HOMOGENIZATION OF AIR TEMPERATURE AND RELATIVE HUMIDITY MONTHLY MEANS OF INDIVIDUAL OBSERVATION HOURS IN THE AREA OF THE CZECH AND SLOVAK REPUBLIC

P. Štěpánek[1] , K. Mikulová[2]
[1] Czech Hydrometeorological Institute, Regional Office Brno, Czech Republic
[2] Slovak Hydrometeorological Institute, Slovak Republic,
petr.stepanek@chmi.cz, katarina.mikulova@shmu.sk

ABSTRACT

Homogenization of monthly averages of air temperature and relative humidity has been carried out for the area of the Czech and Slovak Republics for the period 1961-2005. Because of presence of a noise in the series, statistical homogeneity tests give their results with some portion of uncertainty. Using various statistical tests along with various types of reference series made it possible to considerably increase the number of homogeneity tests results for each tested series and thus to assess homogeneity more reliably. Homogenization was performed on individual hourly observations and comparison demonstrating the improvement of results compared to the homogenization of daily averages was made. Air temperature and relative humidity series were compared to help identify to what extent multi-element processing can help improve the homogenization of individual elements. All data processing and analysis were carried out using AnClim and ProcClimDB (software developed for automatic processing, analyzing, homogeneity testing and adjusting of climatological data).

INTRODUCTION

Long time series often suffer from non-climatic effects. It became well known and accepted fact that such inhomogeneous or erroneous series can lead to biased results in climatological time series analysis. Inhomogeneities may occur when stations are relocated and by changes of observer, instruments and observing procedures. This type of information should be documented in station metadata but there are numerous cases where such metadata is incomplete or missing, so we can rely then only upon statistical test results. A large array of statistical techniques have been developed to detect inhomogeneities in climatological time series. Various methods and different countries approaches are described e.g. in Peterson *et al.* (1998) and Szalai et al. (1999, 2004).

Different tests often identify different inhomogeneities (particularly for smaller amount of change), application of different types of reference series usually leads to different results as well, differences occur also among individual monthly, seasonal or annual series. In a number of instances, particularly where coincide detections in series were identified, adjustments can often be clearly justified, even in the absence of metadata.
The existence of good quality reference series is very important for the detection of real inhomogeneities. That is why it is useful to include as many series for a particular area as possible, to help identify series abnormalities. In the case of air temperature, spatial correlations decrease with distance quite slowly, and so, it makes sense to analyze large areas. In this study the homogenization of both Czech and Slovak Republics series was considered useful and appropriate. To further increase the quality of homogenization, the number of test results were increased by testing monthly means of individual observation

hours (i.e. those taken at 07:00, 14:00, 21:00 hours local time). Additionally, two meteorological elements, air temperature and relative humidity were mutually compared for detected inhomogeneities and were used when considering final adjustments of series.
Single shift inhomogeneities are the most frequent in climatological time series and easiest to detect. Only single shift inhomogeneities were examined and adjusted for in this study so far. Data formatting and processing was performed using database software called ProcClimDB (Štěpánek 2006b). Homogeneity testing and time series analysis was conducted on AnClim software (Štěpánek, 2006a)..

DATA CHARACTERIZATION

The Czech and Slovak Republics cover a total area of 128 km^2. Both Republics are mountainous. The Czech Republic ranges from 115 m to 1602 m at its highest peak (Sněžka). Despite being smaller, the Slovak Republic has a much greater height range, ranging from 94 m to 2655 m at Gerlachovský štít. From west to east the climatic influence of ocean diminishes and the continental influence progressively increases.

Air temperature and relative humidity were analyzed as series of monthly means of observations taken at 07:00, 14:00 and 21:00 hours local time, and daily averages. Stations with a minimum length of 25 years were selected. For the period since 1961, 230 stations measuring air temperature and 217 stations measuring relative humidity were available. The mean minimum distance between stations is 18.6 km, and mean altitude 448 m, (median 380 m). Nine stations are situated above 1000 m a.s.l., and 4 above 1500 m a.s.l.

Because of processing of large number of stations, different methods of homogenization (e.g. different reference series, different tests) were examined in smaller areas such as Southern Moravia (Czech Republic) and Western Slovakia (Slovak Republic), see Fig. 1. From the results of these areas, the most useful types of reference series and homogeneity tests were selected and consecutively applied to the whole area of the Czech and Slovak Republics.

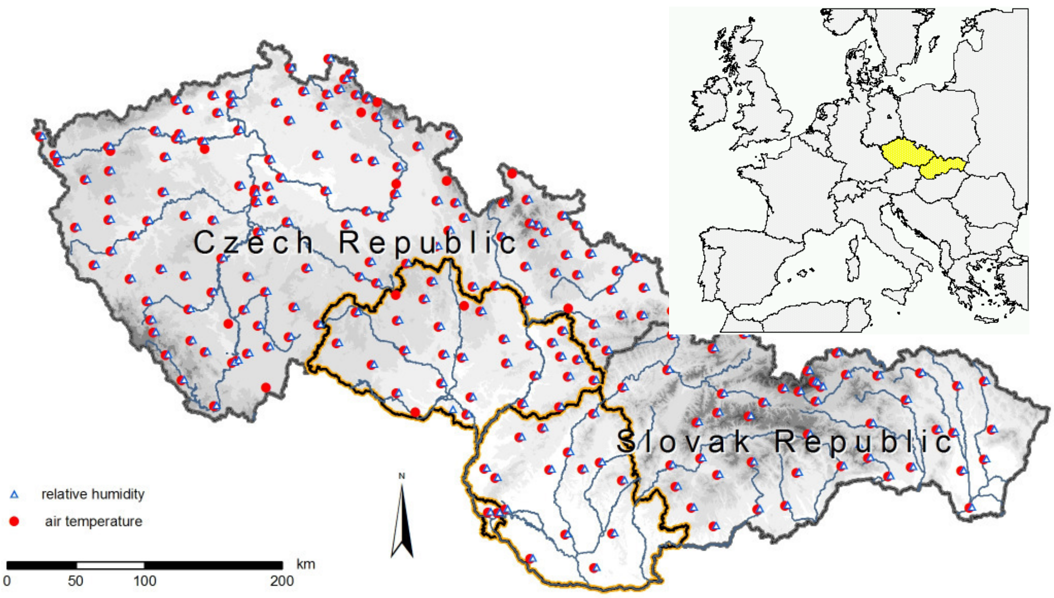


Fig. 1. Area of interest (Czech and Slovak Republics) with marked borders of the test area: Southern Moravia and Western Slovakia. Right top: location within Europe (at a different scale)

Air temperature was found to correlate very well throughout the Czech and Slovak Republics (see Fig. 2). Medians of correlation coefficients (from all the stations) vary only around 0.9 for all months in case of observation hour 14:00,

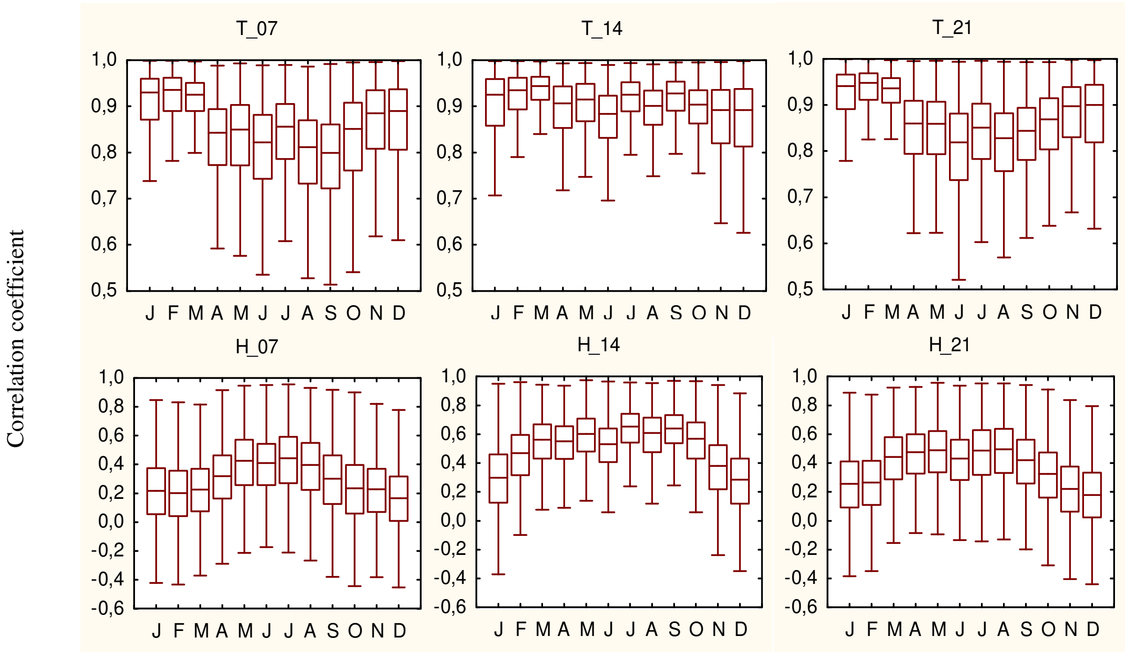


Fig. 2. Correlation coefficients for individual observation hours (07:00, 14:00, 21:00), for air temperature (T) and relative humidity (H) (using 25.420, resp. 22.595 station pairs - values)

and drop to 0.8 in case of hours 07:00 and 21:00. Correlations were lower in summer months and higher in winter. Values of correlation coefficients for daily averages are comparable with the hour 14:00, i.e. values of medians vary around 0.9, and during winter they are even higher. Relative humidity correlates better in summer than in winter, again the best for the hour 14:00. Daily averages are comparable with the hour 14:00, and their values of correlation coefficient are similar or higher than in case of individual observation hours. Relative humidity correlations decrease relatively quickly with distance, but the stations network was sufficiently dense to create a well correlated reference series (see Fig. 8).

1. METHODOLOGY DESCRIPTION

In the case of series with missing or incomplete metadata, only statistical tests for homogeneity are relied upon to indentify inhomogeneities. Unfortunately using solely the results of statistical tests during homogenization is problematic due to the fact that the detected year of inhomogeneity is often given with some error, or not identified. Štěpánek (2004) demonstrated that the determination of the correct year of inhomogeneity for air temperature, where the difference was less than 0.5°C, occurred in less than half of the cases. In the remained of cases, false years of inhomogeneity were given, or the years were not detected. According to this result, inhomogeneities less than 0.5°C are likely to be difficult to detect.

Because of this uncertainty in the result of homogeneity testing, it was attempted to increase the reliability of inhomogeneity determination through processing as many test results for each candidate series as possible. Series of individual observation hours were used, and several statistical tests for homogeneity were applied, various types of reference series were calculated for each candidate series, and monthly series as well as seasonal and annual averages of the series were tested. By combining all of these it was possible to considerably increase the number of test results for each tested series, thereby increasing the reliability of the homogenization process. Through the statistical processing of a large number of test results, it was possible to calculate the probability of each inhomogeneity of a given series (probability calculated as a portion of count of detected inhomogeneities - for each year, group of years or whole series - as an amount of all theoretically possible detections). One of the advantages of this approach is that a sum count of all detected inhomogeneities out of all the theoretically possible detections in the series can be used for assessing quality of measurements of a particular station as a whole.

Processing of the series during quality control and homogenization included the following steps: detection, verification, where necessary the correction of outliers (extreme values), creation of reference series (various ways), homogeneity testing (using 3 homogeneity tests), inhomogeneities (years) determination according to test results and metadata, adjustment of inhomogeneities and, only at the end, filling missing values. These steps are outlined in the Fig. 3. and are further discussed in the text.

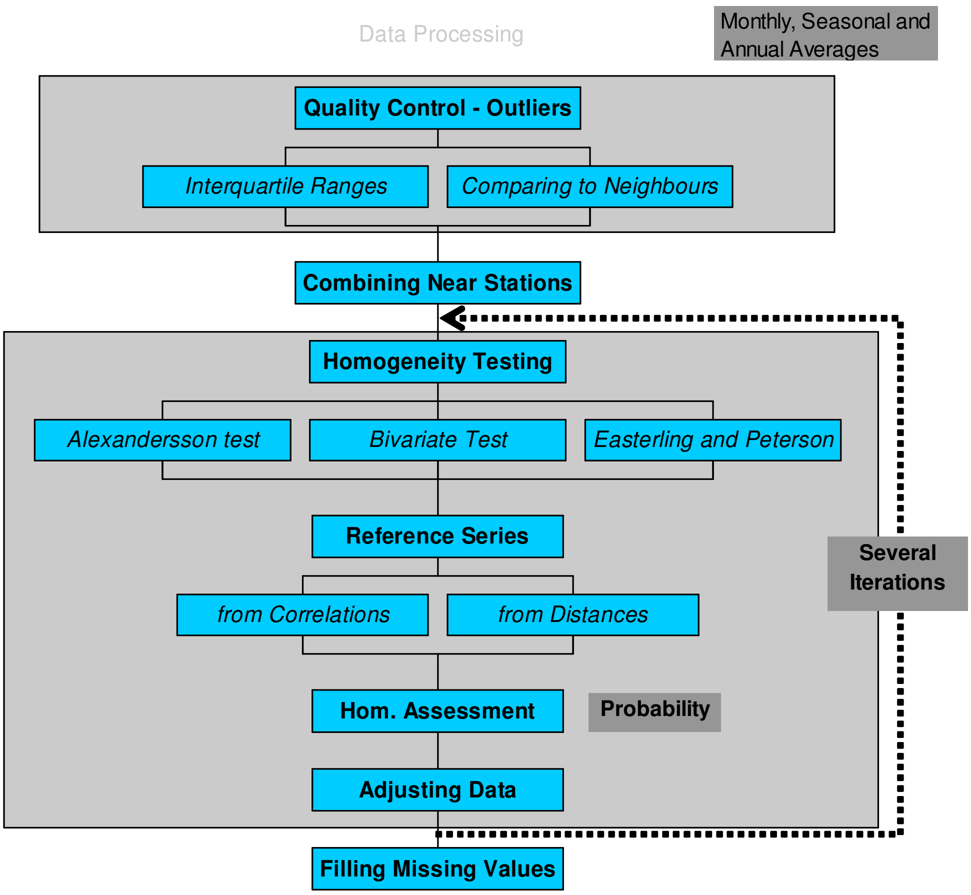


Fig. 3. Scheme of data processing during quality control and homogenization of the series

1.1. OUTLIERS IDENTIFICATION

Data quality control was carried out in two ways: by applying limits derived from interquartile ranges (it can be applied either to individual monthly series, or preferentially, by difference series between candidate and reference series), and, secondly, by comparing values to values of neighbouring stations

Where comparing neighbouring stations, the five best correlated neighbors were selected (correlations calculated form first difference series, see e.g. Peterson 1998), the values of correlation coefficients being at least 0.5, no limit for distance nor for altitude difference was applied. Only series with the same element and observation hour were selected. For outliers evaluation, the following characteristics were considered.

Counts of statistically significant different neighbours (compared to base station) exceeding confidence limit (0.95) were evaluated from difference series (neighbour and base station), the differences standardized to zero mean and standard deviation equal to one (to enable using standardized normal distribution), for each base station and month individually. Cases, where more than 75 % of neighbours significantly differed from the base station value, were visually checked. To help depict outliers, the values of neighbours were standardized with respect to base station average and standard deviation and also a new (theoretical) value for the base station was calculated - as weighted average from the standardized values of the neighbours (using 1/distances as weights, with power 1 which seems to be sufficient in case of air temperature, and 2 in case of relative humidity). Further, coefficient (multiply) of interquartile ranges (q75-q25) above q75 (or below q25) were evaluated (calculated from the standardized neighbours values), and applied to base station value. The reason for this was to assess the similarity of used neighbours values with regard to the outlier test value: the more values of neighbours are similar, the higher is the value of the coefficient.

The final decision on the removal of outliers was based on a percentage count of significantly different neighbours, the difference from the "expected value", coefficient of interquartile range and finally value was conducted by visual (subjective) comparison of the standardized values of neighbours with the base station value.

1.2. COMBINING NEAR STATIONS MEASUREMENTS

In order to produce longer time-series, the neighboring station measurements (within 15 kilometers in case of temperature, and 10 kilometers in case of relative humidity, nearer stations having preference) was merged into one. A maximum gap of two combined series was allowed to be 4 years, the minimum length of reconstructed series was 25 years. In this instance, 14 stations were recorded and the year of merging then used as metadata information during series homogenization.

1.3. HOMOGENEITY TESTING

The AnClim software (Štěpánek, 2006a) was used to identify the inhomogeneities applying following tests for relative homogeneity (significance level α=0,05) on monthly, seasonal and annual data:

- Alexandersson test (SNHT for a single shift) (Alexandersson 1986, 1995)
- Bivariate test of Maronna and Yohai (Maronna and Yohai 1978, Potter 1981)
- Easterling and Peterson test (Easterling and Peterson, 1995)

To ensure that only one inhomogeneity was present in series when using Alexandersson or Bivariate test, a further modification was introduced into the AnClim software, which divides the series at the position of the found inhomogeneity and test the parts before and after the detected inhomogeneity separately. If no additional inhomogeneity was found in these two parts, we can rely upon the results of the given test for a whole length of the series (especially the significance of a test statistic).

These tests were applied for the whole studied area (the Czech and Slovak Republics). For the tested area, additional tests were also used to study differences in detection capabilities of individual tests and influence of various types of reference series (see chapter 4.2).

1.4. REFERENCE SERIES CALCULATION

In order to increase the number of homogeneity test results and thus better assess inhomogeneities in the series, two different calculations of reference series were performed:

- as an average from selected stations based on correlations
- as an average from selected stations based on distances

Each of these types of reference series has both advantages and disadvantages. By using correlations, the reference series created is the most similar to tested series (and thus suppressing variability in the differences/ratios series the best), but stations with similar inhomogeneities to the tested series can be selected. However this effect can be minimized by using first difference series for calculation of correlation coefficients, then inhomogeneities are manifested in one value (see e.g. Alexandersson, Moberg, 1996, Peterson 1998). For the latter type of reference series, by using distances, the geographical vicinity of the selected stations are preserved, but different climatic conditions even for near stations (due to different altitude etc.) can occur. Differences between reference series are further discussed e.g. by Mikulova and Stepanek (2004) or Stepanek (2005).

Weighted averages of neighbour series for reference series calculation were applied. The values of the selected neighbour stations were standardization to base station average and standard deviation to avoid problems with biased reference series. This can often happen in cases of missing data in of one of the neighbour series. No transformation of values (in case of air temperature and relative humidity) was applied to data.

In the first stage, a list of proposed neighbour stations was obtained. The list was subsequently checked, comparing correlation coefficients, distances, and also difference in stations altitude. This approved list was then finally used for the reference series calculation.

1.5. ASSESSMENT OF DETECTED INHOMOGENEITIES

The main criteria for determining the year of inhomogeneity was the probability of the given inhomogeneity, i.e. count of detections of a given year from all the testing of a given station expressed relatively to count of all theoretically possible detections. For detected inhomogeneities, a limit of 20% of all possible detections was used in cases where there was no information in metadata about the change. A limit of 10-15% was sufficient in cases where the inhomogeneity was in agreement with metadata. The count of detections for groups of years was also taken into account (some inhomogeneities started during the course of a year and thus manifested in 2 years at least). In cases where there was no mention in the metadata concerning a detected shift (which was most common), other sources information were used. Distribution of the given year within individual months or seasons, graphs of differences with reference series and some other characteristics, were all used for deciding whether the undocumented inhomogeneity could be regarded as "undoubtfully" proven and thus be corrected.

The mentioned decision limits were estimated subjectively, from selected set of stations, so that only clear inhomogeneities were corrected (an aim being not to "over-homogenize" the series). These limits are appropriate for the studied elements series of the analyzed area, for other elements or areas these would have to be determined according to the given purposes.

1.6. ADJUSTMENT OF INHOMOGENEITIES

Adjustment of the detected inhomogeneities was carried out by means of reference series calculated as an average of five stations with the highest correlation coefficients to the series being adjusted (correlations were calculated again from the first difference series). The amount of change was estimated as a difference between averages calculated from difference series between the candidate and reference series from a period taken 20 years before and 20 years after the year being adjusted. The period was truncated in case another inhomogeneity within the period was encountered. These adjustments were applied to all monthly data. Where possible, the start of inhomogeneity was determined to a particular month.

Inhomogeneities within 4 years of the end of a series could not be adjusted. This happened relatively often in recent years, because of the transition to automatic measurements (being successively introduced since 1995). The parts of series with inhomogeneities near the ends of series had to be removed from further processing.

Various characteristics were analyzed before applying proposed adjustments including: increment of correlation coefficients between candidate and reference series after the adjustments, change of standard deviation in differences before and after the change, presence of linear trend, etc. In case of doubt, adjustments were not applied and the respective series was considered for removal from further processing.

Estimated adjustments are influenced by random errors in the series. To produce a smoother and physically more justifiable annual course of adjustments, weighted averages of the adjacent months were applied (with weights 1:2:1).

1.7. FURTHER CONSIDERATIONS

The above-mentioned steps (creating reference series, homogeneity testing, assessing and adjusting possible inhomogeneities) were performed in several iterations. In each iteration more precise results were obtained. The final adjustments of inhomogeneities were estimated from original data, taking into account inhomogeneities detected in all the previous iterations. It was necessary to use original data for the final correction (but used reference series were calculated from adjusted series in the last iteration), so that the final adjustments were estimated using periods without any inhomogeneities (a period taken for an adjustment was truncated when there was found another inhomogeneity in the series).

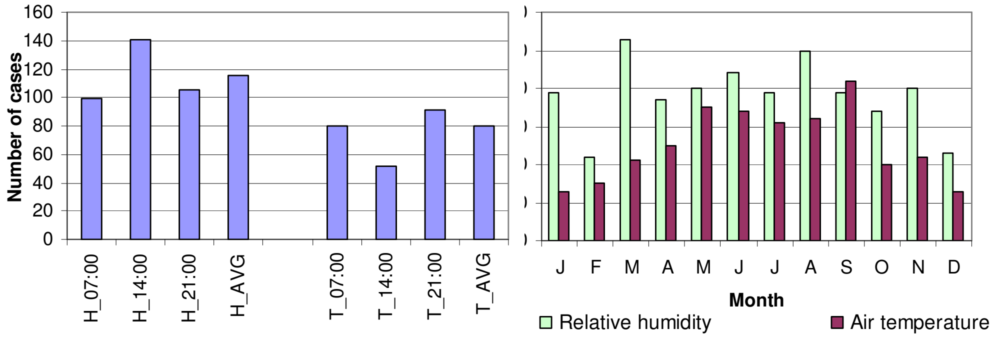
The filling of missing values was performed only after homogenization and adjustment of inhomogeneities in the series. The reason for this was to enable the new values to be estimated from data not influenced by possible shifts in the series. Moreover, when missing data are filled before homogenization, they may influence correct inhomogeneity detection (above all when a gap is longer than one year and there is an inhomogeneity - change of mean - near the position of the missing value). Filling the gaps was done by means of linear regression between filled value series (dependent variable) and a reference series (independent variable). Reference series were calculated as an average of five stations with the highest correlations with respect to the series with filled value. For the linear regression model, values 20 years before and 20 after the value being filled were used. Again, for assessing quality of the process, various statistics were monitored, e.g. differences of averages and standard deviations in periods before and after the gap.

A remaining question on the homogenization is the influence of the transition to automatic measurements which started in the station network of the Czech Republic in 1997 and in the Slovak Republic since 1995. For some of the series it is too early to assess the influence of automation, for other stations the effect was already very well detectable. The crucial problem is that the change do not manifest in the only data characteristic, such as arithmetic mean, but rather influencing several properties of the series, and moreover, these changes in series properties can not be usually linked linearly to the previous segment (before the change). That is why we need to possess long series (after the breaks) to be able to detect all the possible influences and, last but not least, to invent appropriate approaches for such analysis. Another difficulty is that there exist only few stations with comparative measurements, and stations from other sites are influenced by several other factors that are problematic to enumerate and thus to suppress.

2 SERIES FOR HOMOGENIZATION

2.1 Finding and removing outliers

For outlier identification, the approach described earlier in chapter 3.1 was applied. We tried to detect causes of anomalous monthly data by tracing the problems in daily data (the same method as mentioned above, applied to daily data within individual months), but due to huge database of processed values and shortage of time, we were not able to check all the detected values. This area will be the subject of further work. For the purposes of this study, approved errors were removed from further processing and were replaced by missing values. Considerably higher count of outliers occurs during summer months, in both processed elements (see Fig. 4)



.

Fig. 4. ***Left*: count of removed outliers for individual observation hours, for relative humidity (RV) and air temperature (T). *Right*: count of removed outliers for individual months, for relative humidity and air temperature**

2.2 Verification of homogeneity detection procedure in the selected area

Due to the huge database (1.2 millions of monthly values for all the elements and observation hours within the whole area of the Czech and Slovak Republics, compared to “only” 240.000 months in the tested area), some approaches were tested in smaller areas, namely Southern Moravia and Western Slovakia (see Fig. 1).

We analyzed various types of reference series. These were created either by means of correlations or distances (using either simple or weighted mean), using either original series or series with removed outliers, using either the same element and observation hour of the neighbour stations as was that of candidate or using arbitrary observation hour from neighbours, or by using monthly or seasonal and annual averages. Altogether 18 different reference series were analyzed for each candidate. Reference series from distances and using arbitrary elements and observation hours of neighbours were not calculated because this option makes no sense.

Correlation coefficients for the different reference series are shown in Fig. 5. In case of air temperature, the medians of correlations are very similar, in winter there is no difference, in the summer months a maximum difference of five hundredths. The reference series for relative humidity differ, on the contrary, rather in the winter months, during the summer months the differences are again very small.

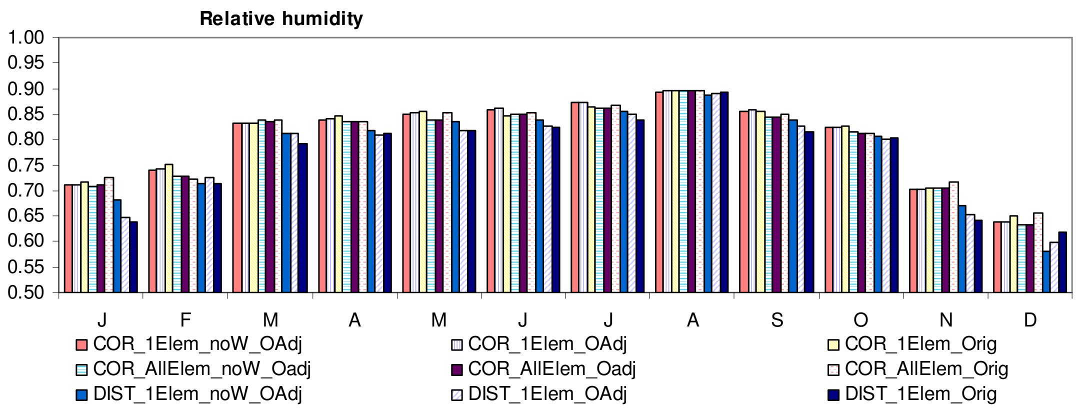


Fig. 5. Correlation coefficients (medians, calculated from 47 values) between various types of reference series and the tested series, for individual months. For explanation of used names of reference series see Fig. 6

Homogenization results for the questioned reference series and the Alexandersson test are shown in Fig. 6. The highest count of detected inhomogeneities occurs in the reference series created by means of distances, the same element and observation hour, using original series and unweighted mean. It is clear that this is caused by higher portion of random error in the tested candidate – reference difference series compared to the other types of reference series. Generally, series where outliers have not been removed give a higher count of detected inhomogeneities. The lowest count of detected inhomogeneities occurs in the reference series created by means of correlations. In Fig. 6 we can also see portion of the count of corrected to all detected inhomogeneities, which can be used for evaluation of the best reference series. The higher values of the portion are achieved by reference series created from correlations; the most efficient method to find true inhomogeneities is by using seasonal and annual averages for reference series calculation, series with removed outliers, and the same meteorological element and observation hour of the neighbour stations as in the tested series.

Following from the presented results, for the homogenization of the series within the whole area of the Czech and Slovak Republics, reference series created both from correlations and distances, using neighbours with the same element and observation hour as that of candidate series and using weighted average were applied.

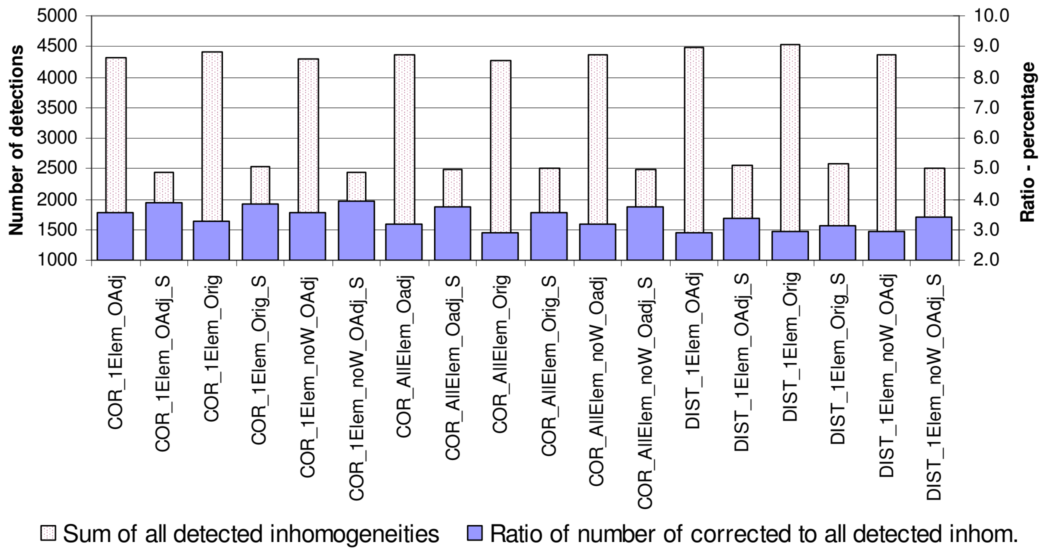


Fig. 6. Count of detected inhomogeneites for various types of reference series and ratio of the number of corrected inhomogeneities (100 adjusted elements and observation hours) to count of all detected inhomogeneities. *COR* – reference series created by means of correlations, *DIST* – by means of distances. *1Elem* – reference series created using neighbours with the same element and observation hour, *AllElem* – neighbours with different hours can be selected. *noW* – simple mean from neighbour series, otherwise weighted mean is used. *OAdj* – series with removed outliers, *Orig* – original series (no removal of detected outliers). *S* – seasonal and annual averages, otherwise monthly averages are used. Results for Alexandersson test

Various types of reference series give slightly different results due to random error present in the series. Where detections coincide, it is possible to better rely upon the test results. Including other types of reference series should lead only to a small improvement, therefore it seems more appropriate to gain further test results through testing individual monthly, seasonal and annual averages, testing individual observation hours etc.

2.3 Homogenization results for the whole studied area

As mentioned in chapter 3.1, we have used three tests for homogeneity, two types of reference series, monthly as well as seasonal and annual averages.

For the limits in which the values of correlation coefficients between candidate and reference series vary, for daily averages, see boxplots (i.e. median, lower and upper quartile and limits for outliers) in Fig. 7. Comparison of individual observation hours is shown in Fig. 8. For air temperature, the lowest values of medians (0.95) occur during summer months (for the hours 07:00 and 21:00), for the hour 14:00 they do not drop below 0.98, the same occurs for daily averages. In case of relative humidity, the correlation coefficients with reference series are markedly lower, mainly during winter months, but still usable for homogeneity testing. Again, values for the observation hour 14:00 are higher than those for 07:00 and 21:00, being comparable with daily averages.

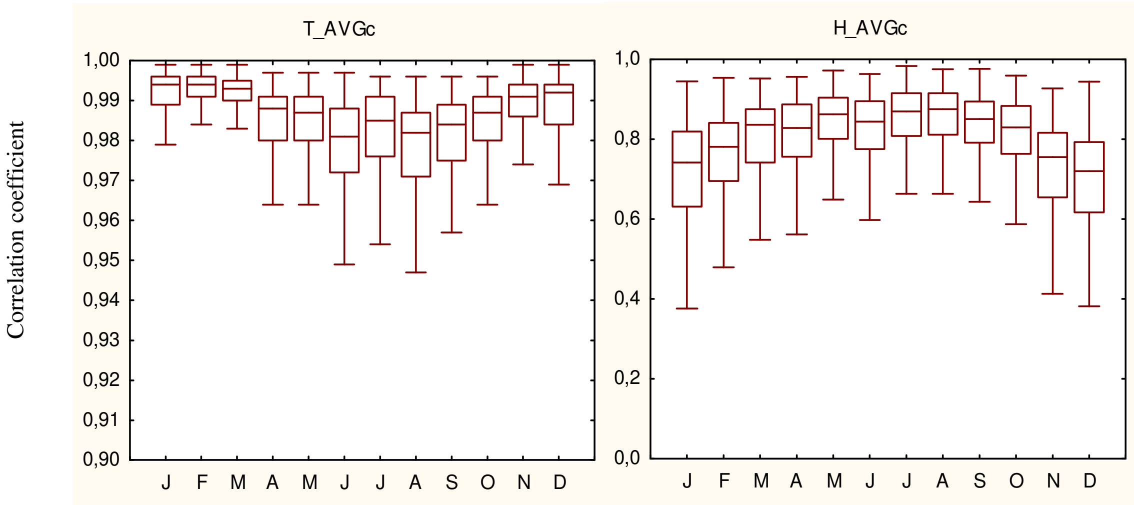


Fig. 7. Boxplots for correlation coefficients between candidate and reference series, applying correlations for reference series calculation, for air temperate (T, *Left*, from 227 values in each category) and relative humidity (H, *Right*, from 214 values)

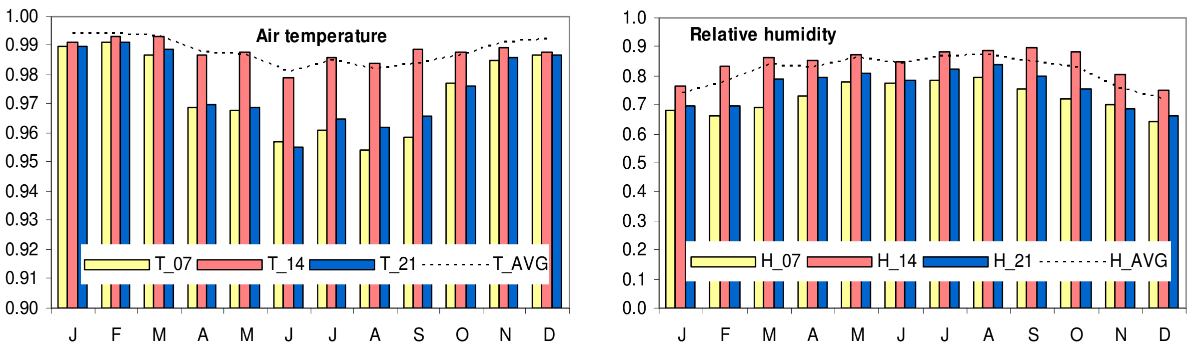


Fig. 8. Correlation coefficients between candidate and reference series (medians), for individual observation hours, applying correlations for reference series calculation (from 227, resp. 214 values for each category)

Fig. 9 shows the count of homogeneity tests detections for individual observation hours. Since data from the same place were used, both for air temperature and relative humidity, it is clear that inhomogeneities are more obviously detected in the air temperature series. One conclusion is that, effects such as station relocation, manifest more profoundly in air temperature series (and within air temperature mainly during summer months). Nevertheless, an important role can also be played by lower correlations between candidate and reference series in case of relative humidity.

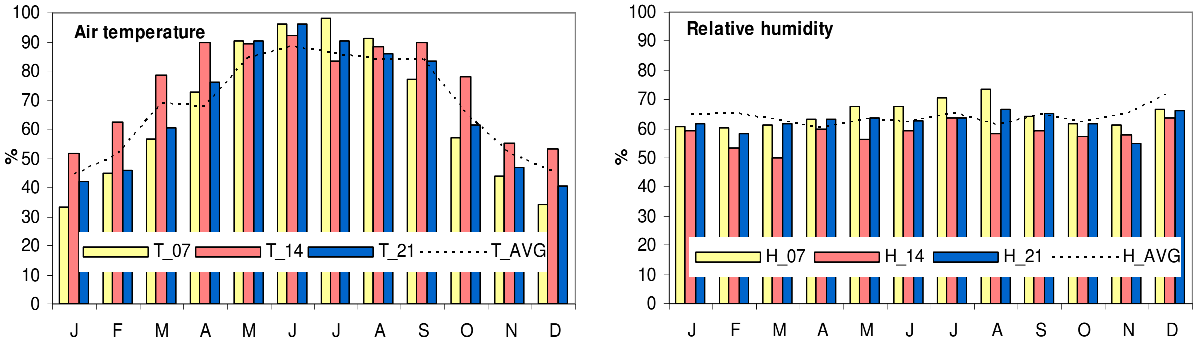
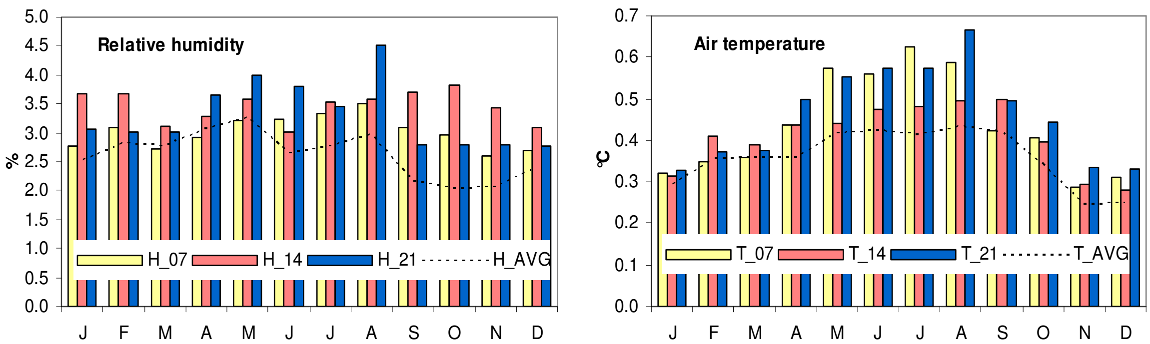


Fig. 9. Percentage of significant inhomogeneities (0.05) detected by used tests (SNHT test, Bivariate test, reference series created both by means of correlations and distances, altogether) related to the total number of series used. For individual months

In individual tests, the count of detections in individual years for Alexandersson, Bivariate and Easterling and Peterson tests is similar, if we do not consider ends of the tested series. The similarity of Easterling and Peterson test with Alexandersson one was a surprise because during simulations (for series with properties of air temperature series in the Czech Republic) that have been done previously (presented e.g. in Stepanek, 2005), the Easterling and Peterson test gave many more inhomogeneities, compared to Alexandersson test, that had no justification. This field is highlighted for further study, to better assess the relative power of these tests.

Counts of inhomogeneity detections – years – for various types of reference series show the same fluctuations, but for references series created by means of distance, the count of detections is about 10 percent lower for air temperature and 20 percent lower for relative humidity compared to reference series created by means of correlations.

Detected inhomogeneities were the subject of careful control before accepting final adjustments; this primarily took account of metadata, but also a number of auxiliary characteristics (see chapter 3.5). Fig. 10 shows final adjustments applied to the processed series. For air temperature, lower values of adjustments were applied for the hour 14:00 compared to the hours 07:00 and 21:00. The same behavior is valid for the improvement of correlation coefficients between candidate and reference series after realizing the adjustments. Relative humidity adjustments differ for individual observation hours and part of a year, e.g. in summer the highest values of adjustments occur for the hour 21:00, while in winter for the hour 14:00, the same course can be seen repeated for the correlations improvements after realizing adjustments.



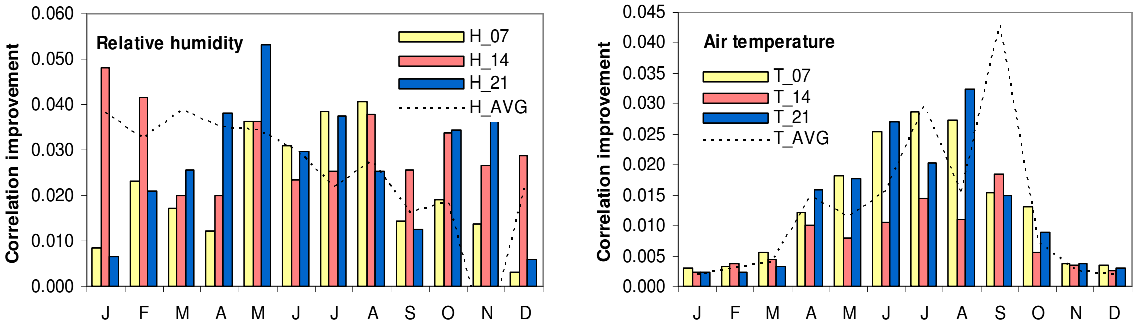


Fig. 10. Adjustments - averages of their absolute values and improvements of correlation coefficients (between tested and reference series) after realizing the adjustments, for air temperature and relative humidity, for individual months (using 40, resp. 32 values for each category)

In respect of air temperature (counts of detected inhomogeneities and amount of adjustments) the difference between summer and winter months can be explained by different influence of active surface upon the formation of air temperature regime in these distinct periods. In winter, prevailing circulation factors and reduced vegetation ensures the influence of effects leading to inhomogeneities (e.g. station relocation) is smaller, while in summer, resulting from prevailing radiation factors and increased volume of vegetation, the influence (of relocation for instance) is greater. The role of different active surfaces are also manifested in the fact that values of correlations are higher in winter months (climate conditions are similar for larger areas) in comparison with summer. Due to this characteristic, correlations were improved mainly in summer, in winter months the effect of homogenization was smaller (both for the adjustments and correlations improvement).
Relative humidity is a complex meteorological element influenced by many factors including air temperature, precipitation, wind speed, evaporation. This has the effect that explanation of inhomogeneity manifestation throughout a year is much difficult than in the case of air temperature. For instance, precipitation in the analyzed area is effect by station relocation primarily in winter, mainly due to the larger error in measurements connected with solid precipitation (manifested both in count of detected inhomogeneities and amount of adjustments).

Homogenization results obtained for air temperature can be generalized to a wider area (outside the analyzed area), since the spatial correlations of air temperature decrease with distance slowly. Contrary, relative humidity correlations decreas rapidly with distance, so the presented results (annual course of inhomogeneities characteristics, etc.) can differ outside this study area.

3. SUMMARY

Our results indicate that analyzing series of individual observation hours improves the detectability of inhomogeneities. This is because inhomogeneities manifest in the different series in a different way: count of inhomogeneities, amount of change, correlations between reference and tested series (and thus delectability of inhomogeneities).

If we compare individual observation hours, inhomogeneities are better detected for the individual observation hour 14:00 – mainly because of higher correlations between candidate and reference series. Since inhomogeneities may manifest only in one of the observation hours and thus be masked in daily averages, performing homogenization on both individual observation hours and daily averages at the same time is recommended.

Valid for both the processed elements: station relocation and other effects that lead to inhomogeneities in the series are more profoundly manifested in summer months than in winter months in this study. In the case of air temperature, large differences of the used inhomogeneity characteristics occurred between individual months; the annual courses for relative humidity are, on the contrary, found to be much smoother. For relative humidity, the height of correlation coefficients (in summer) coincides well with the count of detected inhomogeneities, amount of adjustments, and correlation improvement (after adjustments). Inhomogeneity characteristics for air temperature were found to have a different annual course: higher correlations were found in winter but the count of detected inhomogeneities and amount of change were highest in summer.

Data processing and analysis was conducted using ProClimDB and AnClim software. This software is available from a server at http://www.klimahom.com/software/. Ongoing development of this software, e.g. connection with R software, is planned.

Acknowledgements

The authors would like to acknowledge Mr. Andy Moir, Brunel University, Uxbridge, for the undertaken English editing.

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: Proc. 6th Int. Meeting on Stat. Climatology, Galway, Ireland, 439-441.

Alexandersson, H., Moberg, A. (1996): Homogenization of Swedish Temperature Data. Part I: A Homogeneity Test for Linear Trends. in: Moberg, A: Temperature Variations in Sweden Since the 18th Century. The Department of Physical Geography, Stockholm University. Dissertation Series No. 5, 98 pp.

Easterling, D., R., Peterson, T., C. (1995): A new method for detecting undocumented discontinuities in climatological time series. Int. Journal of Climatology, vol. 15, 369-377.

Maronna, T., Yohai, V. J. (1978): A bivariate test for the detection of a systematic change in mean. J. Amer. Statist. Assoc., 73, 640-645.

Mikulová K, Štepánek P. (2004): Porovnanie referenčných radov pri testovaní homogenity relatívnej vlhkosti vzduchu vzhľadom na rôzne kritéria ich výberu, 7. konferencia mladých meteorológov a klimatológov, Zborník príspevkov, ISBN 80-88907-79-7, Bratislava

Peterson, T.C. (1998): Homogeneity adjustments of in situ atmospheric climate data: a review. Int. J. Climatol. 18, 1493-1517.

Potter, K.W. (1981): Illustration of a New Test for Detecting a Shift in Mean in Precipitation Series. Mon. Wea. Rev., 109, 2040-2045.

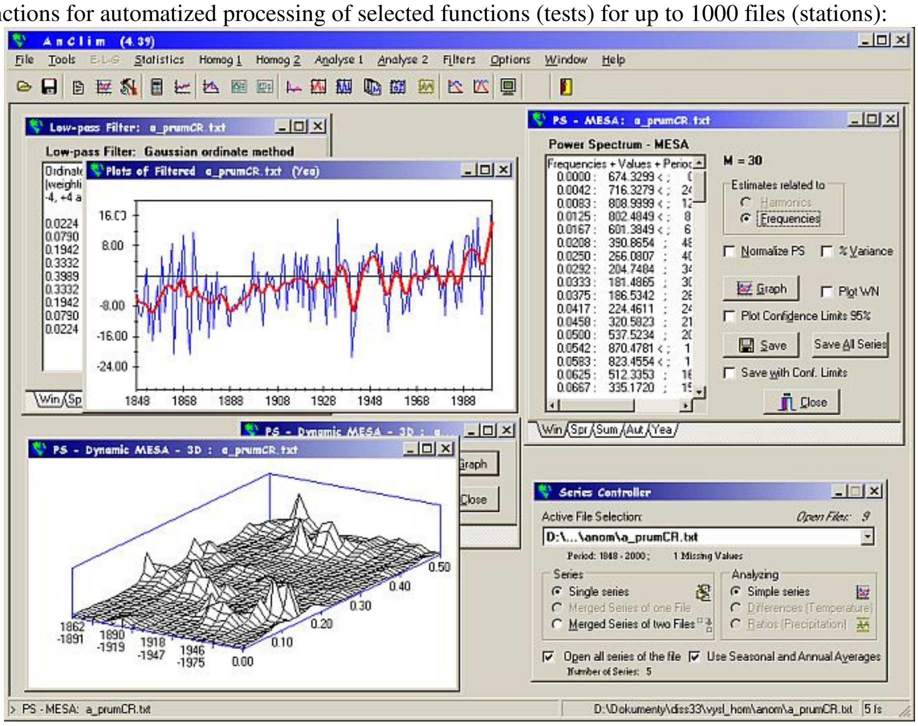
Štěpánek, P. (2004): Homogenization of air temperature series in the Czech Republic during a period of instrumental measurements. In: Fourth seminar for homogenization and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Genova. 117-133.

Štěpánek, P. (2005): Variabilita teploty vzduchu na území České republiky v období přístrojových měření (Air Temperature Fluctuations in the Czech Republic in the Period of Instrumental Measurements). Dissertation, Geographical institute, Faculty of Sciences, Masaryk University, Brno. 136 pp.

Štěpánek, P. (2006a): ProClimDB – software for processing climatological datasets. CHMI, regional office Brno. http://www.klimahom.com/software/ProcData.html
Štěpánek, P. (2006b): AnClim - software for time series analysis. Institute of Geography, Masaryk University, Brno, 1.47 MB. http://www.klimahom.com/software/AnClim.html
Szalai, S., Szentimrey, T. (2004): Proceedings of the Fourth seminar for homogenization and quality control in climatological databases (Budapest, Hungary, 6-10 October 2003), WCDMP-No. 56. WMO, Genova. 243 pp.
Szalai, S., Szentimrey, T., Szinell, C. (1999): Proceedings of the second seminar for homogenization of surface climatological data (Budapest, 9-13 November 1998). WCDMP-No. 41. WMO, Genova. 214 pp.

APPENDIX I. ANCLIM SOFTWARE DESCRIPTION

- **General characteristics**
  - A comprehensive tool for processing monthly time series (from transformations through quality control and homogenization to time series analysis)
  - Operates under Windows 95/98/NT/ME/2000/XP
  - User friendly: a lot of graphical components, graphs clarifying the results, etc.
  - Continuous development 1995-2006
- **Functionality**
  - Series overview:
    - basic statistical characteristics, tests of randomness, outliers detection etc.
    - normal distribution testing, histograms
    - graphs of the series
  - Regression models:
    - linear, polynomial regression.
    - multivariate linear regression graphs of the series
  - Adjusting data:
    - replacing outliers, filling missing values
    - various transformations, converting series into anomalies from a mean, etc.
    - calculating differences/ratios of two series
    - switching between monthly or seasonal and annual averages
  - Homogeneity testing:
    - absolute homogeneity tests
    - relative homogeneity tests (Alexandersson SNHT – various modifications, Bivariate test, Easterling and Peterson test, Vincent MLR, and others), creating reference series
    - adjustment of the inhomogeneities
  - Time series analysis:
    - one series analysis (autocorrelations, power spectrum – MESA, dynamic MESA, etc.)
    - two series analysis (coherency analysis, etc.)
    - filtering the series (low-pass, band-pass, high-pass filters)
  - Automation:
    - functions for automatized processing of selected functions (tests) for up to 1000 files (stations):



APPENDIX II. PROCLIMDB SOFTWARE DESCRIPTION

- **General characteristics**
  - Database software for processing climatological datasets (supports dbf IV files)
  - Two modes of processing: monthly or daily data
  - Automation of the processing (processing for a given list of stations, using all the stations in database)
  - Full control of the processing: many parameters for each option can be set, various outputs are created
  - Flexibility in modifying or adding new functions
- **Functionality**
  - Basic statistical characteristics computation, normal distribution testing, etc.
  - Finding outliers and extreme values
  - Neighbours stations analysis (reconstructions, quality control, etc.)
  - Calculating correlation coefficients between all the pairs of a given list of stations
  - Reference series calculated as:
    - an average from the best correlated stations
    - an average from the nearest stations
    - an average of all stations available for a given year and month (regionally)
  - Processing output from the AnClim software homogeneity testing
  - Adjusting the series for inhomogeneities
  - Filling missing values:
    - from differences
    - using linear regression
  - Calculating monthly or seasonal and annual averages, calculating differences with a given reference series, etc.
  - Export to txt files, Excel, import from txt to dbf and other formats