Software solution for data quality control and homogenization of time series

Petr Štěpánek

Czech Hydrometeorological Institute,
regional office Brno

e-mail: petr.stepanek@chmi.cz

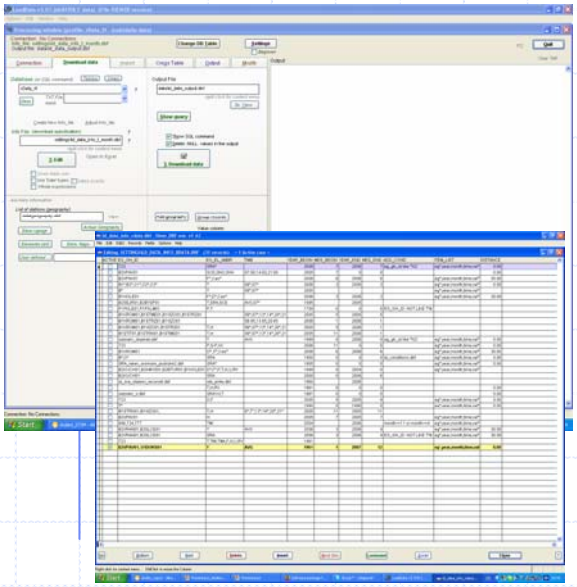
Software package to be prezented

- Originally created for homogeneity testing and time series analysis (trends, cycles, correlation analysis)
- Recently added functions for extreme values analysis (GEV, GPD), RCM outputs validation and correction, multivariate analysis (connection with R software)

http://www.climahom.eu

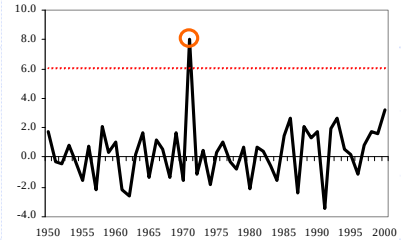
Download data from database (e.g. Oracle)

(LoadData)



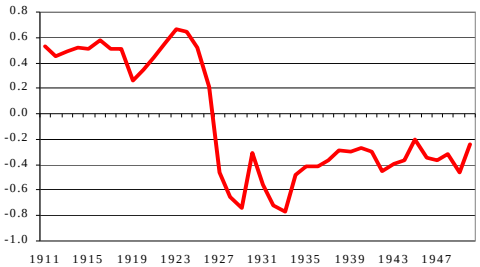
Quality control

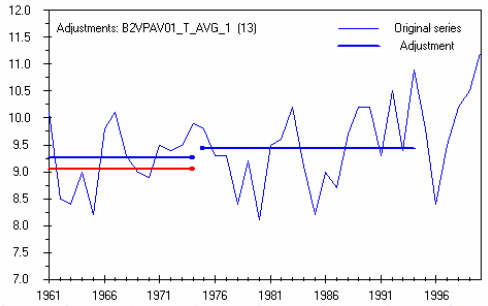
(ProClimDB)



Homogenization

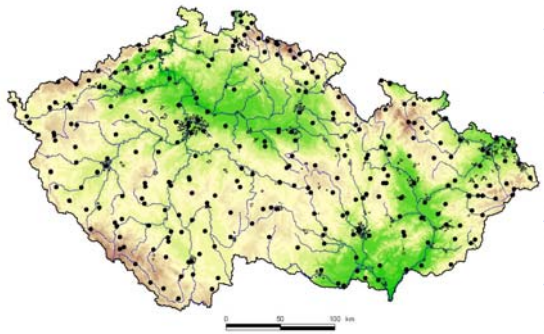
(ProClimDB/AnClim)

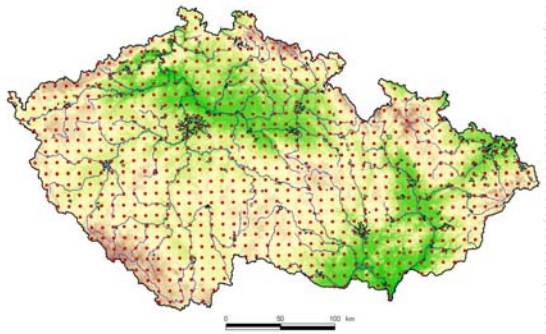




„Technical“ series and grid points calculation

(ProClimDB)





Statistical analysis

...

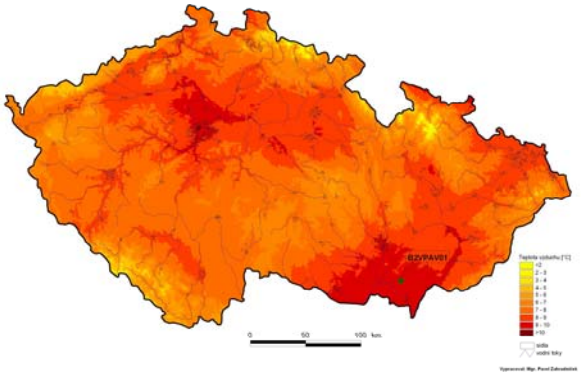
Validation of RCM outputs

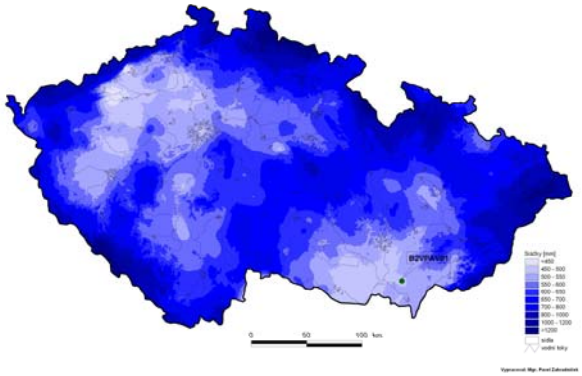
Extreme value analysis

Correction of RCM outputs

Spatial analysis

(connection ProClimDB - ArcView)





Further tools:

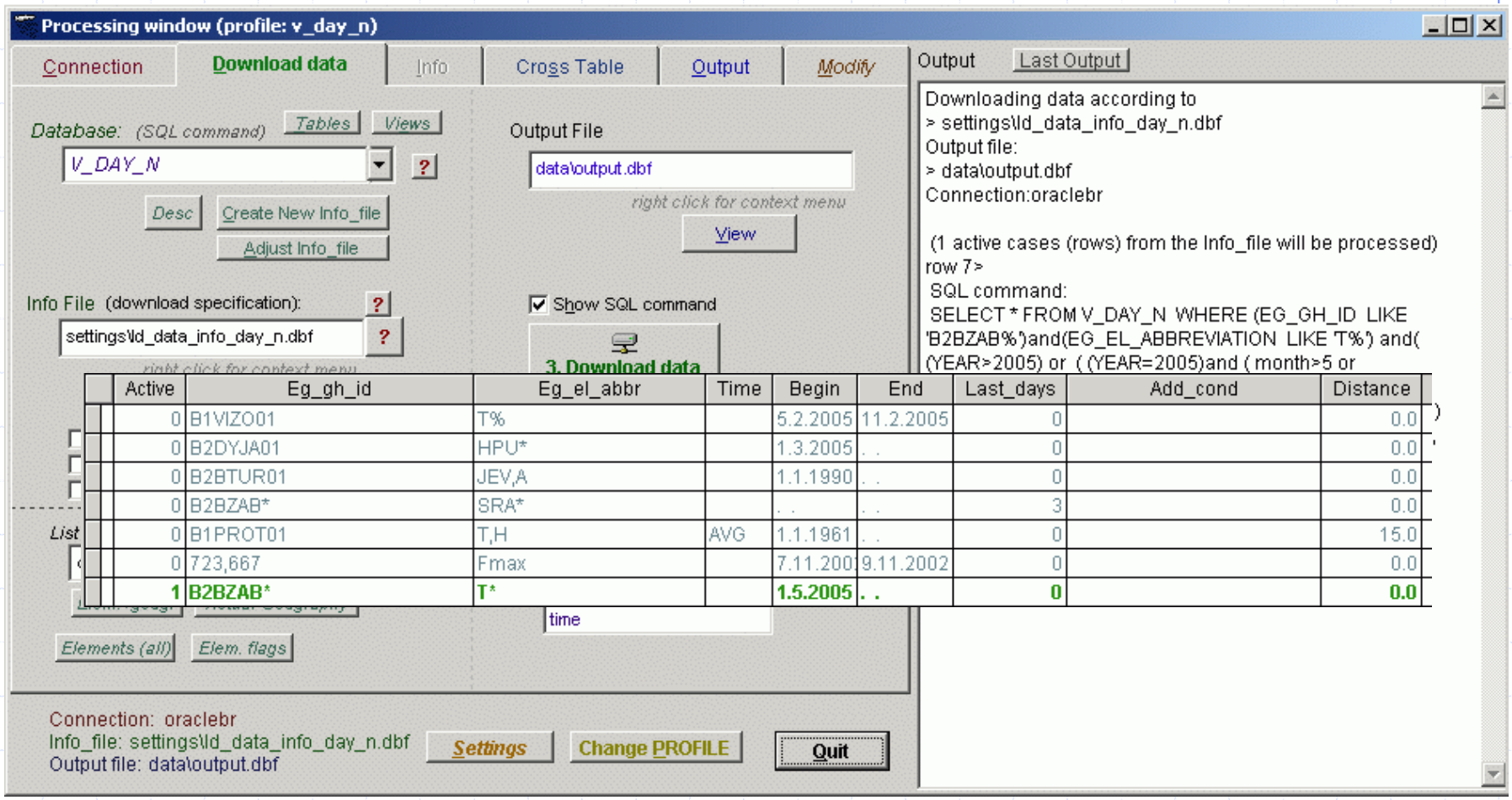
(connection ProClimDB - R)

Software Package for Processing Climatological Data

- 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, analysis ...)
- AnClim software for homogeneity testing

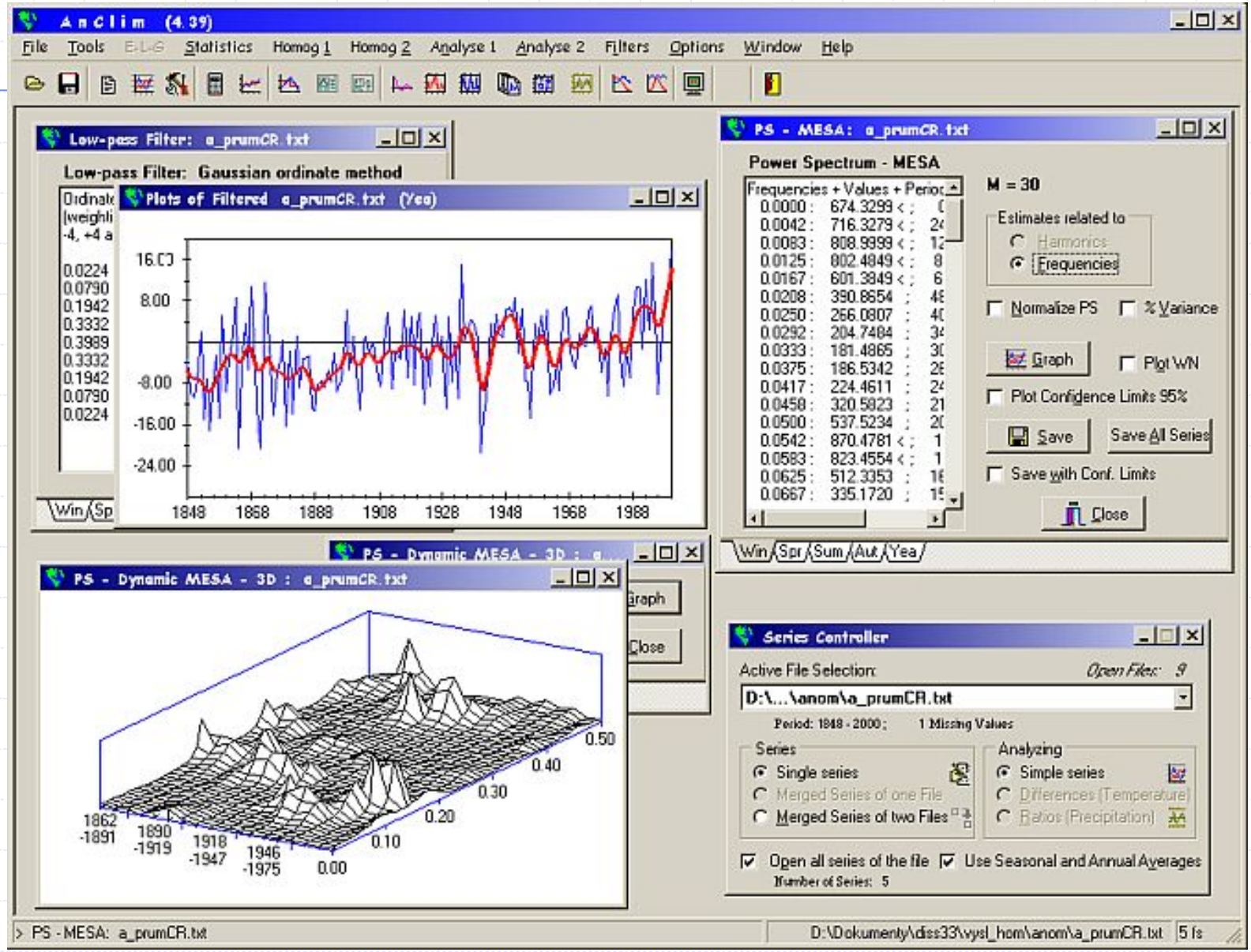
http://www.climahom.eu

LoadData software, SQL commands generator (based on given *Database Table* and *Info_file*)



AnClim software, TXT files (each station has its own text file)

Monthly (seasonal, annual) or **daily** data processing
conveniant for learning of statistical methods in climatology (tutorials)



Examples of Data formats – AnClim, monthly data

Edit precip_bohemia.txt

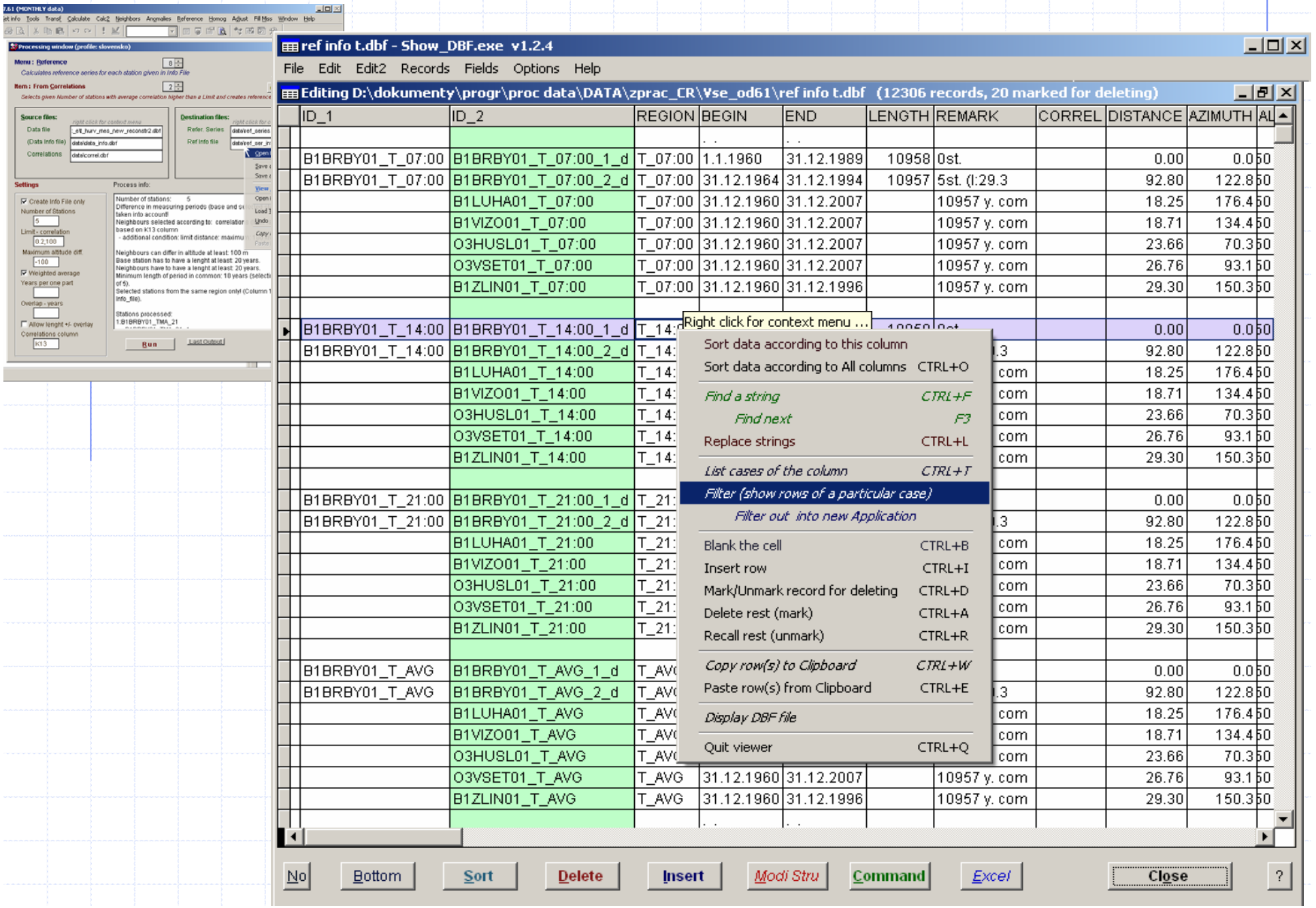
1876	19.0	78.0	76.0	37.0	42.0	69.0	55.0	60.0	91.0	16.0	39.0	50.0	632.0
1877	56.0	76.0	46.0	30.0	50.0	43.0	100.0	62.0	53.0	24.0	32.0	41.0	613.0
1878	48.0	19.0	82.0	52.0	68.0	49.0	69.0	88.0	46.0	49.0	37.0	30.0	637.0
1879	29.0	52.0	33.0	49.0	70.0	114.0	86.0	78.0	40.0	36.0	63.0	29.0	679.0
1880	32.0	22.0	34.0	48.0	114.0	82.0	93.0	117.0	53.0	85.0	41.0	98.0	819.0
1881	13.0	20.0	72.0	28.0	88.0	93.0	79.0	98.0	77.0	52.0	18.0	24.0	662.0
1882	14.0	25.0	23.0	55.0	84.0	87.0	139.0	109.0	84.0	52.0	85.0	73.0	830.0
1883	31.0	19.0	33.0	18.0	33.0	123.0	110.0	50.0	66.0	36.0	24.0	59.0	602.0
1884	44.0	12.0	40.0	27.0	32.0	121.0	84.0	60.0	36.0	104.0	36.0	61.0	657.0
1885	11.0	16.0	48.0	14.0	51.0	36.0	76.0	65.0	89.0	48.0	55.0	51.0	560.0
1886	35.0	12.0	52.0	59.0	53.0	147.0	93.0	51.0	42.0	42.0	31.0	83.0	700.0
1887	10.0	18.0	56.0	29.0	117.0	39.0	45.0	74.0	34.0	15.0	60.0	44.0	541.0
1888	38.0	47.0	80.0	56.0	28.0	104.0	95.0	123.0	67.0	65.0	32.0	15.0	750.0
1889	20.0	62.0	41.0	59.0	68.0	71.0	103.0	53.0	80.0	82.0	18.0	20.0	677.0
1890	65.0	7.0	23.0	89.0	74.0	105.0	92.0	145.0	111.0	48.0	85.0	6.0	850.0
1891	49.0	12.0	44.0	47.0	50.0	95.0	136.0	72.0	39.0	16.0	41.0	64.0	665.0
1892	52.0	47.0	34.0	48.0	52.0	97.0	71.0	27.0	85.0	62.0	12.0	38.0	625.0
1893	48.0	53.0	39.0	2.0	71.0	47.0	80.0	41.0	52.0	56.0	50.0	18.0	557.0
1894	7.0	47.0	63.0	64.0	99.0	84.0	98.0	82.0	73.0	113.0	15.0	30.0	775.0
1895	58.0	32.0	54.0	50.0	95.0	80.0	84.0	79.0	31.0	51.0	46.0	78.0	738.0
1896	34.0	14.0	56.0	55.0	96.0	103.0	92.0	98.0	75.0	28.0	29.0	24.0	704.0
1897	38.0	48.0	67.0	39.0	125.0	50.0	182.0	101.0	74.0	26.0	26.0	27.0	803.0
1898	42.0	45.0	46.0	73.0	80.0	63.0	61.0	51.0	43.0	61.0	29.0	36.0	630.0
1899	50.0	18.0	15.0	78.0	136.0	62.0	102.0	65.0	127.0	22.0	28.0	55.0	758.0
1900	91.0	81.0	81.0	51.0	52.0	95.0	99.0	40.0	31.0	64.0	55.0	64.0	804.0
1901	34.0	38.0	56.0	67.0	53.0	74.0	105.0	74.0	37.0	67.0	46.0	51.0	702.0
1902	37.0	35.0	49.0	47.0	74.0	82.0	92.0	75.0	64.0	50.0	39.0	44.0	688.0
1903	31.0	34.0	17.0	63.0	49.0	61.0	115.0	93.0	52.0	67.0	89.0	20.0	691.0
1904	21.0	57.0	31.0	68.0	61.0	39.0	23.0	16.0	54.0	66.0	74.0	35.0	545.0
1905	46.0	40.0	49.0	54.0	61.0	50.0	86.0	105.0	53.0	68.0	56.0	32.0	700.0
1906	45.0	29.0	70.0	23.0	85.0	77.0	71.0	61.0	129.0	31.0	33.0	45.0	699.0
1907	59.0	27.0	52.0	59.0	48.0	68.0	137.0	62.0	37.0	31.0	27.0	67.0	674.0
1908	32.0	56.0	28.0	51.0	65.0	66.0	69.0	84.0	49.0	1.0	30.0	12.0	543.0
1909	32.0	76.0	28.0	58.0	50.0	100.0	138.0	45.0	58.0	24.0	52.0	71.0	732.0
1910	34.0	28.0	22.0	44.0	86.0	121.0	93.0	95.0	105.0	27.0	104.0	40.0	799.0
1911	31.0	51.0	35.0	32.0	74.0	32.0	28.0	35.0	58.0	29.0	36.0	64.0	505.0

Examples of Data formats – AnClim, daily data

Data of B1BYSH01_T_07_00_3.txt

Year	Day	1	2	3	4	5	6	7	8	9	10	11	12
1917	11	0.0	-9.8	-4.6	2.0	9.2	16.2	15.0	15.4	11.8	8.4	5.3	-2.5
1917	12	-4.6	-7.6	-4.6	8.4	11.2	17.0	13.8	17.2	18.9	5.2	3.2	0.0
1917	13	-2.0	-6.8	2.4	10.0	11.0	17.8	14.4	19.2	13.0	7.7	1.2	-2.0
1917	14	-6.0	-13.2	0.8	4.0	12.4	17.6	18.0	24.6	9.0	12.3	0.2	-2.0
1917	15	-0.6	-6.8	3.4	4.6	14.2	15.4	18.0	18.7	8.6	8.7	1.0	2.0
1917	16	0.0	-10.4	-6.8	6.2	15.8	18.6	21.4	15.4	7.3	6.0	0.5	-1.2
1917	17	8.4	-11.6	-3.8	3.2	13.8	19.2	19.4	14.0	11.4	4.6	-0.8	-0.5
1917	18	-2.6	-2.6	2.2	3.2	12.2	21.8	16.4	15.2	17.3	12.8	0.2	-0.6
1917	19	-2.6	-1.6	2.2	0.0	21.0	22.4	20.0	15.4	14.6	6.4	4.4	-3.8
1917	20	-9.0	0.0	3.4	0.4	19.0	25.0	21.0	21.4	18.0	4.3	4.6	-9.0
1917	21	-14.6	-1.2	1.0	3.0	12.4	26.0	17.0	17.8	15.0	9.1	5.8	-10.8
1917	22	-13.8	-2.8	-2.8	2.4	6.6	25.0	14.2	15.6	9.3	5.6	1.8	-3.2
1917	23	-11.6	-7.0	-2.8	0.1	11.4	16.2	14.8	14.2	6.6	6.6	1.0	-4.6
1917	24	-10.4	-11.6	-1.8	1.4	15.2	12.6	16.2	23.3	8.4	7.6	6.5	-8.4
1917	25	-10.4	-0.1	-1.6	2.2	17.4	11.2	17.0	17.2	7.4	4.1	7.8	-6.0
1917	26	-11.8	0.4	-1.2	1.6	13.8	20.0	17.8	15.2	7.7	2.8	1.4	-5.0
1917	27	-11.0	0.6	3.0	5.2	11.0	20.0	19.0	18.6	8.8	6.1	-2.4	-7.1
1917	28	-13.0	-1.0	0.0	5.0	18.6	17.2	17.2	13.4	14.8	11.8	5.8	-7.2
1917	29	-11.0	missing	-1.0	4.8	21.0	22.0	23.0	21.6	11.2	14.6	8.4	-2.4
1917	30	-11.2	missing	3.8	11.4	20.0	23.4	20.6	12.4	10.2	4.2	5.6	-2.0
1917	31	-8.6	missing	4.6	missing	23.6	missing	24.0	14.2	missing	7.8	missing	-5.0
1918	1	-5.6	-3.2	4.0	3.6	9.4	12.3	12.2	13.0	13.2	8.3	6.0	-2.8
1918	2	-7.4	-1.0	7.0	12.6	7.6	9.8	11.8	17.0	13.4	7.0	4.4	-3.6
1918	3	-10.8	-1.8	10.0	8.4	9.4	9.8	15.0	16.8	11.8	4.1	4.4	-6.0
1918	4	-13.6	-2.6	6.3	5.1	11.6	7.0	13.4	17.3	10.8	5.4	6.3	0.0

ProClimDB software



ProcData software, only one Data file, accompanied by Info_file

database processing

Processing window (profile: slovensko)

Menu : Reference

Calculates reference series for each station giv

Item : From Correlations

Selects given Number of stations with average corr

Source files:

right click for context menu

Data file: :_e\t_hurv_mes_new_reconstr2

(Data Info file): data\data_info.dbf

ID	EG_EL_ABBR	YEAR	DAY	TIME	N1	N2	N3	N4	N5	N6	N7	N8
B1BYSH01_SCE_07:00	SCE	2006	24	07:00	30.00	10.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	25	07:00	28.00	10.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	26	07:00	28.00	12.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	27	07:00	28.00	9.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	28	07:00	28.00	9.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	29	07:00	28.00	-999.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	30	07:00	28.00	-999.00	0.00	0.00	0.00	0.00	0.00	-999.00
B1BYSH01_SCE_07:00	SCE	2006	31	07:00	27.00	-999.00	0.00	-999.00	0.00	-999.00	0.00	-999.00
B1BYSH01_SNO_07:00	SNO	1961	1	07:00	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	2	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	3	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	4	07:00	0.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	5	07:00	0.00	5.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	6	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	7	07:00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
B1BYSH01_SNO_07:00	SNO	1961	8	07:00	1.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00
									0.00	0.00	0.00	0.00

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

Allow length +/- overlay

Correlations column: K13

Run | Last Output | **Quit**

Data formats - ProClimDB

- DBF files (the only DBF file for data + Info file)
- Macro in MS-Excel to load TXT,XLS,... files and to create a DBF data file
- function in ProClimDB to import from TXT,DBF files / export to TXT,... files
- **Monthly** (seasonal, annual) or **daily** (even individual time) data processing

Examples of Data formats – ProClimDB, **monthly data**

Id	Year	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	Remark
11801_RV_07:00	1961	88.0	89.0	86.0	74.0	81.0	80.0	75.0	72.0	67.0	67.0	76.0	73.0	
11801_RV_07:00	1962	87.0	81.0	79.0	68.0	75.0	68.0	70.0	78.0	80.0	87.0	89.0	87.0	
11801_RV_07:00	1963	83.0	86.0	84.0	80.0	84.0	79.0	74.0	80.0	84.0	89.0	82.0	87.0	
11801_RV_07:00	1964	85.0	78.0	84.0	75.0	77.0	79.0	80.0	83.0	83.0	87.0	89.0	92.0	
11801_RV_07:00	1965	91.0	88.0	87.0	86.0	81.0	82.0	82.0	83.0	85.0	92.0	86.0	87.0	
11801_RV_07:00	1966	87.0	86.0	88.0	84.0	77.0	80.0	85.0	88.0	90.0	88.0	89.0	88.0	
11801_RV_07:00	1967	86.0	88.0	85.0	83.0	75.0	80.0	78.0	82.0	90.0	88.0	90.0	87.0	
11801_RV_07:00	1968	87.0	91.0	82.0	75.0	74.0	73.0	77.0	87.0	89.0	92.0	90.0	88.0	
11801_RV_07:00	1969	89.0	88.0	89.0	79.0	74.0	86.0	81.0	86.0	88.0	88.0	86.0	93.0	
11801_RV_07:00	1970	90.0	92.0	89.0	84.0	78.0	78.0	84.0	88.0	89.0	93.0	87.0	91.0	
11801_RV_07:00	1971	90.0	92.0	87.0	78.0	80.0	82.0	80.0	80.0	91.0	91.0	90.0	92.0	
11801_RV_07:00	1972	88.0	86.0	75.0	85.0	84.0	78.0	85.0	86.0	88.0	88.0	87.0	87.0	
11801_RV_07:00	1973	85.0	90.0	82.0	79.0	75.0	79.0	82.0	81.0	85.0	85.0	81.0	82.0	

(ID, Year, Months in columns: very useful format > easy processing of individual months)

Id	Year	Pav_4h	Pdsav_4h	Pdssdv_4h	Pf20_4h	Pf40_4h	Pf50_4h	Pf60_4h	Pf80_4h	Pf90_4h
ADAMCLISI	1961	1.221	6.886	6.355	0.957	0.880	0.814	0.756	0.542	0.3
ADAMCLISI	1962	0.966	6.383	6.149	0.944	0.861	0.762	0.729	0.489	0.3
ADAMCLISI	1963	1.079	6.522	6.306	0.950	0.878	0.804	0.737	0.545	0.3
ADAMCLISI	1964	1.051	6.756	5.713	0.936	0.884	0.835	0.772	0.575	0.3
ADAMCLISI	1965	1.055	7.119	7.178	0.925	0.843	0.796	0.721	0.511	0.3
ADAMCLISI	1966	1.723	6.796	7.322	0.959	0.860	0.800	0.710	0.472	0.3
ADAMCLISI	1967	0.976	6.864	5.201	0.949	0.865	0.782	0.709	0.510	0.3
ADAMCLISI	1968	1.117	7.625	9.771	0.955	0.880	0.823	0.749	0.522	0.3
ADAMCLISI	1969	1.493	7.317	10.978	0.963	0.904	0.855	0.799	0.600	0.4
ADAMCLISI	1970	1.633	6.348	5.941	0.966	0.906	0.840	0.782	0.562	0.3
ADAMCLISI	1971	1.670	6.042	5.694	0.964	0.899	0.841	0.789	0.612	0.4
ADAMCLISI	1972	1.533	7.974	7.103	0.967	0.911	0.861	0.803	0.615	0.4

(ID, Year, Annual data (e.g. various indexes) in columns: e.g. individual months, seasons and year can be used > easy processing of individual columns)

Examples of Data formats - ProClimDB, **daily data**

	Id	Year	Day	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12
▶	B1BYSH01_T_07:00	1866	1	-3.7	-0.6	2.8	3.8	11.6	16.0	17.4	9.3	11.3	9.7	4.1	-2.1
	B1BYSH01_T_07:00	1866	2	-3.2	2.7	2.5	4.5	15.0	16.0	15.8	9.7	11.4	12.0	0.5	-3.4
	B1BYSH01_T_07:00	1866	3	-3.0	5.7	0.8	5.8	4.7	15.7	17.0	12.8	12.3	9.1	3.5	1.0
	B1BYSH01_T_07:00	1866	4	-1.3	1.0	-3.2	8.0	4.8	14.5	10.5	13.0	8.8	8.0	5.0	1.5
	B1BYSH01_T_07:00	1866	5	-4.5	1.0	0.5	5.3	10.7	16.4	14.0	11.8	10.5	8.0	4.5	4.0
	B1BYSH01_T_07:00	1866	6	-6.5	1.1	-0.1	5.6	5.0	14.4	14.0	11.5	11.3	7.3	3.8	1.5
	B1BYSH01_T_07:00	1866	7	-3.9	5.0	2.9	8.1	4.1	14.5	11.7	9.0	12.6	1.0	6.8	3.3
	B1BYSH01_T_07:00	1866	8	-4.4	3.9	-1.1	8.7	5.6	14.8	10.6	13.8	14.2	0.8	4.5	2.2
	B1BYSH01_T_07:00	1866	9	-2.0	0.0	0.3	11.6	9.5	14.0	10.7	15.8	14.2	0.7	5.0	-1.5
	B1BYSH01_T_07:00	1866	10	-1.7	1.5	2.7	11.2	11.9	13.5	11.9	11.7	12.4	3.0	0.8	-2.0
	B1BYSH01_T_07:00	1866	11	-1.8	1.4	-0.6	6.8	6.8	14.6	12.3	10.7	12.5	0.5	-4.0	0.0
	B1BYSH01_T_07:00	1866	12	2.3	4.5	0.0	5.8	9.5	16.7	11.8	8.7	12.5	3.2	0.5	-5.5
	B1BYSH01_T_07:00	1866	13	-1.9	2.1	1.6	6.4	6.0	16.4	14.5	8.9	10.5	6.0	4.0	0.6
	B1BYSH01_T_07:00	1866	14	-3.6	-1.7	2.4	5.3	6.2	15.7	15.0	9.5	6.5	8.5	6.1	4.0
	B1BYSH01_T_07:00	1866	15	1.1	-3.0	-3.7	9.4	6.8	13.0	16.2	10.5	11.4	5.0	1.9	-6.2
	B1BYSH01_T_07:00	1866	16	0.0	0.0	-4.3	4.8	5.5	11.4	16.7	11.3	13.5	2.8	-0.3	-6.0
	B1BYSH01_T_07:00	1866	17	1.0	0.5	-1.6	6.9	3.5	15.5	16.2	10.5	7.7	0.0	4.0	-2.2
	B1BYSH01_T_07:00	1866	18	0.0	1.9	4.0	6.7	4.2	8.8	15.7	10.0	10.5	-2.1	-5.0	-1.4
	B1BYSH01_T_07:00	1866	19	3.0	3.3	2.4	6.9	3.0	11.6	13.5	10.5	8.8	-0.1	-1.0	-0.9
	B1BYSH01_T_07:00	1866	20	1.0	-2.0	6.0	1.7	2.1	14.7	12.8	10.5	9.0	-1.5	-6.2	-3.9
	B1BYSH01_T_07:00	1866	21	0.9	-9.3	0.9	7.0	1.8	11.8	10.4	12.5	7.5	-1.1	-6.0	-4.9

(ID, Year, Day, Months in columns: very useful format > easy processing of individual months)

	Year	Month	Day	Id	Value2
	1961	1	1	T1HOLE01	-0.4
	1961	1	1	T1IVAN01	-1.6
	1961	1	1	T1KIOM01	-1.0
	1961	1	1	T1LUHA01	-0.6
	1961	1	1	T1TYSH01	-1.2
	1961	1	2	T1HOLE01	-2.3
	1961	1	2	T1IVAN01	-2.9
	1961	1	2	T1KIOM01	-3.5
	1961	1	2	T1LUHA01	-1.1
	1961	1	2	T1TYSH01	-3.5
	1961	1	3	T1HOLE01	-2.0
	1961	1	3	T1IVAN01	-2.2
	1961	1	3	T1KIOM01	-1.5
	1961	1	3	T1LUHA01	-2.9
	1961	1	3	T1TYSH01	-2.9
	1961	1	4	T1HOLE01	3.5

(ID, Year, Month, Day, Value: very space consuming > long time calculations …)

Examples of Data formats - ProClimDB, **daily data**

ID	YEAR	MONT	VAL01	VAL02	VAL03	VAL04	VAL05	VAL06	VAL07	VAL08	VAL09	VAL10	VAL11	VAL12	VAL13	VAL14	VAL15	VAL16	VAL17	VAL18	VAL19	VAL
B2DVES02	2001	3	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999
B2DVES02	2001	4	4.2	0.0	12.1	13.5	8.8	9.7	11.0	7.7	7.4	7.8	9.5	8.4	3.6	2.6	4.3	7.1	6.3	8.3	6.6	6
B2DVES02	2001	5	18.4	-999.0	21.2	19.5	18.7	-999.0	12.6	15.0	16.5	16.5	15.3	13.2	14.4	15.3	15.5	18.8	20.1	14.2	13.0	14
B2DVES02	2001	6	13.2	14.6	13.0	11.1	13.2	14.8	17.9	18.0	14.7	14.7	12.3	14.9	17.0	17.4	20.0	18.3	16.8	16.6	14.9	17
B2DVES02	2001	7	18.7	18.2	15.9	19.0	20.8	22.3	23.0	20.4	20.9	23.4	19.2	20.3	22.0	24.8	27.2	21.6	15.3	19.6	20.4	17
B2DVES02	2001	8	-999.0	22.5	25.9	22.1	19.3	21.3	22.0	21.6	21.0	17.7	17.0	17.7	19.7	-999.0	23.7	24.5	-999.0	24.4	25.1	21
B2DVES02	2001	9	15.9	16.9	18.5	16.4	13.9	14.8	14.1	15.4	11.9	11.9	12.8	-999.0	13.5	13.3	12.9	11.8	11.3	10.6	12.0	12
B2DVES02	2001	10	17.4	19.8	15.5	14.7	13.6	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	12.9	14.3	11.1	12.0	12.1	12.4	12
B2DVES02	2001	12	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	0.0	0
B2DVES02	2002	1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0
B2DVES02	2002	2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999
B2DVES02	2002	3	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999.0	-999
B2DVES02	2002	4	0.0	0.0	0.0	0.0	2.5	0.8	2.7	6.2	7.0	5.2	7.8	8.3	9.9	9.4	8.5	10.0	11.8	13.1	11.5	11
B2DVES02	2002	5	16.1	19.5	22.6	22.8	13.3	15.2	16.7	18.1	18.5	17.3	18.1	18.9	18.6	17.3	16.4	18.1	21.8	18.5	17.9	16
B2DVES02	2002	6	15.8	13.8	17.4	20.6	19.8	17.3	15.4	15.6	16.5	16.7	16.7	19.8	21.7	21.9	21.4	22.4	22.6	25.4	25.7	26
B2DVES02	2002	7	22.6	21.5	25.3	17.1	19.5	22.4	20.8	22.8	24.6	28.5	21.2	22.3	22.1	22.5	23.9	21.7	21.4	20.4	19.1	21
B2DVES02	2002	8	22.7	10.5	0.0	0.0	17.8	20.9	19.7	20.7	20.5	21.0	19.8	17.0	17.0	17.3	19.5	19.1	21.0	21.7	22.2	21
B2DVES02	2002	9	18.9	16.6	20.1	21.3	20.4	18.6	19.2	20.4	21.2	18.7	17.9	14.0	13.8	15.2	12.9	12.8	13.6	15.5	16.1	15
B2DVES02	2002	10	11.1	11.3	12.0	13.1	10.8	11.7	7.4	7.7	8.1	7.7	5.4	3.8	5.6	6.9	9.8	10.4	11.6	8.5	7.3	6

(ID, Year, Month, Days in columns)

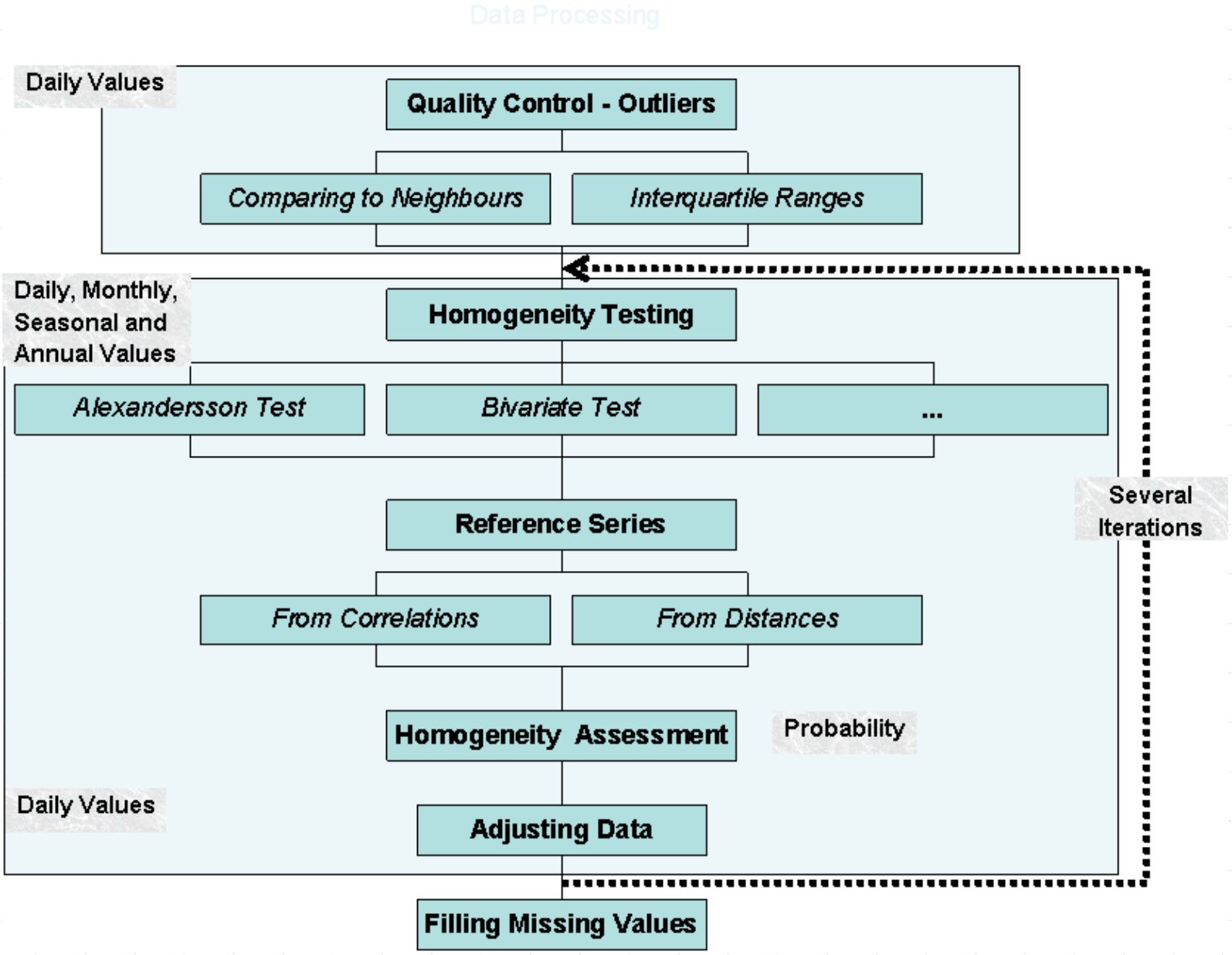
Year	Month	Day	T1hole01	T1ivan01	T1kiom01	T1luha01	T1piot01	T1tity01	T1tysh01
1961	1	1	-0.4	-1.6	-1.0	-0.6	-999	-999	-1.2
1961	1	2	-2.3	-2.9	-3.5	-1.1	-999	-999	-3.5
1961	1	3	-1.1	-1.3	-0.6	-2.0	-999	-999	-2.0
1961	1	4	3.5	0.3	1.0	3.1	-999	-999	2.6
1961	1	5	0.3	-1.4	-1.0	0.4	-999	-999	-0.5
1961	1	6	-3.4	-3.5	-8.0	-3.9	-999	-999	-4.5
1961	1	7	1.4	-1.9	-2.5	1.1	-999	-999	0.5
1961	1	8	-1.6	-2.4	-2.6	-1.4	-999	-999	-2.5
1961	1	9	-0.6	-1.4	-4.0	-0.8	-999	-999	-1.0
1961	1	10	-0.9	-1.1	-1.0	-0.9	-999	-999	-0.5
1961	1	11	0.2	0.1	-0.3	0.1	-999	-999	-2.0
1961	1	12	-0.3	-1.4	-0.3	-0.3	-999	-999	-2.2
1961	1	13	-6.2	-8.5	-7.5	-4.4	-999	-999	-8.5

(Year, Month, Day, ID's – stations in individual columns > suitable in case of the same period of measurements)

ProcData software, info_file

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

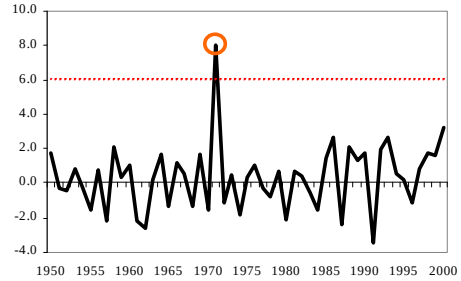
General scheme of data processing before time series analysis



Data Quality Control
Finding suspicious values

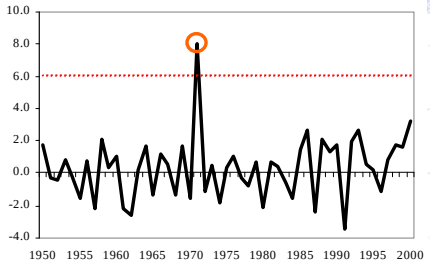
Two main approaches:

- Using limits derived from interquartile ranges (time series)

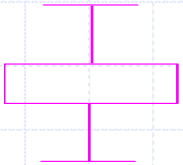


- comparing values to values of neighbouring stations (spatial analysis)

Data Quality Control
Finding Outliers



- 1. Limits derived from interquartile range
  - relatively, series of ratios (logarithms) of tested and reference series
    - reference series created as an average of 5 mostly correlated stations, max. distance 35 km (precipitation)
    - limits: coefficient (multiple) = 3.0
  - absolutely, in the past when only one station is available
    - in cases when less than three neighbours have been found
    - limits: coefficient (multiple) = 5.0



Data Quality Control
Finding Outliers

- **2. comparing values to values of neighbouring stations**
  - **comparing to <u>min. 3</u> to 5 best correlated stations**
  - **series of standardized differences (logarithms of ratios)**
  - **number of cases exceeding 95% confidence limits**
  - **Comparison to „expected“ value – calculated from neighbours (using distances or correlations)**

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**	

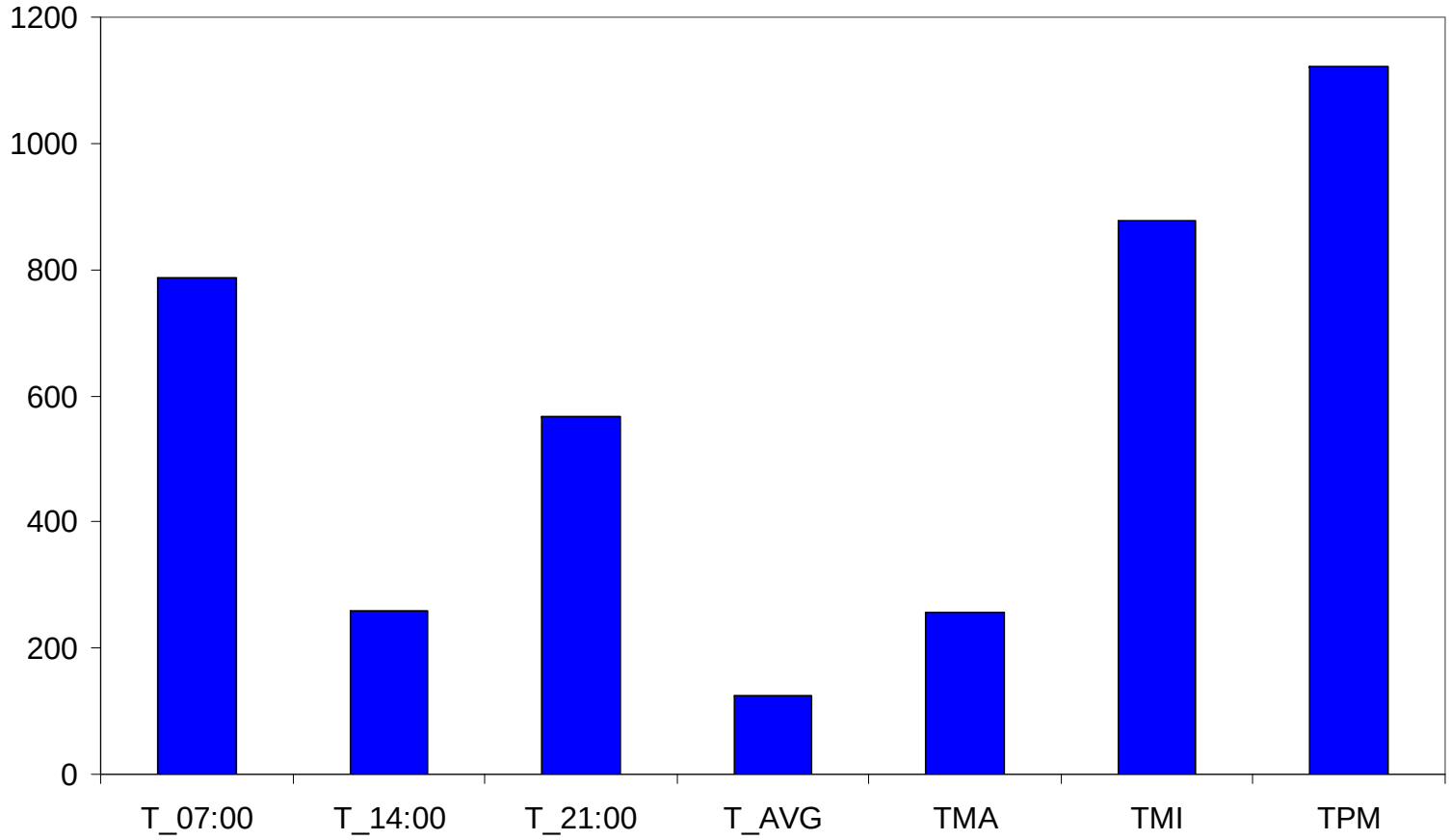
Quality control

- Run for CZ period 1961-2007, daily data (measured values in observation hours)
- All stations (CZ: 200 climatological stations, 800 precipitation stations)
- All meteorological elements (T, TMA, TMI, TPM, SRA, SCE, SNO, E, RV, H, F)
- Found optimal settings in the software for each met. element

- Historical records will follow now

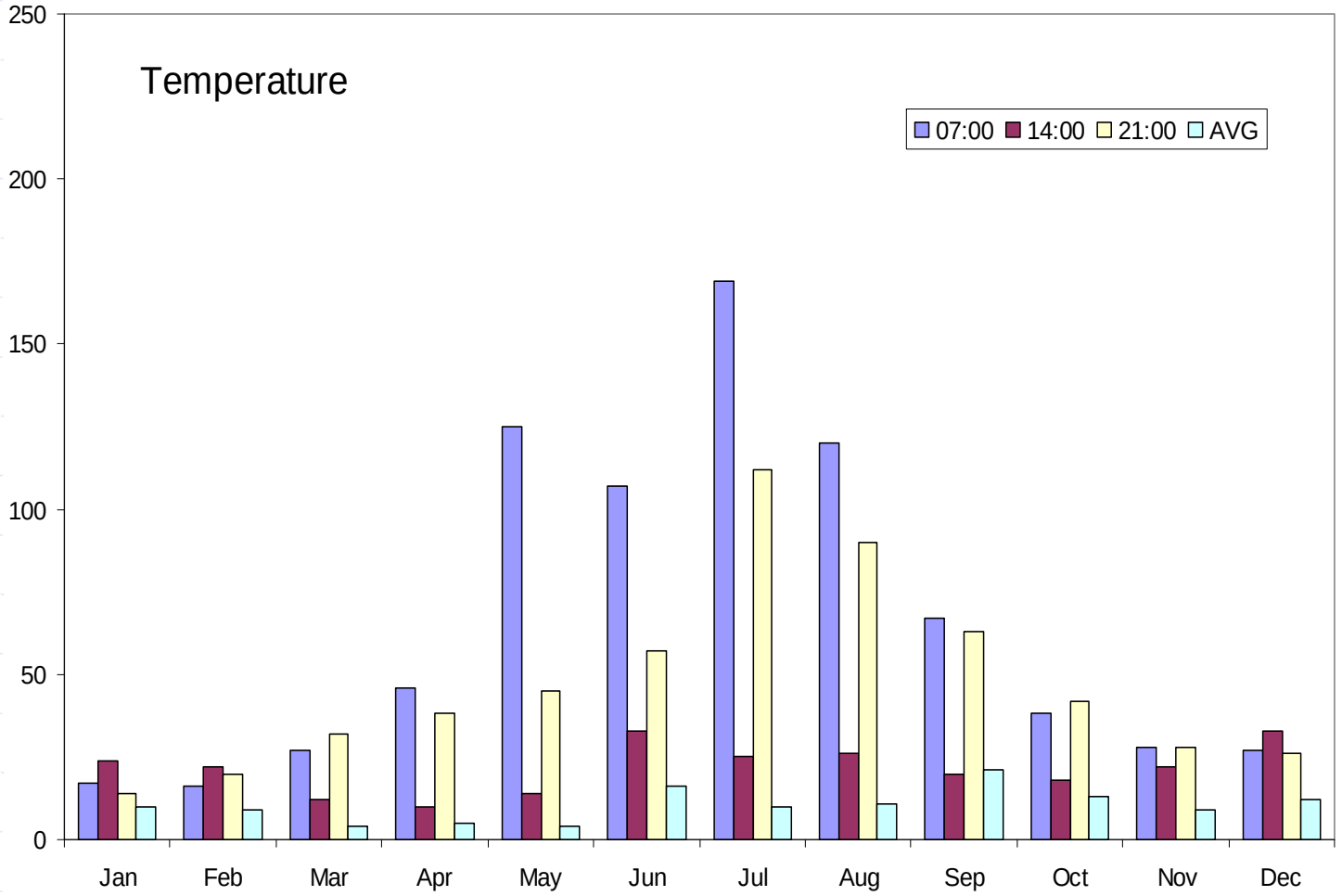
Air temperature, number of outliers 1961-2007, from 3.431.000 station-days

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.

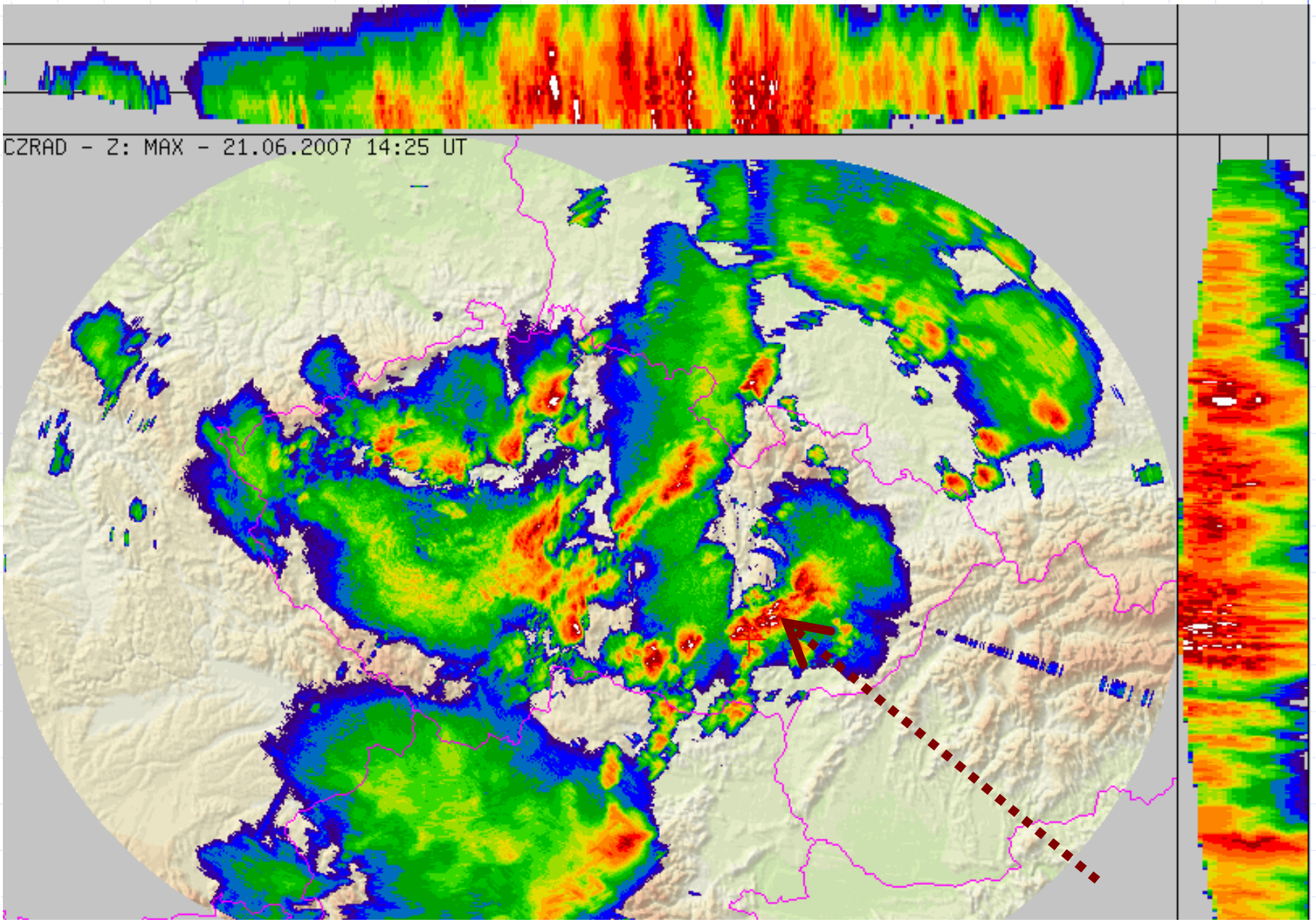


Problematic detections (heavy rainfall)

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),

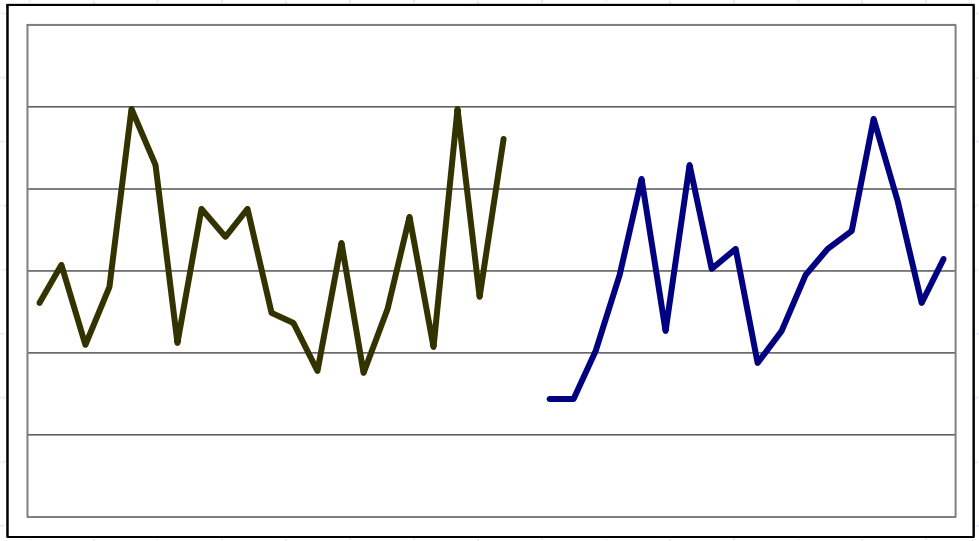
combination with Radar information



Remarks to QC

- Only combination of several methods for outliers detection leads to satisfying results ("real" outliers detection, supressing fault detection -> Emsemble approach) and makes it possible to automatize QC
- Parameters (settings) has to be found individually for each meteorological element, maybe also region (terrain complexity) and part of a year (noticeable annual cycle in number of outliers)
- Similar to homogenization of time series, it is important to use measured value (e.g. from observation hours) - outliers are masked in daily average (and even more in monthly or annual ones)

Combining measurements of neighbouring stations



Combining measurements of neighbouring stations

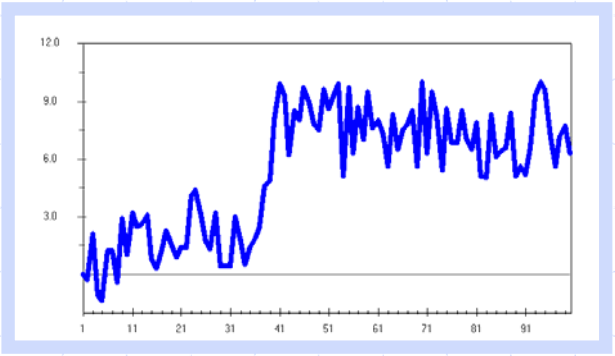
- **Selecting neighbours till ...** km
- **gap between two series: maximum ...** years
- **resulting series: at least ...** years long

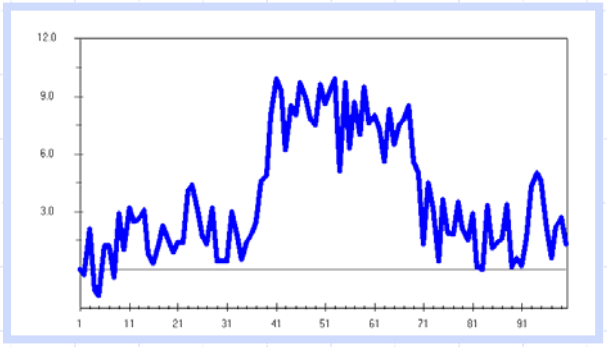
Example:
Combining series of neighbouring stations

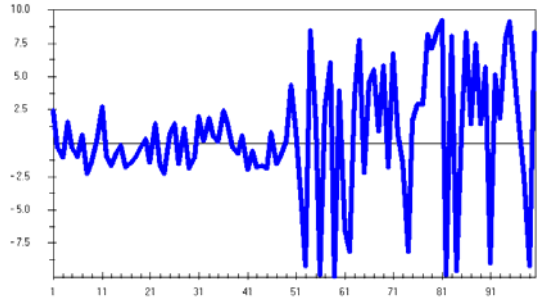
ID	REMARK	BEGIN	END	LENGTH	*REMARK*	DISTANCE	ALT_1	ALT_2	ALL_MONTHS
B1POZL01		1961	1993	33					
	List of neighbours>								
	1.>B1HLHO01	1961	2000	40	*49.16;17.80*	**4.68**		340	
	2.>B1LUHA01	1961	2000	40	*49.11;17.77*	**1.33**		254	
	Correlations of Overlap>								
	1.>B1HLHO01	1961	1993	33					0.944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

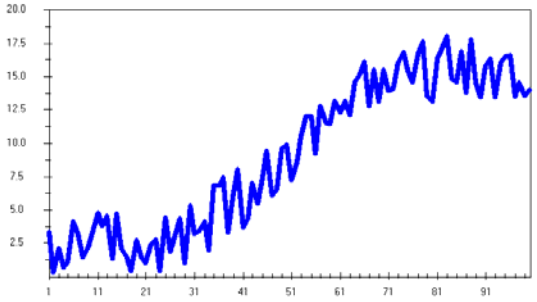
Homogeneity testing

- Change in shift or variance, trend detection, p=0.05









Creating Reference Series

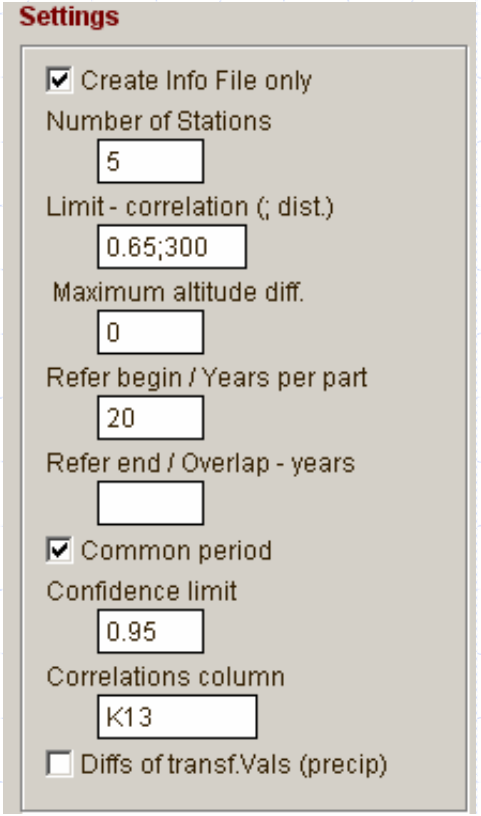
- for monthly, daily data (each month individually)
- weighted/unweighted mean from neighbouring stations
- criterions used for stations selection (or combination of it):
  - best correlated / nearest neighbours

    (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	LENG	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

Relative homogeneity testing

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

40 year parts of the series **(10 years overlap),**

in SNHT splitting into subperiods in position of detected significant changepoint

(30-40 years per one inhomogeneity)

Homogeneity assessment

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

Example I:

Homogeneity assessment

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

Example II:

Homogeneity assessment

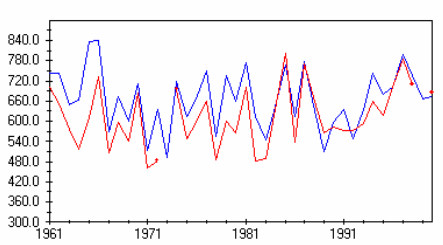
ID	BEGIN	END	LENGTH	YEAR_INHOM	YEAR_COUNT	Y_PORTION	Y_POSSIBL
B1KROM01	1961	2000	40		104	36.24	36.11
B1KROM01				1977	18	6.27	6.25
B1KROM01				1978	13	4.53	4.51
B1KROM01				1975	11	3.83	3.82
B1KROM01				1974	10	3.48	3.47
B1KROM01				1983	8	2.79	2.78
B1KROM01				1987	8	2.79	2.78
B1KROM01				1989	7	2.44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

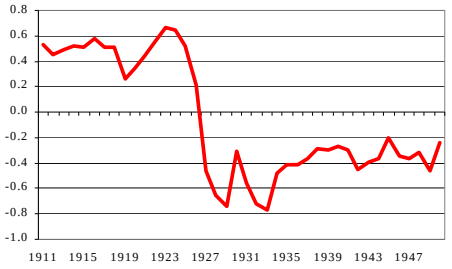
Summed numbers of detections for individual years

Homogeneity assessment

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

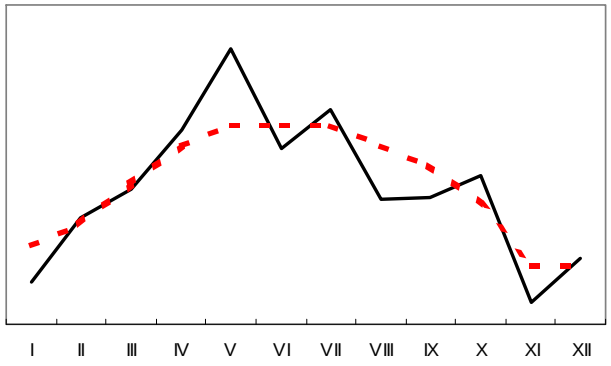
	ID	EL	YEAR	BEGIN	END	YEAR_COUN	Y_POSSIBL	YEA	MIS	X_BEGIN_D	X_END_DA	X	X	L	L	A	E	REMARK	C	C
x	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.3.2003	#	#				E	change		
	B1BOJK01	x	1985			41	14.24		12	23.3.1984	31.12.9999	#	#					obs	V	B
	B1BYSH01	x	1978			37	12.85													
?	B1BYSH01	x	1979			33	11.46													
?	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	#	#				L	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	#	#				L	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	#	#				E	change		
	B2BRUM01	x	1989			51	17.71			1.2.1989	31.3.1994	#	#					obs	M	B





Adjusting data

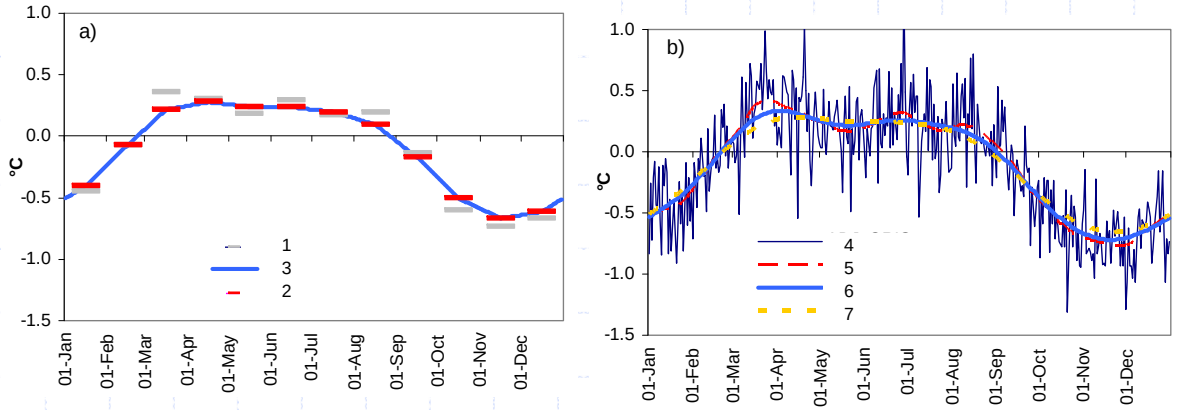
- using reference series based on correlations
- adjustment: from differences (ratios) ... years before and after a change, monhtly
- smoothing monthly adjustments (low-pass filter for adjacent values)



Adjustments (Delta method)

- **The same final adjustments may be obtained from either monthly averages or through direct use of daily data**

(for the daily-values-based approach, it seems reasonable to smooth with a low-pass filter for 60 days. The same results may be derived using a low-pass filter for two months (weights approximately 1:2:1) and subsequently distributing the smoothed monthly adjustments into daily values)

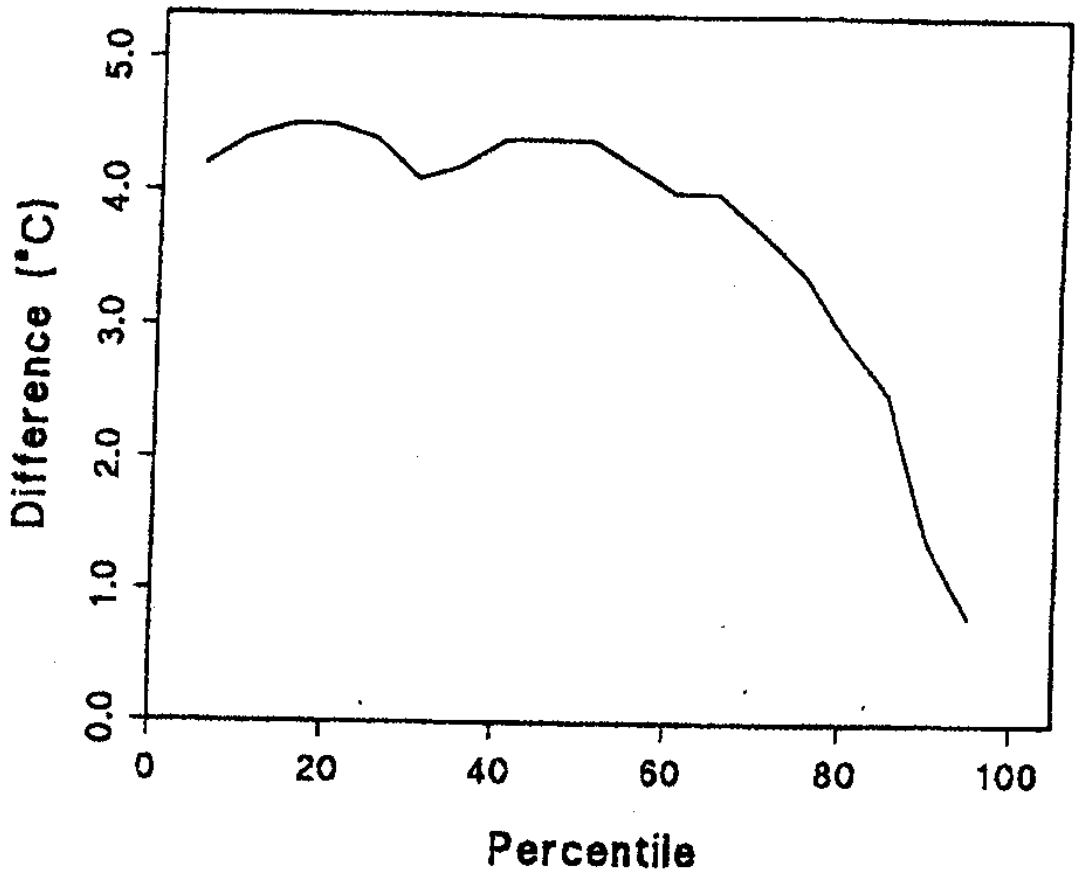


(1 – raw adjustments, 2 – smoothed adjustments, 3 – smoothed adjustments distributed into individual days), b) daily-based approach (4 – individual calendar day adjustments, 5 – daily adjustments smoothed by low-pass filter for 30 days, 6 – for 60 days, 7 – for 90 days)

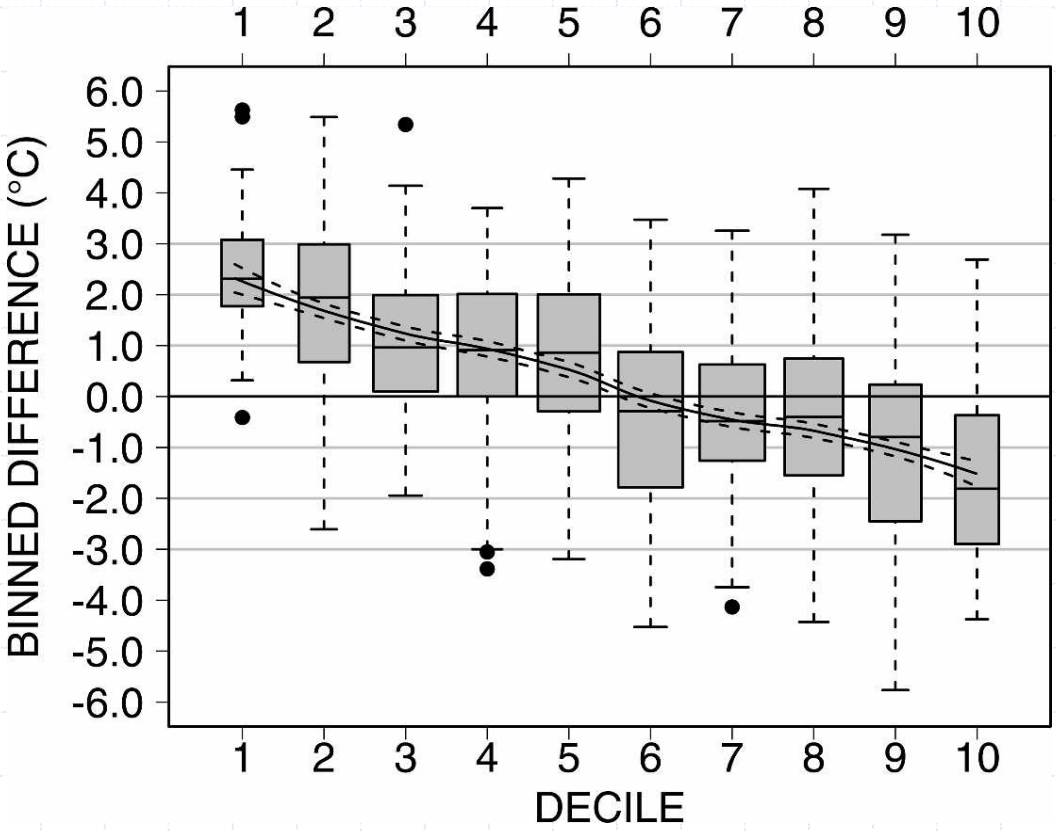
Variable correction

B. C. TREWIN AND A. C. F. TREVITT

1996



Variable correction,
The higher-order moments method



DELLA-MARTA AND WANNER, JOURNAL OF CLIMATE 19 (2006) 4179-4197

Example:

Adjusting values - evaluation

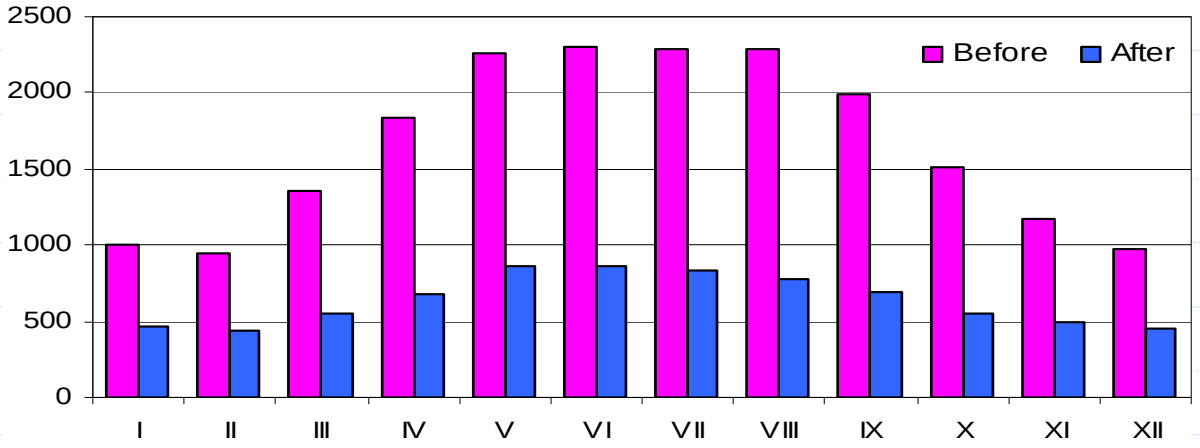
ID_1	pII	BEGIN	END	YEAR	MONTH	REMARK	C(	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K11	K12
B1RYCH01	E	1961	1992	1973	5	ADJust		1.135	1.197	1.155	1.333	1.149	1.070	1.088	1.354	1.145	1.116	1.136	1.265
B1RYCH01						DIFF1		0.905	0.875	0.912	0.813	0.906	0.956	0.896	0.786	0.912	0.956	0.908	0.855
B1RYCH01						DIFF2		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

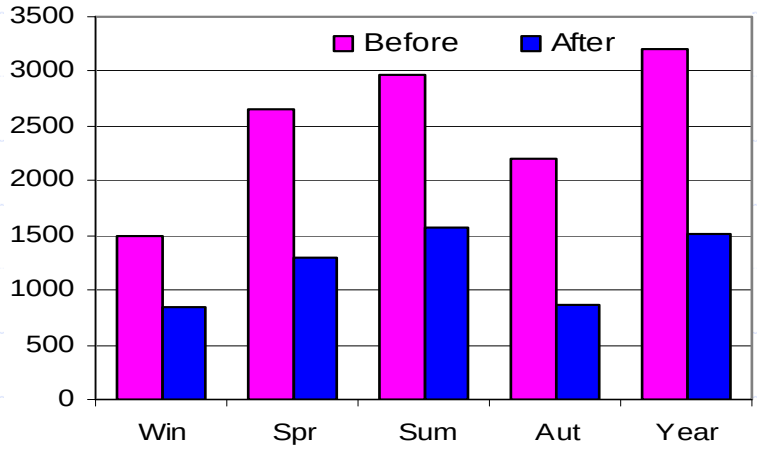
Iterative homogeneity testing

- 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

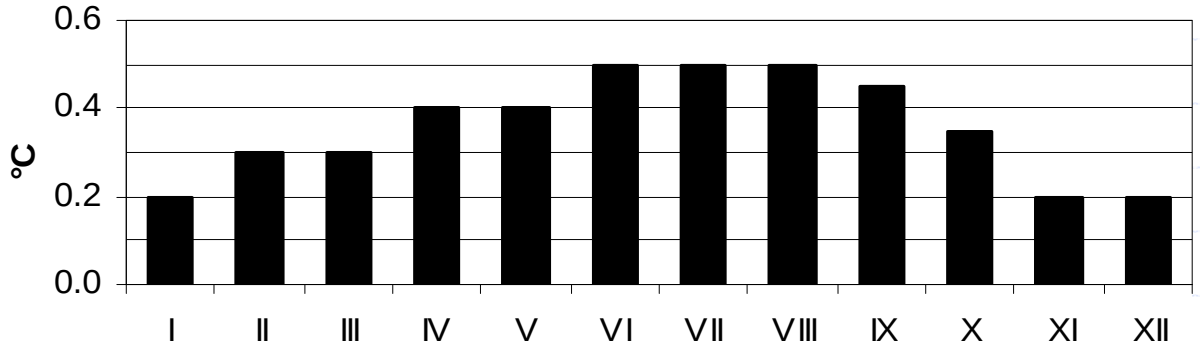
Example: CZ, air temperature (200 stations, 1848-2000)

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

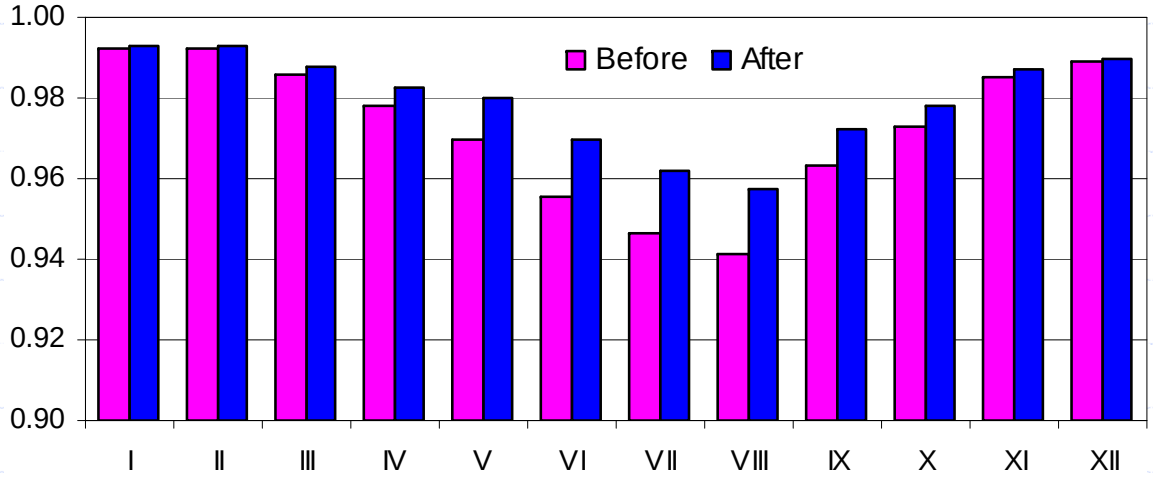




Amount of adjustments for homogenised series (absolute values) - median



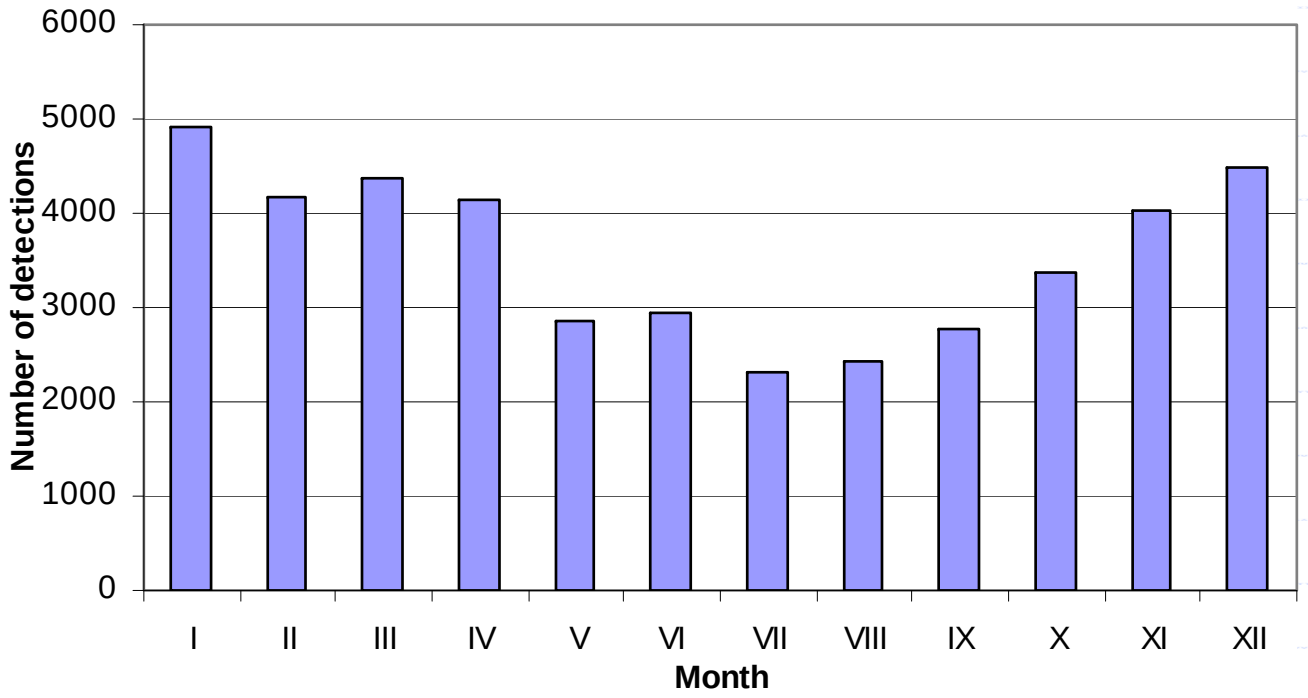
Correlation coefficients between candidate and reference series before and after homogenization (median)



Example: CZ, precipitation

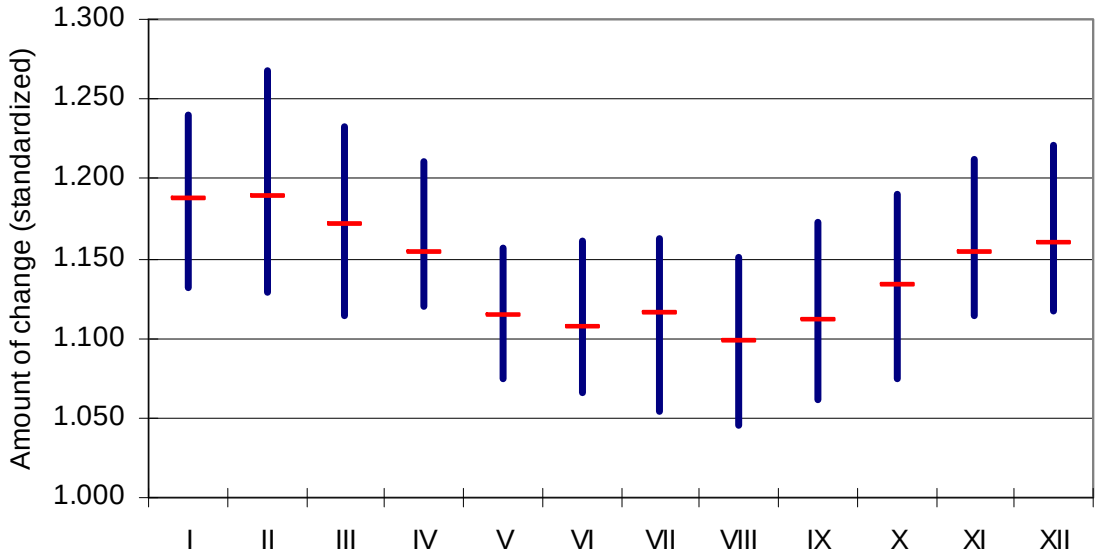
(800 stations, 1961-2000)

- 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)

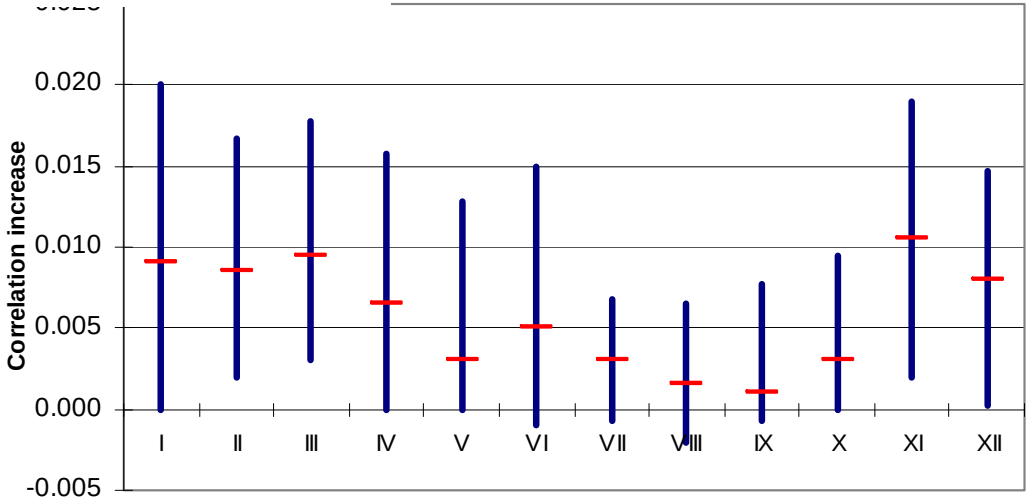


Boxplots:

- Median
- Upper and lower quartiles

(for 589 testes series)

Correlation improvement



Remarks - Homogenization 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, …

Remarks - Homogenization 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

Technical series calculation

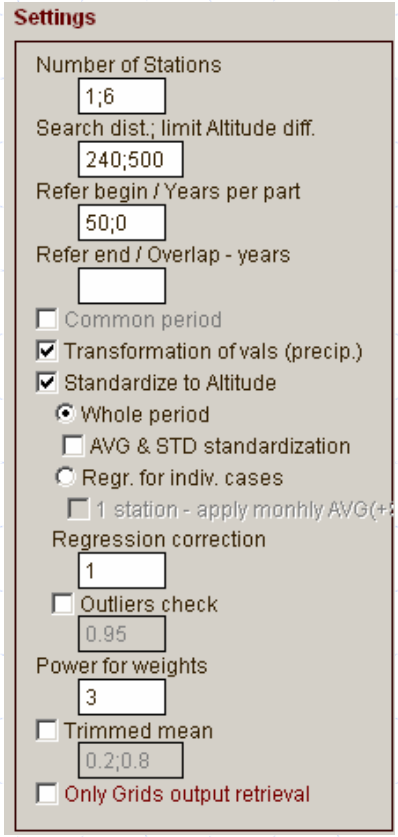
Output:

- Station technical series (QC, homogenized, filled gaps)
- Grid points series (regular network)

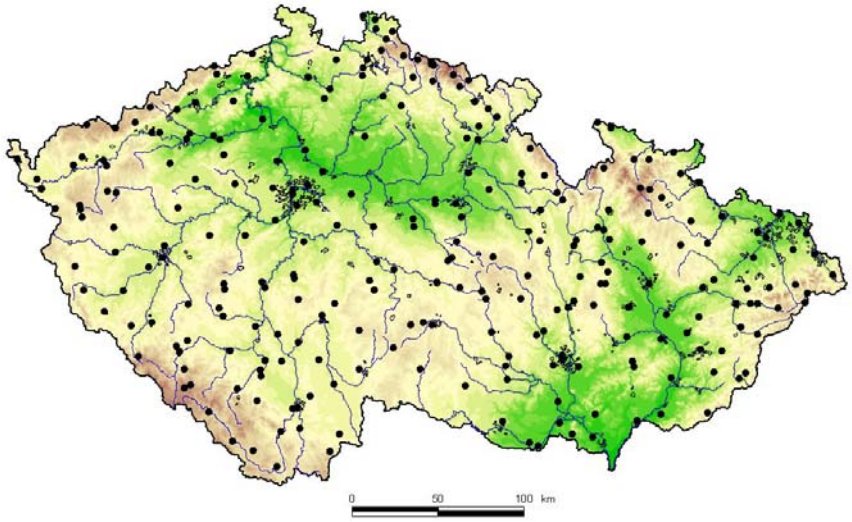
- Daily scale
- 1961-2000 / 2008 (adding new years)
- Various elements (T,TMA,TMI,SRA,SSV,E,F)
- Various regions:
  - whole CZ, various spatial resolution (10, 25km)
  - whole SK (10 km)
  - CECILIA region (10 km)

Technical series calculation

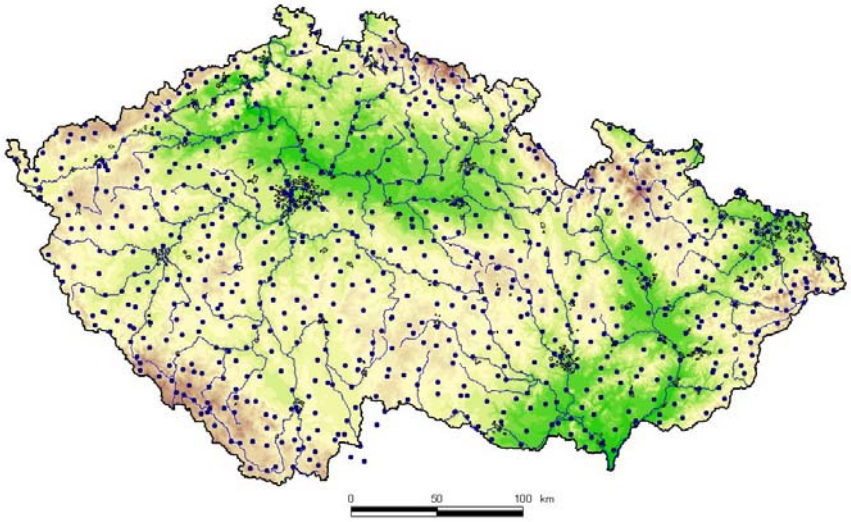
- Method of interpolation:
  - Local linear regression – standardization of neighbour stations values to altitude of given location
  - IDW method for "expected" value calculation (applied different power of inverse distance for various met. elements)
- Time series calculated for arbitrary point, in daily scale
- Single realization in time (each day individually) (it solves inversion etc...),



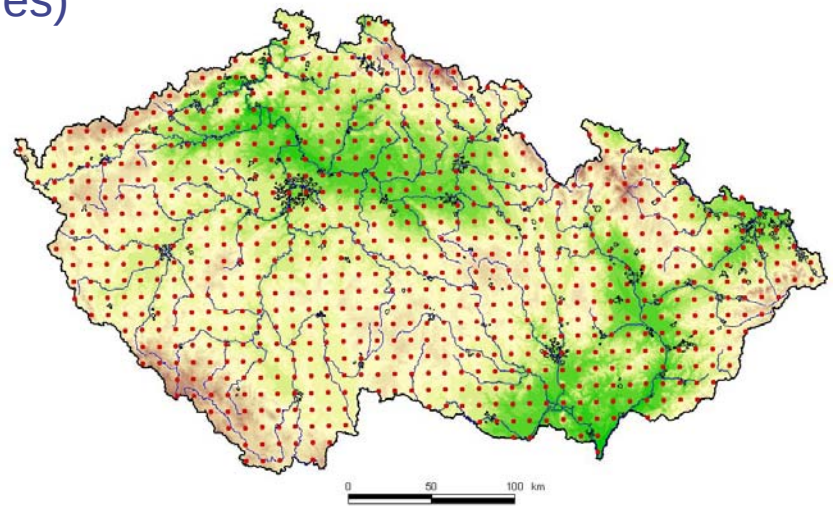
Stations (268) with calculated technical series, climatological stations



Stations with technical series, precipitation stations

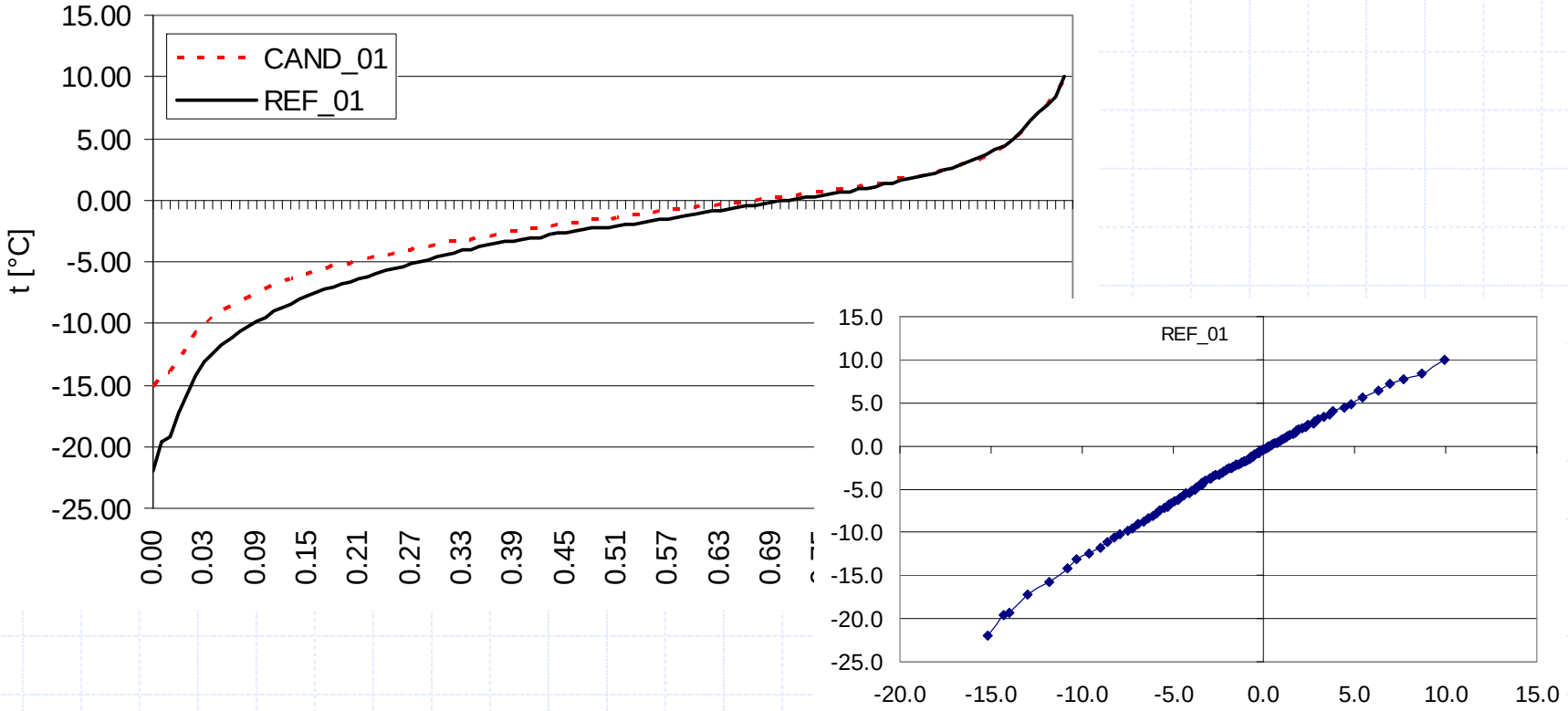


Stations with technical series, for grid points of ALADIN-Climate/CZ (1961-2008, 7 met. variables)



Model output correction

- an approach of Michel Déqué (2007) based on variable correction using individual percentiles
- for each grid point and month individually



Software package – conclusions

- Learning the methods in climatology with **AnClim**
- Processing whole national datasets using LoadData and **ProClimDB**
- Freeware versions
- Continuous improvement via user **feedbacks** (adding new functionality)

http://www.climahom.eu