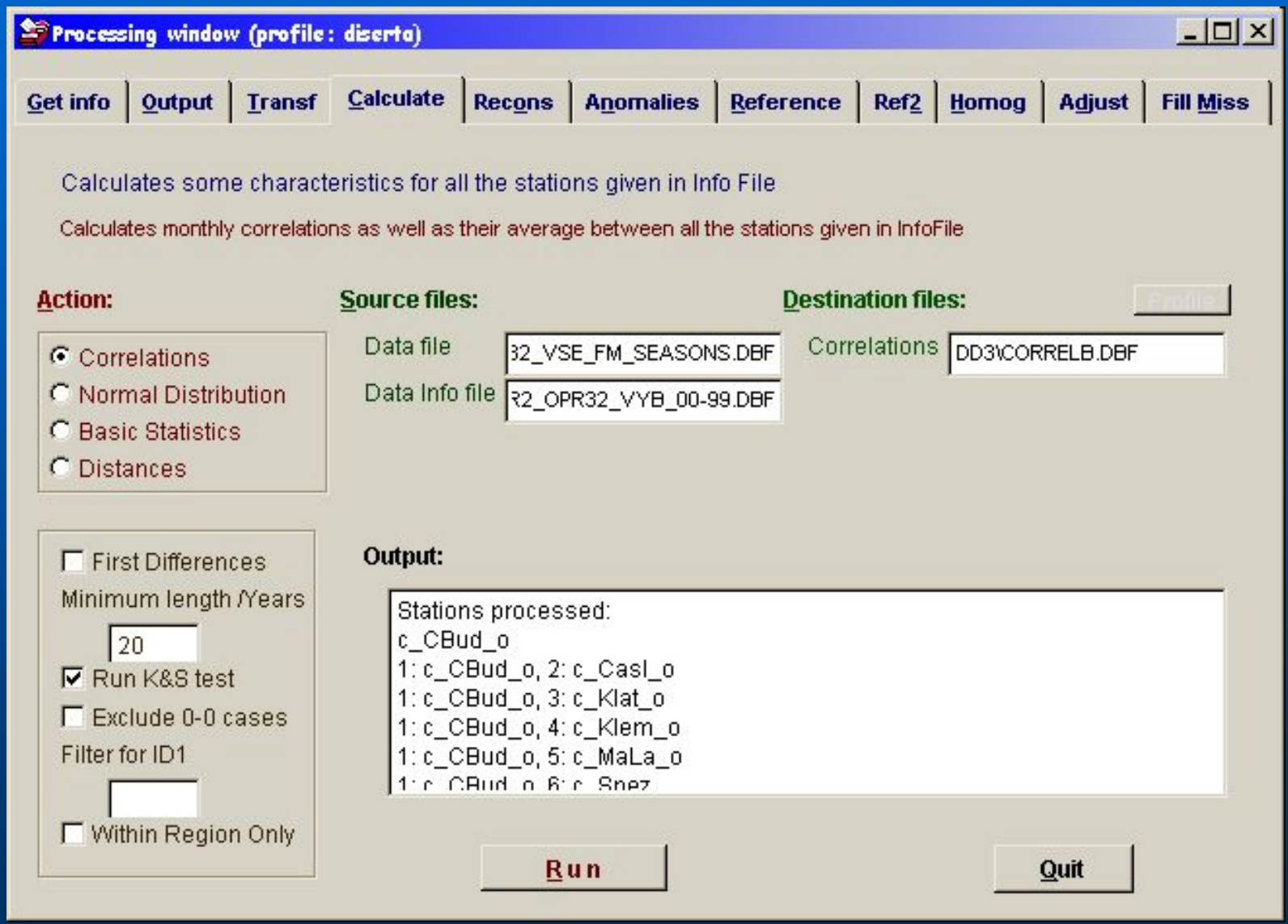
- **1995-2003**
- **Continuous development**
- **Comprehensive tool**
- **For free use**

http://www.sci.muni.cz/~pest

AnClim software



ProcData software



Thank you for your attention