IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 117, No. 1, January–March 2013, pp. 123–141

Experiences with data quality control and homogenization of daily records of various meteorological elements in the Czech Republic in the period 1961–2010

Petr Štěpánek[1,2] ***, Pavel Zahradníček**[1,2] and **Aleš Farda**[2]

[1]*Czech Hydrometeorological Institute, regional office, Kroftova 43, 616 67 Brno, Czech Republic*

[2]*Global Change Research Centre AS CR, v.v.i, Bělidla 986/4a, 603 00 Brno, Czech Republic*

**Corresponding author E-mail: petr.stepanek@chmi.cz*

(Manuscript received in final form December 14, 2012)

Abstract–Quality control and homogenization has to be undertaken prior to any data analyses in order to eliminate any erroneous values and non-climatic biases in time series. In recent years, considerable attention was paid to daily data since it can serve, among other conventional climatological analyses, as non-biased input into extreme value analysis, correction of RCM outputs, etc. In this work, we describe and then apply our own approach to data quality control of station measurements, combining several methods: (i) by analyzing difference series between candidate and neighboring stations, (ii) by applying limits derived from interquartile ranges, and (iii) by comparing the series values tested with "expected" values – technical series created by means of statistical methods for spatial data (e.g., IDW, kriging). Because of the presence of noise in series, statistical homogeneity tests render results with some degree of uncertainty. In this work, the use of various statistical tests and reference series made it possible to increase considerably the number of homogeneity test results for each series and, thus, to assess homogeneity more reliably. Inhomogeneities were corrected on a daily scale. In the end, missing values were filled applying geostatistical methods; thus, the so-called technical series for stations were constructed, which can finally be used as quality input into further time series analysis. These methodological approaches are applied to daily data, for various meteorological elements within the area of the Czech Republic in the period 1961–2010, which allows demonstrate their usefulness. Series were processed by means of the developed ProClimDB and AnClim softwares (http://www.climahom.eu).

Key-words: data quality control, homogenization, statistical correction of inhomogeneities, daily data processing, climatological time series

1. Introduction

For any meaningful climate analysis, investigated time series should be homogeneous, which means that their variations are caused solely by variations in weather and climate (*Conrad* and *Pollak*, 1950). Thus, prior to any analyses, the need to homogenize data and check their quality arises. Unfortunately, most of the climatological series that span from decades to centuries, contain inhomogeneities caused by station relocations, change of observers, changes in the vicinity of the stations (e.g., urbanization), changes in instruments, observing practices (e.g., different formulas for calculating daily means, different observation times), etc. (*Aguilar et al.*, 2003). Another important requirement for climatological analyses is the quality of the individual values, where series should be free of errors and have a low number of missing values (*Vicente-Serrano et al.*, 2010).

In the Czech Republic, this topic has been a focus of interest for several years. The first studies devoted to the homogenization of long series of air temperature, precipitation, and relative humidity for individual stations (e.g., *Macková*, 1997; *Brázdil* and *Štěpánek*, 1998; *Brázdil et al.*, 1996, 2000, 2001), which makes their use difficult (their availability, purpose of the given study, etc.). Later, studies devoting to the whole county have emerged (*Štěpánek*, 2003; *Štěpánek* and *Mikulová*, 2009; *Štěpánek et al.*, 2009), and this interest has continued up to the time of this study dealing with all the basic climatological characteristics throughout the whole territory of the Czech Republic. In recent years, considerable attention has also been devoted to the analysis of daily data (e.g., *Klein Tank et al.*, 2002; *Vincent et al.*, 2002; *Wijngaard et al.*, 2003; *Brunet et al.*, 2006; *Brandsma* and *Können*, 2006; *Della-Marta* and *Wanner*, 2006; *Vicente-Serano et al.*, 2010), which then may be used for various analyses, including those that were not possible to apply when homogenizing only monthly data, like the analysis of extreme value (*Sacré et al.*, 2007; *Kyselý* and *Picek*, 2007; *Costa* and *Soares*, 2009).

The organization of meteorological observations (i.e., maintaining the station network) and administration of collected data belong among the main duties of the Czech Hydrometeorological Institute (CHMI). The climatological database CLIDATA (*Tolasz*, 2008) serves very well for the usual quality control of the collected data (using GIS), but for the historical records, we face a lack of human resources, since the system requires user input. The software tools ProClimDB and AnClim (*Štěpánek*, 2010a, 2010b) allow automation of the process for quality control, homogenization, and filling of missing data but, at the same time, give the user a variety of outputs from which he can easily read what happened during the data processing and can track back all the important changes made to data, and also he can change the parameters and re-run the calculation. Thus, the ProClimDB and AnClim complement the aforementioned CHMI database system very well. However, it can also be used as a stand-alone application.

The general scheme of data processing that we advise being performed before any analyses includes the detection, verification, and possible correction of outliers (at the sub-daily scale – using the measured values), creation of reference series, homogeneity testing applying various statistical tests to better account for uncertainties in the results, determination of inhomogeneities in the light of test results and metadata (in monthly, seasonal and annual scale), adjustment of inhomogeneities (in daily scale) and, finally, the filling of missing values (see *Fig. 1*). We applied this approach to various meteorological elements available in our area in the period 1961–2010: mean, maximum and minimum air temperature, precipitation totals, water vapor pressure, wind speed and sunshine duration.

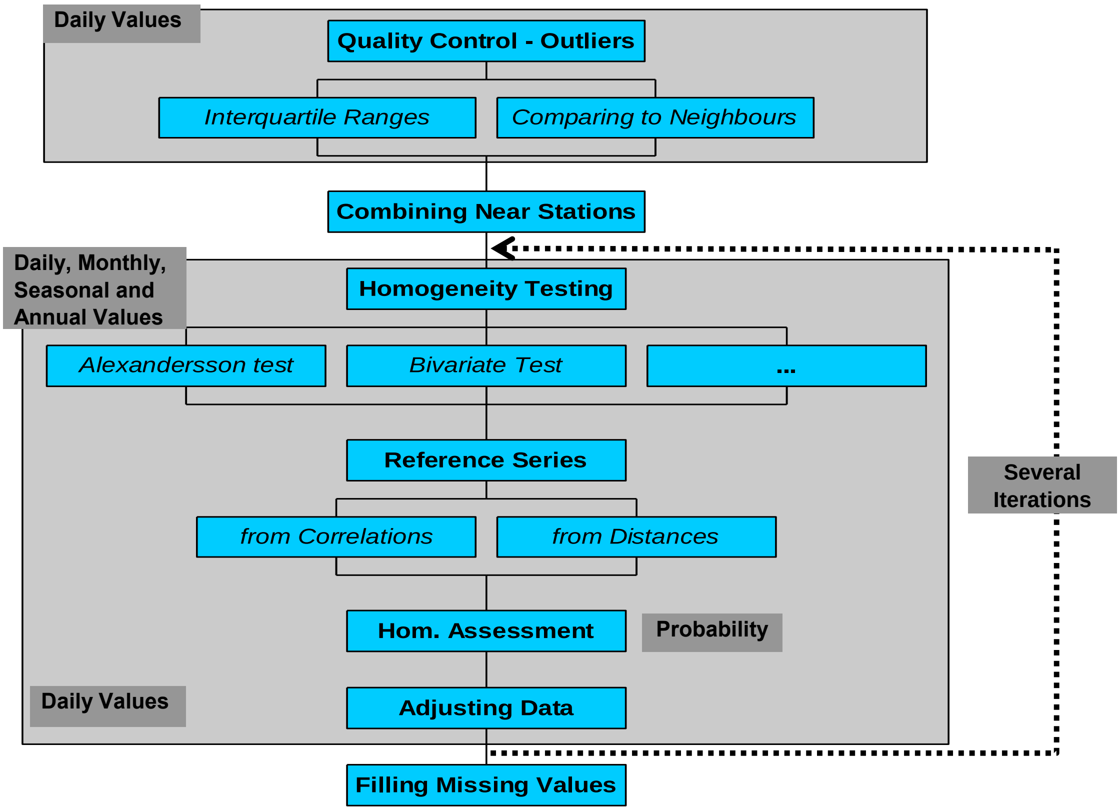

Fig. 1. Scheme of data processing – data quality control and homogenization.

2. *Data Quality control*

There is a lack of a generally accepted methodology for data quality control (contrary to homogenization). But, without outliers being properly treated, homogenization and analysis may render misleading results. Therefore, we devoted considerable attention to the methodology of outlier detection, to something that could, moreover, be automated to process large datasets of daily/sub-daily values (whole country dataset). This quality control was then applied to historical records which had not yet been processed by the methods used nowadays (using GIS and user interaction).

In our approach, data quality control is carried out by combining several methods:

(i) analyzing series of differences between candidate and neighboring stations (i.e. pairwise comparisons);

(ii) applying limits derived from interquartile ranges (either to individual series, i.e. absolutely, or better, to series of the difference between candidate and reference series, i.e. relatively); and

(iii) comparing the series of tested values with “expected” (theoretical) values – “technical” series created by means of statistical methods for spatial data (e.g. IDW, kriging).

Neighboring stations (method, (i)) or reference series (method, (ii)) may be selected either by correlations or distances (in the case of temperature, the results are different, while for precipitation, the selection coincides). Correlation coefficients can be applied either to raw series or to series of first differences (see, e.g., *Peterson*, 1998). In our case, for comparison with neighbor stations, up to six of the nearest stations were selected, with significant correlation coefficients, a distance limit of 400 km and an altitude difference restricted to 500 m. The distance limit was set with the help of a preceding analysis about how correlation coefficients drop with distance and change in altitude.

A method for outlier detection that could be automated to the greatest extent was a priority, since millions of values had to be processed for each meteorological element. Such a method was finally found and successfully applied. It utilizes a combination of several characteristics and their limits are based on the methods mentioned above (details on the quality control process may also be found in the documentation for the ProClimDB software, see *Štěpánek*, 2010b). No method on its own was found adequate; only their combination leads to satisfying results, i.e. the discovery of real outliers and suppression of false alarms. Parameters (settings appropriate to methods) had to be individually found for each meteorological element. The setting of parameters for outlier detection was validated using stations selected within different parts of the Czech Republic and also representing different altitudes.

As for the number of found suspicious values, the wind speed seems to be the most problematic variable, while air temperature has a relatively low number of problematic values (see *Fig. 2*). The number of outliers has a clear annual cycle. For most of the elements (e.g., air temperature), a higher number of outliers was detected in the summer months than in the winter months (larger spatial differences in summer are related to the increased influence of radiation factors compared to winter patterns, prevailingly influenced by circulation factors). More outliers were detected in the morning (7:00 local mean time –

LMT) and evening (19:00 LMT) observation terms compared to 14:00 LMT (associated with steeper gradients in the former case). For precipitation, there are two maxima per year: in the summer months and in December–January (this pertains to problems with solid precipitation measurements in winter), while during spring and autumn, a lower number of outliers were detected. The number of detected outliers also changes with time. The higher number of temperature outliers since the late 1990s coincides well with the transition to automatic measurements. Our explanation is that all values coming from automatic measurements (including errors) are stored directly into the database, while in the case of manual measurements, the observer revises the measured values before sending them to CHMI. In the last years, the number of outliers is again lower, owing to improved data quality control in the database. Conversely, in the case of precipitation, no increase of errors after automation was encountered (*Fig. 3*).

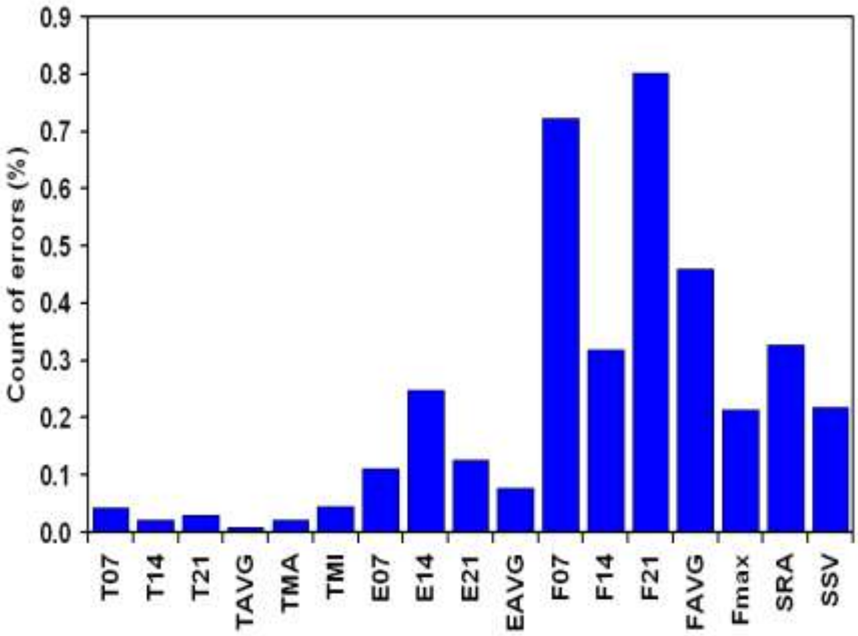

Fig. 2. Percentage portions of the number of errors detected in the total number of tested values for meteorological stations in the territory of the Czech Republic in the 1961–2009 period. Explanations: T – air temperature (T07, T14, T21 – observation terms, TAVG – daily mean), TMA – daily maximum air temperature, TMI – daily minimum air temperature, E – water vapor pressure (E07, E14, E21 – observation terms, EAVG – daily mean), F – wind speed (F07, F14, F21 – observation terms, FAVG – daily mean), Fmax – maximum daily wind gust, SRA – daily precipitation total, SSV – daily sunshine duration.

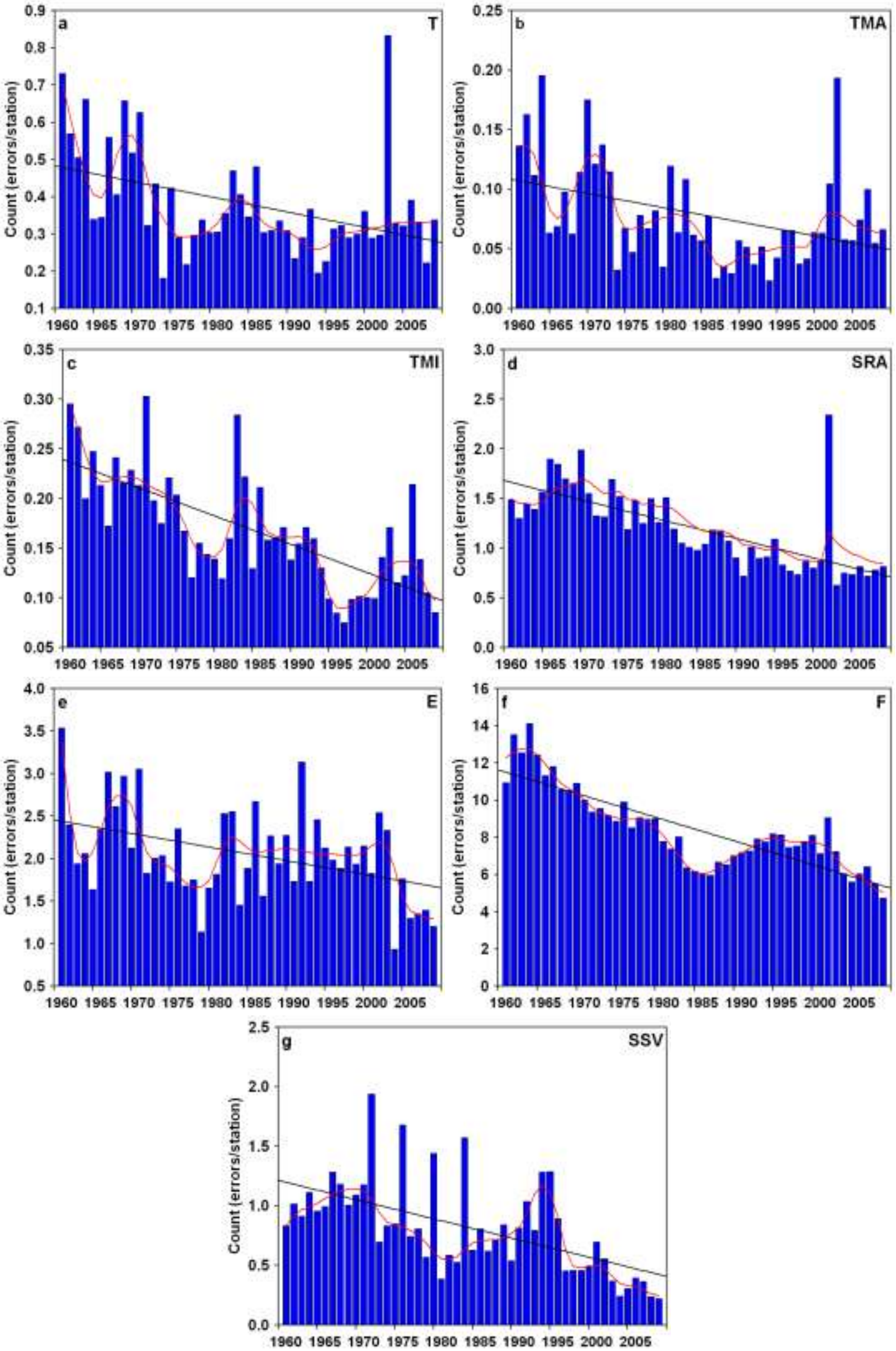

Fig. 3. Number of detected problematic values re-calculated per one meteorological station in the territory of the Czech Republic in the individual years of the 1961–2009 period: a) air temperature (observation terms and daily mean), b) maximum air temperature, c) minimum air temperature, d) precipitation total, e) water vapor pressure (observation terms and daily mean), f) wind speed (observation terms and daily mean), g) sunshine duration. The values are smoothed with a low-pass Gaussian filter for 10 years (red line) and complemented by the linear trend.

3. *Methodology of homogenization*

The general steps to be taken during homogenization consist of reference series creation (serving for comparison with tested series; this is a principal point of relative homogenization, see, e.g., *Conrad* and *Pollak*, 1950), applying statistical tests for testing the homogeneity of candidate series, homogenization (correction of inhomogeneities detected) and filling missing values (some prefer to fill missing values before homogenization). The individual steps are discussed, e.g., in *Štěpánek et al.* (2012), including a comparison of the results for various parameter settings (methods of weighting, number of stations used, individual statistical tests applied, method of correlation calculation for selection of neighbors, etc.)

Because of noise in time series, statistical homogeneity tests render results with some degree of uncertainty (see *Fig. 4*). In this work, the use of various statistical tests, types of reference series and time frames (monthly, seasonal, and annual series) allowed a considerable increase in the number of homogeneity test results for each series tested and thus to assess the homogeneity more reliably.

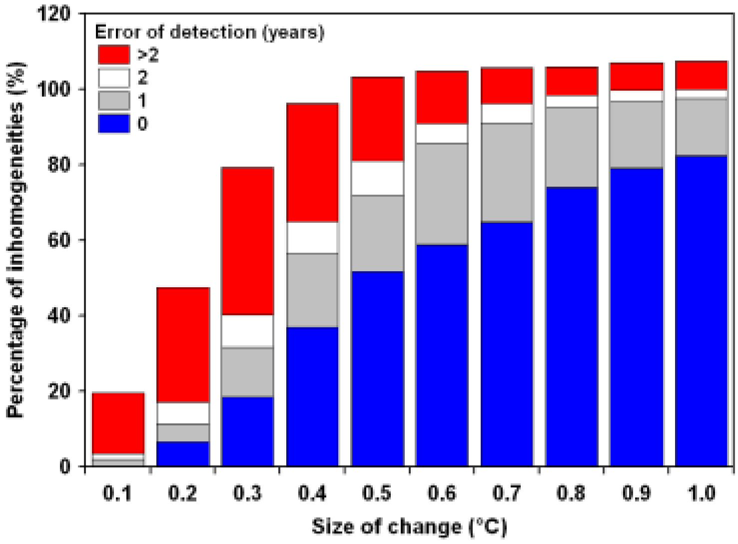

Fig. 4. The relative proportion (%) of the number of detected inhomogeneities of various sizes in the theoretically possible number of all inhomogeneities detected by the Alexandersson's SNHT test for the significance level of α = 0.05. Generated series shorter than 50 years with annual standard deviation were used. Zero false detection (blue) corresponds to the exact inhomogeneity estimation in the given year; further false detections are given for 1, 2, or more years apart (grey, white, and red). A total of 180 series were used for each category of the inhomogeneity size in mean (shift). The proportion of inhomogeneities exceeding 100% is due to dividing series into more parts during the testing.

The relative homogeneity tests applied were as follows: the standard normal homogeneity test [SNHT] (*Alexandersson*, 1986, 1995), the Maronna and Yohai bivariate test (*Potter*, 1981), and finally, the Easterling and Peterson test (*Easterling* and *Peterson*, 1995). Reference series were calculated as weighted means from the five nearest stations (measuring within the same period as the candidate series, they were also newly applied individually), with statistically significant correlations, a distance limit of 300 km, and an altitude difference limit of 500m. The weight (inverse distance) for temperature was taken as one and for precipitation as three. Neighbouring station values were standardized to the mean and standard deviation of the candidate station. The detection of inhomogeneities was performed for series divided into a maximum duration of 40 years, with an overlap for two consecutive periods of 10 years (due to requirements of SNHT to test only one shift in a series). The tests were applied for series of monthly, as well as seasonal and annual means (totals in the case of precipitation and sunshine duration).

The main criterion for determining a year of inhomogeneity was the probability of detection of a given year, i.e., the ratio between the count of detections for a given year from all test results for a given station (using type of reference series, range of tests applied, monthly, seasonal, and annual series) and the count of all theoretically possible detections (for more details of reference series creation and testing, see *Štěpánek et al.*, 2012).

After the evaluation of detected breaks and a comparison with metadata, a final decision on the correction of inhomogeneities was made. Data were corrected on a daily scale. The adjustment of such inhomogeneities was addressed by means of a reference series calculated in a similar way as described above.

We created our own correction method, an adaptation of a method for the correction of regional climate model outputs by *Déqué* (2007), itself based on assumptions similar to those implicit in methods described by *Trewin* and *Trevitt* (1996) and *Della-Marta* and *Wanner* (2006), which apply variable correction according to individual percentiles (or deciles). Our process is based on a comparison of percentiles (empirical distribution) of differences (or ratios) between candidate and reference series before and after a break. Percentiles are estimated from candidate and references series separately (not for the same date). Each month is processed individually, but the values of adjacent months before and after are also taken into account to ensure smoother passage from one month to another. Differences of candidate and reference series for individual percentiles are treated before and after a break and smoothed by low-pass filter to obtain a final adjustment based on a given percentile (see *Fig. 5* for illustration). Values before a break are then adjusted in such a way that we find a value for the candidate series before a break (interpolating between two percentile values if needed) and the corresponding correction factor, which is then applied to the values to be adjusted. Special treatment is needed for extremes at the ends of distribution. A comparison of the DAP approach and the "classic" one using monthly values is shown in *Fig. 6*.

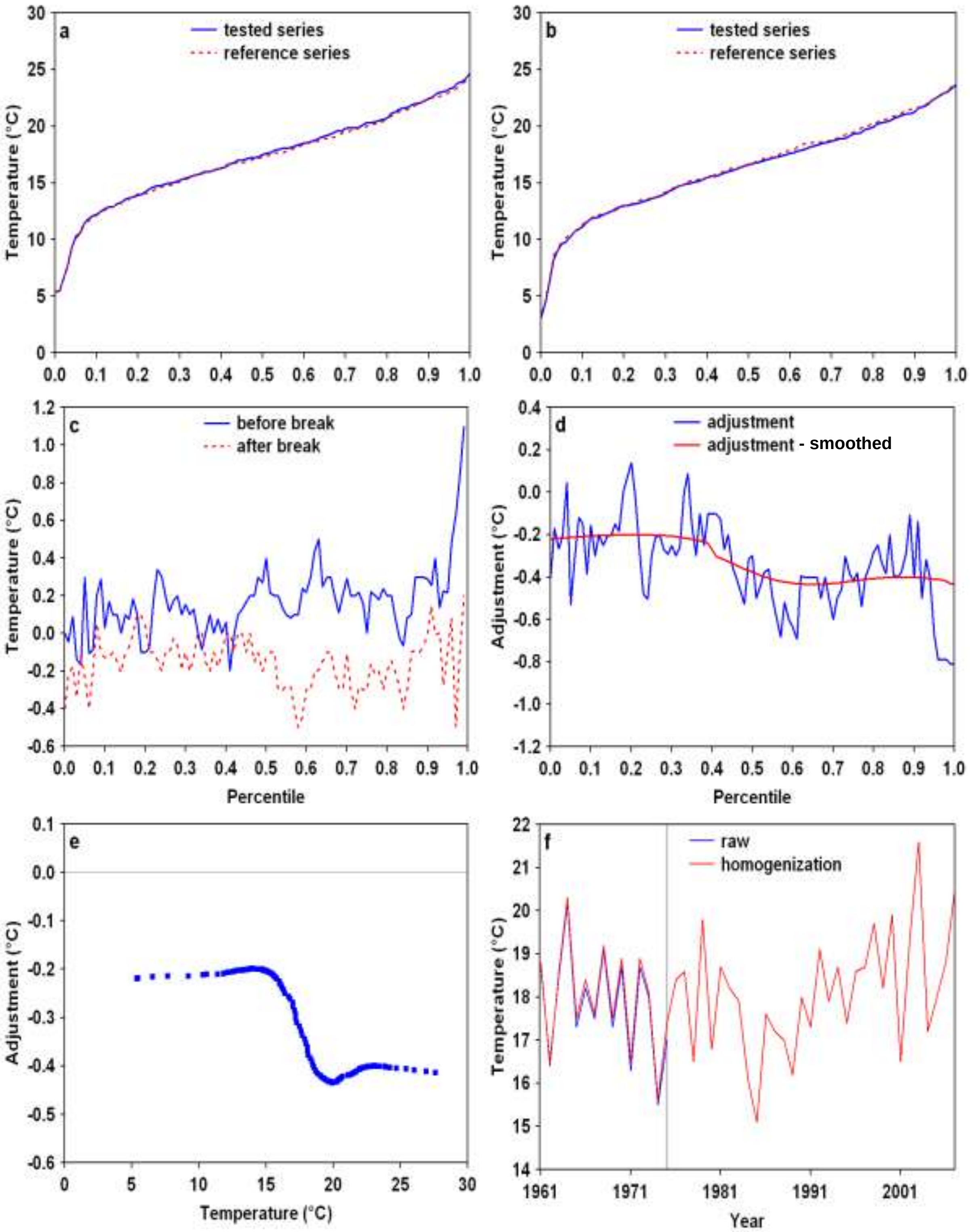

Fig. 5. Adjustment of series of daily mean air temperatures from the Velké Pavlovice station with an inhomogeneity detected in April 1975: a) quantiles (empirical distribution) of tested and reference series before the break, b) quantiles (empirical distribution) of tested and reference series after the break, c) the difference between tested and reference series for quantiles before and after the break, d) values of adjustments for quantiles (difference of tested and reference series differences before and after the break) and their smoothing by low-pass filter (final value of adjustments), e) adjustment values for a specific air temperature, f) original and homogenized series (monthly means).

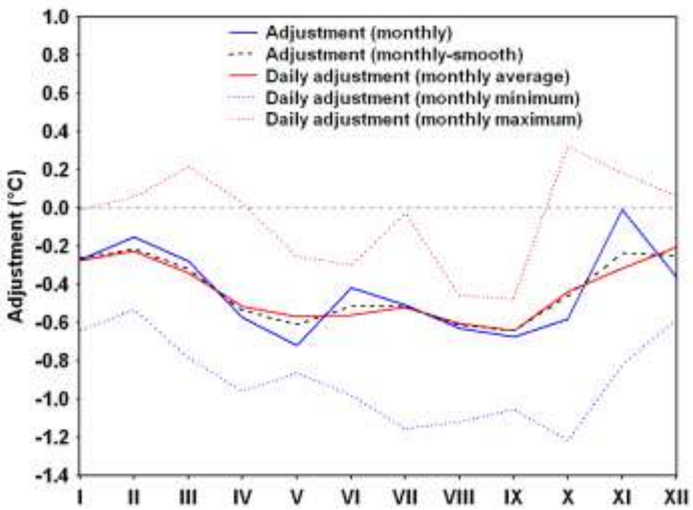

Fig. 6. Example of the series adjustment for inhomogeneity applying the classic approach with use of monthly data (solid blue line; smoothed corrections: dashed black line) and monthly means of daily corrections using DAP – distribution adjusting by percentiles method (solid red line; dotted red (blue) lines show the monthly maxima (minima) from daily corrections). The series is mean daily air temperature measured at station Bystřice pod Hostýnem with inhomogeneity on Januar 1, 1985.

Various characteristics were analyzed before applying the adjustments: the increment of correlation coefficients between candidate and reference series after adjustments; any change of standard deviation in differences before and after the change; the presence of linear trends, etc. In the event of any doubt, the adjustments were not applied.

Homogeneity testing, evaluation and correction of inhomogeneities detected were performed by several iterations, in which more precise results are gradually obtained. Missing values were filled after the homogenization and adjustment of inhomogeneities in the series. This means that the new values filled are estimated from data which are not influenced by possible shifts in the series. Filling missing data before homogenization may negatively influence inhomogeneity detection.

A preference for testing individual observation term series, if available, belongs among the recommendations for further homogenization improvement, since inhomogeneities are manifested in a different way in them (see *Fig. 7*). Further improvement can potentially be achieved by grouping values into categories, e.g., using weather types and testing individual categories alone.

Within the COST Action ES0601 ("Advances in homogenization methods of climate series: an integrated approach – HOME", 2006–2011), parameter settings used in this work were verified (at least for creation of reference series and detection of inhomogeneities that were run in the monthly mode) for air temperature and precipitation. The best parameter settings for air temperature were achieved applying a probability of detection equal to 20%, using correlations for neighbors selection calculated from the first differenced series, pairwise comparison with neighbors, and running several iterations of homogeneity testing and correction. The second and third iterations improved the series the best, while further iterations meant only negligible improvement. As for precipitation parameter settings, the best results were gained applying probability of detection equal to 10%, using correlations for neighbors selection calculated from the first differenced series, and pairwise comparison with neighbors; however, improvement of statistical characteristics of the series after homogenization was not as profound as in the case of air temperature. These settings will be applied in the follow-up work dealing with historical records (before 1961).

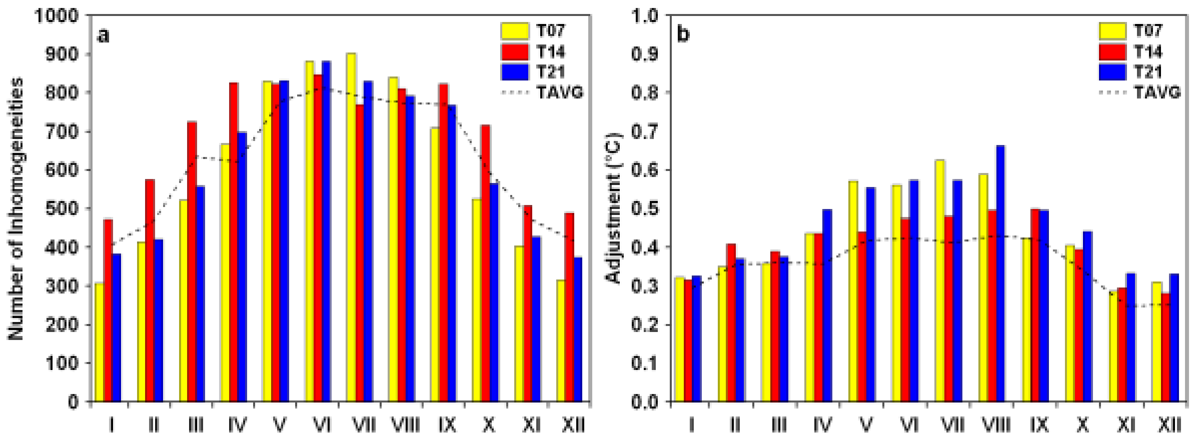

Fig. 7. a) Annual variation of the number of statistically significant inhomogeneities detected in air temperature series of observation terms (T07, T14, T21) and daily mean (TAVG) (Alexandersson's test – SNHT, bivariate test, reference series calculated using distances and correlations); b) annual variation of the size of corrections for individual observation terms and daily mean. Data refer to 230 stations analyzed in the Czech Republic and Slovak Republic in the 1961–2005 period.

4. Homogenization results for the Czech Republic

In the 1961–2007 period, 1750 series of seven climatological characteristics were tested and some inhomogeneities were found in 42% of them (*Table 1*). This value is underestimated, due to the low number of detections in precipitation series, in which breaks were detected in only 15% of series. For all other characteristics, this number is above 50%. The number of detected inhomogeneities varies according to the meteorological element (*Fig. 8*). For homogenization, just as for data quality control, the most problematic meteorological element is wind speed, where 75% of series were detected as inhomogeneous. Wind speed is a very specific meteorological element because, before automation, which took place since about 2000, it was estimated subjectively by observers using the Beaufort wind force scale.

Table 1. Number of breaks detected at meteorological stations in the Czech Republic in the 1961–2007 period for selected characteristics of meteorological variables: T – mean air temperature, TMA – maximum air temperature, TMI – minimum air temperature, SRA – precipitation total, E – mean water vapor pressure, F – mean wind speed, SSV – sunshine duration.

Meteorological element	**Number of series**	**Number of series with break**	**Ratio (%)**	**Number of breaks in series**			
				0	**1**	**2**	**3**
T	181	100	55,2	81	77	21	2
TMA	178	122	68,5	56	88	32	2
TMI	179	92	51,4	87	68	23	1
SRA	761	117	15,4	644	110	7	0
E	173	123	71,1	50	83	34	6
F	176	132	75,0	44	85	39	8
SSV	102	55	53,9	47	49	5	1
Total	1750	741	42,3	1009	560	161	20

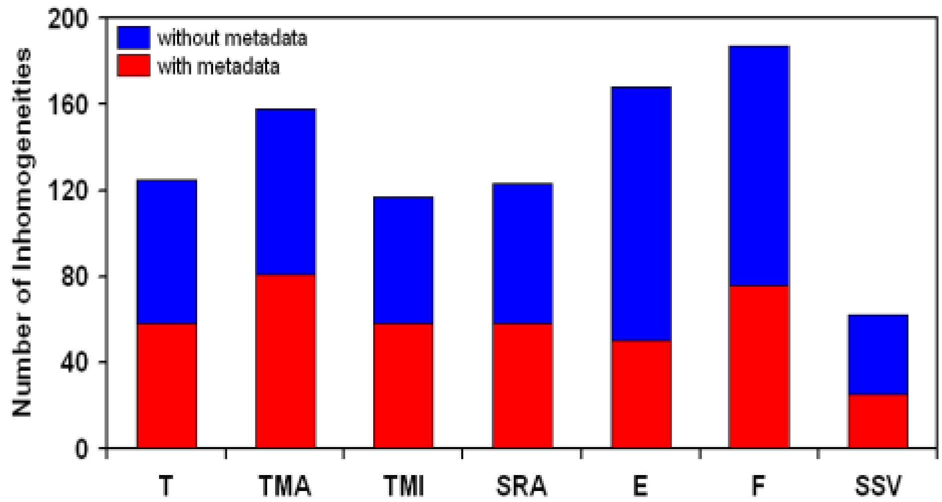

Fig. 8. Number of corrected inhomogeneities of selected characteristics of meteorological variables at stations in the territory of the Czech Republic in the 1961–2007 period: T – mean air temperature, TMA – maximum air temperature, TMI – minimum air temperature, SRA – precipitation total, E – mean water vapor pressure, F – mean wind speed, SSV – sunshine duration (the number of series tested for the individual characteristics is given in *Table 1*). Explanations for inhomogeneities: red – clarified by metadata, blue – no metadata.

For monthly values of air temperature and precipitation over the Czech Republic, the correlation coefficients between candidate and reference series are very high (median above 0.95 or 0.90, respectively; note that the rain-gauge station network is much denser than the climatological one). Along with mean wind speed, correlations are also very high in the case of the other characteristics.

As for inhomogeneity detection itself, more breaks occur in the summer months for air temperature and sunshine duration (the influence of relocation and other artificial changes is greater, resulting from the influences of the active surface, such as prevailing radiation factors and increased volume of vegetation), while for precipitation, it appears in the winter months (mainly due to problems associated with the measurement of solid precipitation). Water vapor pressure and wind speed do not show such a clear annual cycle (*Fig. 9*).

An annual variation is also clearly manifested in the correction of inhomogeneities. Considering the absolute values of corrections, the number of adjustments was higher during the summer months for temperature characteristics and water vapor pressure. After corrections, air temperature correlation coefficients increased mainly in the summer months and those for precipitation in the winter months. The largest increase in correlation coefficient after homogenization was observed in the case of wind speed.

The knowledge of metadata is an import factor for the proper correction of detected inhomogeneities. Out of all corrected breaks, 44% can be explained by metadata (*Fig 8*). There are some differences in the size of corrections according to the causes of the inhomogeneity: the size of correction was higher for inhomogeneities explained by metadata for all characteristics except minimum temperature and sunshine duration, where the mean size of corrections was similar to the case of missing metadata, and for precipitation, where it is even lower (however, for precipitation, only a small percentage of breaks can be detected). As it is evident from the results, the automation of measurements had a very strong influence on the homogeneity of series, as well as on the occurrence of outliers: for mean and maximum temperature and water vapor pressure, the size of the corrections was higher in the case of automatic measurements than the mean of over-all corrections. For example, the inhomogeneities in the series of maximum air temperature caused by automation are higher on average by 0.1°C than breaks not confirmed by metadata. Because the automation of measurements was introduced in the CHMI station network successively from the mid-1990s (see *Fig. 10*), it was possible to detect and make corrections without major problems.

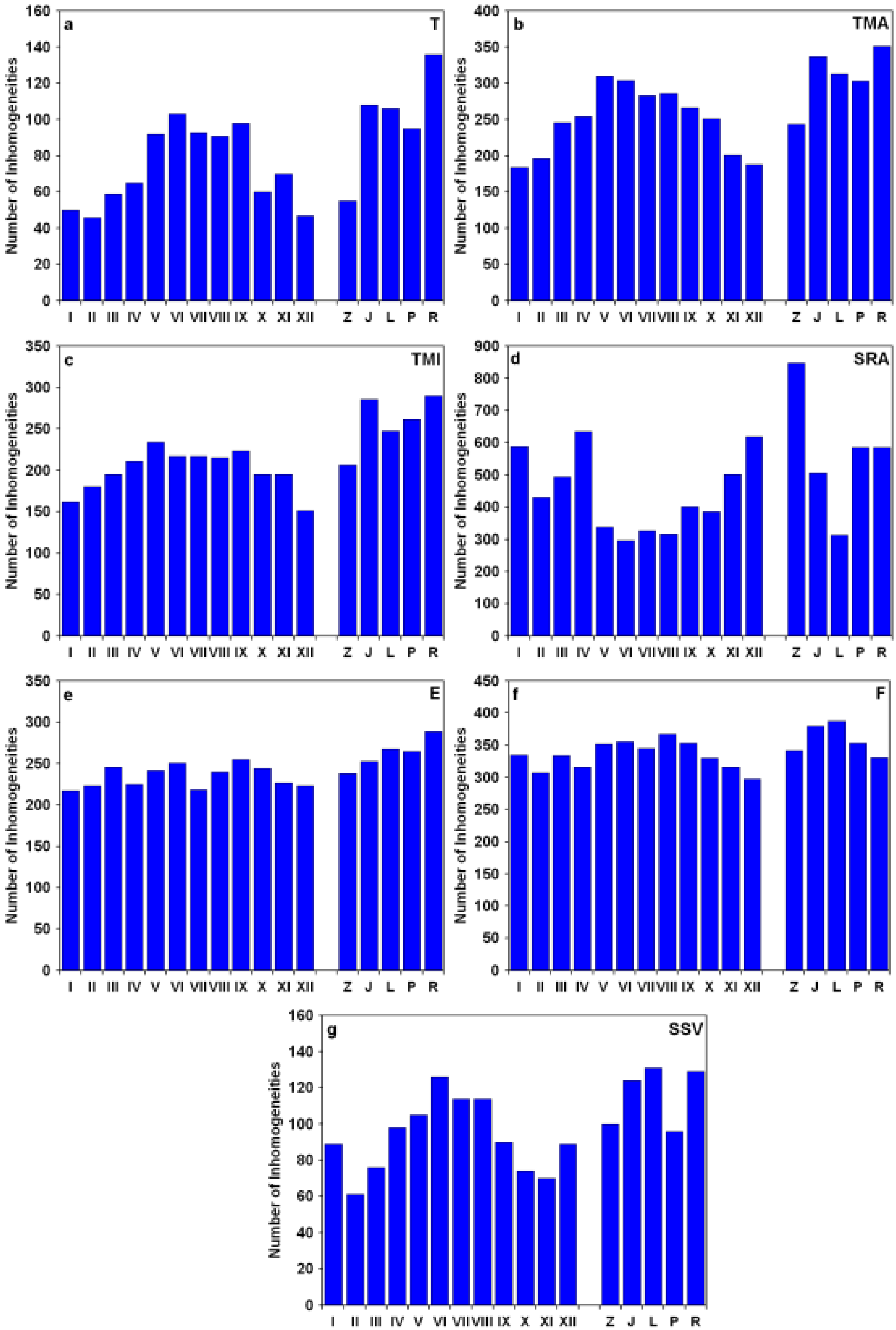

Fig. 9. Annual variation in the number of detected statistically significant inhomogeneities ($\alpha = 0.05$) for selected climatological and rain-gauge stations in the territory of the Czech Republic in the 1961–2007 period: a) mean air temperature, b) maximum air temperature, c) minimum air temperature, d) precipitation total, e) mean water vapor pressure, f) mean wind speed, g) sunshine duration (Z – winter, J – spring, L – summer, P – autumn; R – year).

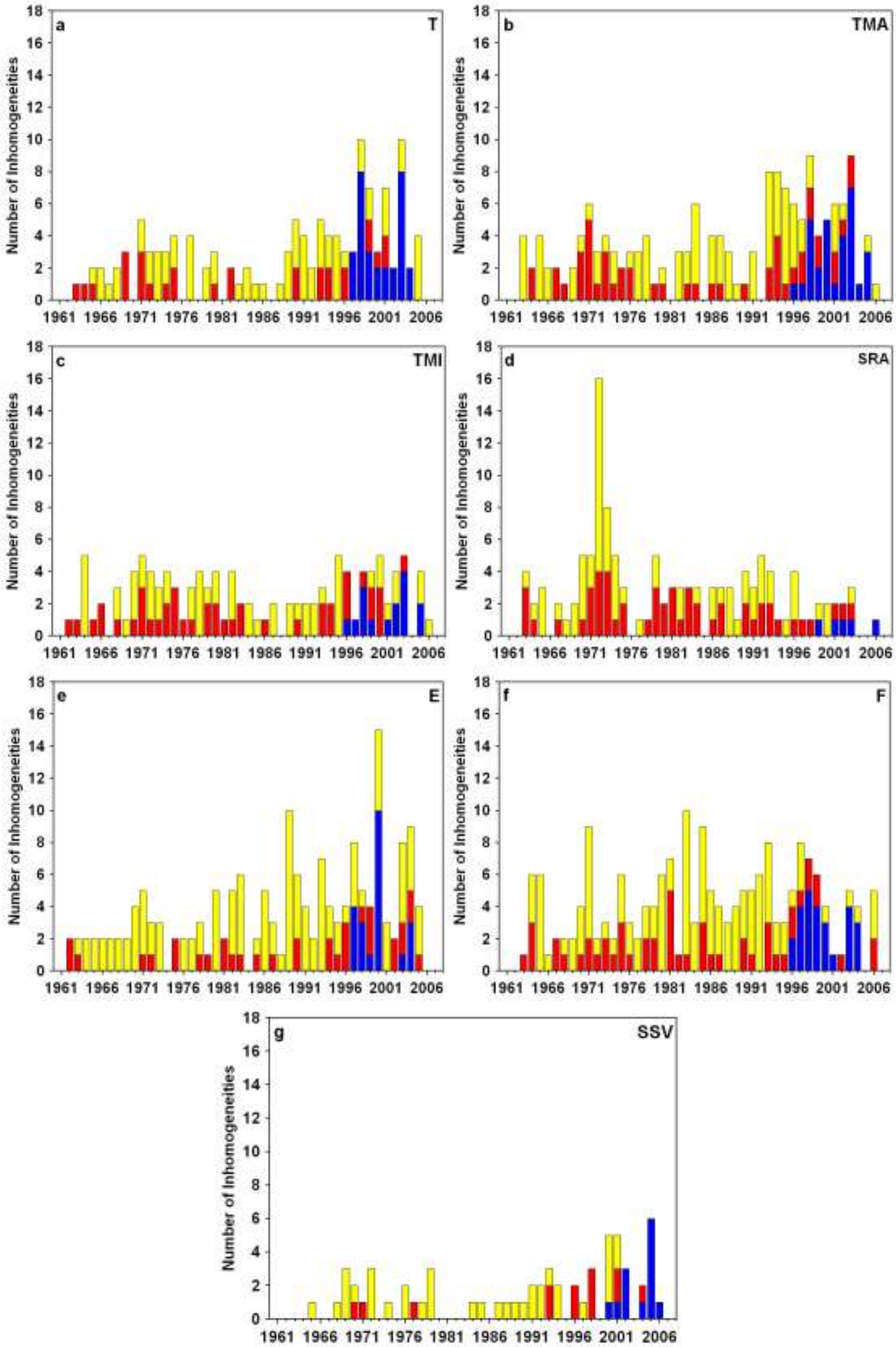

Fig. 10. Number of inhomogeneities detected in the series of climatological and rain-gauge stations in the territory of the Czech Republic in the 1961–2007 period: a) mean air temperature, b) maximum air temperature, c) minimum air temperature, d) precipitation total, e) mean water vapor pressure, f) mean wind speed, g) sunshine duration (yellow – break without metadata, red – break with metadata, blue – break with automation of measurements); AMS express a change to automatic measurements.

5. Technical series

Data quality control, homogenization and filling missing values lead to the creation of the so-called "technical" series for mean, maximum and minimum temperatures, precipitation totals, sums of sunshine duration, mean water vapor pressure, and wind speed. Such series may be used for further data analysis, because their values are consistent and complete over a given period. They were calculated for 268 climatological and 787 rain-gauge stations of the CHMI network in the 1961–2010 period, and actual values are continually added. Despite the fact that a smaller number of stations were available for some of the studied climatological characteristics (e.g. 196 stations for sunshine duration or 257 stations for water vapor pressure), "technical" series were completely calculated (for arbitrary station location or regular gridded network). In this way, we have a complex set of meteorological variables for each position of climatological stations, which can easily be further used (e.g., for evapo-transpiration calculation).

The possibility of calculating "technical" series for new positions, either in irregular or regular network, e.g. for grid points of regional climate model (RCM) outputs, allow their use for validation and correction of RCM outputs in each grid point. In the case of the RCM ALADIN-Climate/CZ (*Farda et al.*, 2010), series were calculated with 10x10 km resolution, specifically for 789 grid points over the Czech Republic (*Fig. 11*). The method for the "technical" series calculation is similar to the calculation of theoretical values during the data quality control (for more details, see, e.g., *Štěpánek et al.*, 2011).

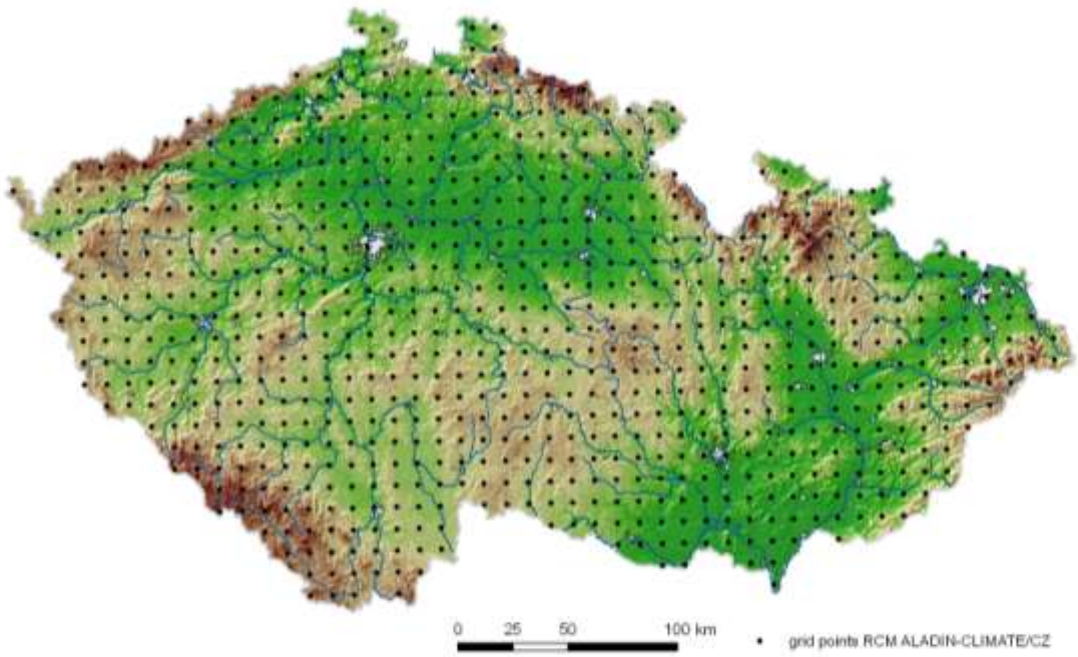

Fig. 11. Grid points of the outputs of RCM ALADIN-Climate/CZ for which "technical" series for the 1961–2009 period were calculated.

6. Conclusions

In the Czech Republic, experience with data quality control and homogenization has existed for several years. For data processing, software packages AnClim (*Štěpánek*, 2010a), LoadData and ProClimDB (*Štěpánek*, 2010b) were created. They offer complex solution, from tools for handling databases, through data quality control to homogenization of time series, as well as time series analyses, extreme value evaluation, and model output verification.

In this work, we summarize the effort and methodology behind outlier detection, series homogenization, and interpolation techniques for various climatological characteristics in the territory of the Czech Republic in the 1961–2010 period. In total, over 62 million values were data quality checked, for which the automation of the process was crucial. The final results are acceptable only because of the combination of several methods. The approach became part of the ProClimDB software (*Štěpánek*, 2010b). For correct outlier detection, it is necessary to work directly with measured values in the standard observing terms (e.g. 7:00, 14:00, 21:00 LMT), since possible errors can be masked in the "aggregated" values (daily means, monthly means, or sums).

Similar to the quality control, the aim of the created software for homogenization was to provide the user with support information for making quick, efficient, and correct decisions. Thanks to the COST Action ES0601 benchmark dataset, various parameter settings were checked in the software and recommended: finding neighbor stations using correlations of series of the first differences, performing homogeneity testing individually with each of the neighbors (pairwise comparison), weighting of the number of detected inhomogeneities for the homogeneity evaluation (weights of five for annual values, two for seasonal values, and one for monthly values). The correction of inhomogeneities was performed on a daily basis using our own approach (DAP –distribution adjusting by percentiles).

Quality control and correction of inhomogeneities have been performed on a daily (sub-daily) basis for all key meteorological variables over the territory of the Czech Republic in the 1961–2010 period. The homogenization of data before 1961 has only been carried out on the monthly basis so far (*Štěpánek*, 2003).

Due to the "technical" series calculated in both station and various regular grid point locations, we have gained a sufficiently large number of climatological series for subsequent analysis. These series are free of detectable outliers and inhomogeneities, have had their gaps filled, and are being applied to research in various projects in climatology and hydrology. From the "technical" series, we have also created maps for various meteorological variables for each month and day in the period of 1961–2010 (i.e., more than 130,000 maps).

Further steps will lead to the processing of series of individual observation terms and daily historical records before 1961, as well.

Acknowledgement–This paper was prepared with financial support from the Grant Agency of the Czech Republic for project No. P209/10/0605. The present work was also supported by the CzechGlobe project – the Centre for Global Climate Change Impacts Studies, Reg.No. Cz. 1.05/1.1.00/02.0073.

References

Alexandersson, A., 1986: A homogeneity test applied to precipitation data. *J. Climatol. 6*, 661–675.

Alexandersson, A., 1995: Homogenity testing, multiple breaks and trends. In: *Proceedings of the 6th International Meeting on Statistical Climatology*, Galway, Ireland, 439–441.

Aguilar, E., Auer, I., Brunet, M., Peterson, T.C., and *Wieringa, J.*, 2003: Guidelines on Climate Metadata and Homogenization. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organisation, Geneva.

Brandsma, T. and *Können, G.P.*, 2006: Application of nearest-neighbour resampling for homogenizing temperature records on a daily to sub-daily level. *Int. J. Climatol. 26*, 75–89.

Brázdil, R. and *Štěpánek, P.*, 1998: Kolísání teploty vzduchu v Brně v období 1891–1995. *Geografie – Sborník České geografické společnosti 103*, 13–30, (in Czech).

Brázdil, R., Štěpánek, P., and *Budíková, M.*, 1996: Homogenized air temperature series in Brno, 1891–1994. Zeszyty Naukowe Uniwersytetu Jagiellońskiego, MCLXXXXVI, *Prace Geograficzne 102*, 85–91.

Brázdil, R., Štěpánek, P., and *Květoň, V.*, 2000: Air temperature fluctuation in the Czech Republic in the period 1961–1999. Instytut Geografii UJ – *Prace Geograficzne 107*, 173–178.

Brázdil, R., Štěpánek, P., and *Květoň, V.*, 2001: Temperature series of the Czech Republic and its relation to Northern Hemisphere temperatures in the period 1961–1999. In: *Detecting and Modelling Regional Climate Change*(eds.: Brunet India, M., López Bonillo, D.), Springer, Berlin, 69–80.

Brunet, M., Saladie, O., Jones, P., Sigro, J., Aguilar, E., Moberg, A., Lister, D., Walther, A., Lopez, D., and *Almarza, C.*, 2006: The development of a new dataset of Spanish Daily Adjusted Temperature Series (SDATS) (1850–2003). *Int. J. Climatol. 26*, 1777–1802.

Conrad, V., Pollak, L.W.,1950: Methods in Climatology. Harvard University Press, Cambridge.

Costa, A.C. and *Soares, A.*, 2009: Trends in extreme precipitation indices derived from a daily rainfall database for the south of Portugal. *Int. J. Climatol. 29*, 1956–1975.

Della-Marta, P.M. and *Wanner, H.*, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate 19*, 4179–4197.

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global Planet. Change 57*, 16–26.

Easterling, D.R. and *Peterson, T.C.*, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol. 15*, 369–377.

Farda, A., Déqué, M., Somot, S., Horányi, A., Spiridonov, V., and *Tóth, H.* ,2010: Model ALADIN as Regional Climate Model for Central and Eastern Europe. *Studia Geophysica et Geodetica 54*, 313–332.

Klein Tank, A. M. G., Wijngaard, J. B., Können, G. P., Böhm, R., Demarée, G., Gocheva, A., Mileta, M., Pashiardis, S., Hejkrlik, L., Kern-Hansen, C., Heino, R., Bessemoulin, P., Müller-Westermeier, G., Tzanakou, M., Szalai, S., Pálsdóttir, T., Fitzgerald, D., Rubin, S., Capaldo, M., Maugeri, M., Leitass, A., Bukantis, A., Aberfeld, R., van Engelen, A. F. V., Forland, E., Mietus, M., Coelho, F., Mares, C., Razuvaev, V., Nieplova, E., Cegnar, T., López, J. A., Dahlström, B., Moberg, A., Kirchhofer, W., Ceylan, A., Pachaliuk, O., Alexander, L. V., and *Petrovic, P.*, 2002: Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *Int. J. Climatol 22*, 1441–1453.

Kyselý, J. and *Picek, J.*, 2007: Regional growth curves and improved design value estimates of extreme precipitation events in the Czech Republic. *Climate Res. 33*, 243–255.

Macková, J., 1997: Homogenizace dlouhých teplotních řad v České republice. Geografický projekt. Katedra geografie PřF MU, Brno. (in Czech)

Peterson, T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E. J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and *Parker, D.*, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *International J. Climatol. 18*, 1493–1517.

Potter, K.W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Weather Rev. 109*, 2040–2045.

Sacré, C., Moisselin, J.M., Sabre, M., Floria, J.P., and *Dubuisson, B.*, 2007: A new statistical approach to extreme wind speeds in France. *J. Wind Engineer. Indust. Aerodyn. 95*, 1415–1423.

Štěpánek, P., 2003: Homogeneización de las series de temperatura del aire en la República Checa durante el período instrumental. *Geographicalia 43,* 5–24.

Štěpánek P., 2010a: AnClim – software for time series analysis. Department of Geography, Faculty of Natural Sciences, MU, Brno, 1.47 MB (http://www.climahom.eu/AnClim.html).

Štěpánek, P., 2010b: ProClimDB – software for processing climatological datasets. CHMI, regional office Brno (http://www.climahom.eu/ProcData.html).

Štěpánek, P. and *Mikulová, K.*, 2009: Homogenization of air temperature and relative humidity monthly means of individual observation hours in the area of the Czech and Slovak Republic. In: (Eds: Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S.) *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases* (Budapest, Hungary, 29 May – 2 June 2006). WCDMP-No. 71. World Meteorological Organization, Geneva, 149–164.

Štěpánek, P., Řezníčková, L., and *Brázdil, R.*, 2009: Homogenization of daily air pressure and temperature series for Brno (Czech Republic) in the period 1848–2005. In: (Eds: Lakatos, M., Szentimrey, T., Bihari, Z., Szalai, S.) *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases* (Budapest, Hungary, 29 May – 2 June 2006). WCDMP-No. 71. World Meteorological Organization, Geneva, 107–122.

Štěpánek, P., Zahradníček, P., and *Huth, R.*, 2011: Interpolation techniques used for data quality control and calculation of technical series: an example of Central European daily time series. *Időjárás 115*, 87–98.

Štěpánek, P., Zahradníček, P., Brázdil, R., and *Tolasz, R.*, 2012: *Metodologie kontroly a homogenizace časových řad v klimatologii* (Methodology of quality control and homogenisation of time series in climatology - in Czech). ČHMÚ (in print) (in Czech)

Tolasz, R., 2008: *Databázové zpracování klimatologických dat*. Sborník prací ČHMÚ 52. Český hydrometeorologický ústav, Praha. (in Czech)

Trewin, B.C. and *Trevitt, A.C.F.*, 1996: The development of composite temperature records. *International Journal of Climatology, 16*, 1227–1242.

Vicente-Serrano, S., Beguería, S., Lopez-Moreno, J. I., García-Verac, M. A., and *Stepanek, P.*, 2010: A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *Int. J. Climatol. 30*, 1146–1163.

Vincent, L.A., Zhang, X., Bonsal, B.R., and *Hogg, W.D.*, 2002: Homogenization of daily temperatures over Canada. *J. Climate 15*, 1322–1334.

Wijngaard, J.B., Klein Tank, A.M.G., and *Können, G.P.*, 2003: Homogeneity of 20th century European daily temperature and precipitation series. *Int. J. Climatol. 23*, 679–692.