TECHNICAL SERIES FOR VARIOUS METEOROLOGICAL ELEMENTS IN THE CZECH REPUBLIC, 1961-2010

Petr Štěpánek[1], Pavel Zahradníček[1], Petr Skalák[1]
[1] Czech hydrometeorological institute, Czech Republic

Abstract. For various studies, it is necessary to work with sufficiently long series of daily data that are processed in the same way for the whole area. For analysis based on these series it is necessary that the series are homogeneous, i.e. that their fluctuations are only due to variations in weather and climate. For this purpose the „technical“ series of meteorological elements were created, which means that data are quality controled, homogenized and with filled gaps. These series were calculated for the locations of existing climatological and precipitation stations of the CHMI network in the period 1961-2010. Final series were interpolated to a regular grid network of 10×10 km. The technical series are currently used for various purposes

Introduction

In many scientific disciplines (inculding bioclimalogy) it is needed to process long time series of meteorogological elements. In recent years considerable attention has been devoted also to analysis of daily data. Prior to any analysis, the need to homogenize data and check their quality arises. Unfortunately, most of the time series of atmospheric data with a resolution of decades to centuries contains inhomogeneities caused by station relocations, exchange of observers, changes in the vicinity of stations (eg urbanization), changes of instruments, observing practices (eg a new formula for calculating daily average, different observation times), etc. Another requirement is that the spatial distribution of these series for the whole area is sufficient. For this purpose we created the „technical“ series of meteorological elements (mean, maximum and minumum temperature, precipitation, sunshine duration, water vapour, wind speed). By technical series we mean quality controled, homogenized and series with filled gaps. These series were calculated for the location of the existing climatological (268) and rain-gauge (787) stations of the CHMI network in the period 1961-2010 but also in the locations of regular grid points (10×10 km resolution), e.g. for purposes of validation of regional climate model (RCM) outputs..

Data and methods

Considering quality control, we faced the lack of a generally accepted methodology (contary to homogenization). But without treating outliers properly, homogenization and analysis may render misleading results. We therefore devoted considerable time to the methodology of detecting outliers, something that could moreover be automated to process large datasets of daily (subdaily) values (whole country dataset).

In this work, data quality control was carried out by combining several methods: (i) by analyzing difference series between candidate and neighbouring stations – i.e. pairwise comparisons (ii) by applying limits derived from interquartile ranges (this can be applied either to individual series, i.e. absolutely or, better, to difference series between candidate and reference series, i.e. relatively) 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). 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.

Detection of inhomogeneities was performed using monthly means (or sums in the case of precipitation). In the homogenization of the time series, the use of various statistical tests and types of reference series made it possible to increase considerably the number of homogeneity tests results for each series tested and thus to assess homogeneity more reliably. The relative homogeneity tests applied were: Standard Normal Homogeneity Test [SNHT], the Maronna and Yohai bivariate testand the Easterling and Peterson test. Reference series were calculated as weighted average from the five nearest stations (measuring within the same period as candidate series), with statistically significant correlations.Neighbouring station values were standardized to average and standard deviation of candidate station. Detection of inhomogeneities was performed for series divided to a maximum duration of 40 years, with the overlap for two consecutive periods of 10 years (due to requirements of SNHT to test only one shift). The tests were applied on monthly as well as seasonal and annual averages (sums in case of precipitation). After evaluation of detected breaks and comparison with metadata, a final decision on correction of inhomogeneities was made. Data were corrected on a daily scale.

We created our own correction method (called DAP – Distribution Adjusting by Percentiles), an adaptation of a method for the correction of regional climate model outputs by Deque (2007). Our process is based on comparison of percentiles (empirical distribution) of differences (or ratios) between candidate and reference series before and after a break. Percentiles are estimated from candidate series and values for differences of candidate and references series are taken from the same time (date). Each month is processed individually, but also taking into account the values of adjacent months before and after it to ensure smoother passage from one month to another. Candidate – reference differences for individual percentiles are then differenced before and after a break and smoothed by low-pass filter to obtain a final adjustment based on a given percentile. 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 value to be adjusted. Special treatment is needed for outlier values at the ends of distributions.

The above-mentioned steps (homogeneity testing, evaluation and correction of inhomogeneities detected) were performed in several iterations. At each iteration, more precise results were obtained (Štěpánek et al. 2009).

After quality control and homogenization, gaps have been filled for the stations that have not measured in the whole period of 1961-2010. In this way we processed stations which measured at least 20 years.

In the end, the technical series of daily values at a particular grid point (or station location) were calculated from up to 6 neighboring (nearest) stations. Before applying inverese distance weighting, data at the neighor stations were standardized relatively to the altitude of the base grid point (station location). The standartization was carried out by means of linear regression and dependence of values of a particular meteorological element on altitude for each day, indivudually and regionally. For the weighted average (using inverse distance as weights), the power of weights equal to 1 (all meteorological elements expect precipitation) and 3 (precipitation) were applied. From these technical series new values to the regular grid network 10×10 km (ALADIN-Climate/CZ RCM outputs) were interpolated by the same method (Štěpánek et al. 2011).

Results and disscussion

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 summer months than in winter months (connected with larger neighbour differences variations due to influence of active surface). More outliers were detected in the morning and evening measurements compared to noon (associated with steeper gradients in the former case). For precipitation there are two maxima per year, in the summer months and then in January and December (in winter it is pertinent to problems with solid precipitation measurements) while during spring and autumn a lower number of outliers was detected. The number of detected outliers also changes with time. For air temperature, the higher number of outliers since the late 1990s coincides well with transition to automatic measurements. Our explanation is that all values coming from automated measurements (including errors) are stored straight into database while in the case of manual measurements observer revises read values before sending them to meteorological office. In the last years the number is low again, thanks to internal CHMI data quality control. On the contrary, in the case of precipitation no increase of errors after automation was encountered.

As for inhomogeneities detection itself, for air temperature more breaks occur in the summer months (the influence of relocation and other artificial changes is greater resulting from influence of active surface – prevailing radiation factors and increased volume of vegetation) while this occurs for precipitation in the winter months (mainly due to problems associated with measurement of solid precipitation). An annual cycle is also clearly manifested in the correction of inhomogeneities. Considering the absolute values of corrections, the degrees of adjustment were higher during the summer months for air temperature. For precipitation, major corrections (ratios) were applied in winter months (reasons being the same as for number of inhomogeneities, see previous paragraph). After correction, for air temperature correlation coefficients increased mainly in the summer months, for precipitation in the winter months. Automation of measurements had very strong influence on the homogeneity of station time series (and also on the occurrence of outliers). Fortunately, automation was introduced successively into the station network so it was possible to detect it and make corrections without major problems.

After quality control and homogenization, technical series by above described method for the 7 meteorological elements (mean, maximum and minumum temperature, precipitation, sunshine duration, water vapour, wind speed) were created. All these series are avaible for 268 climatological stations and 787 precipitation stations. The same elements are avaible for the 789 grid points. .

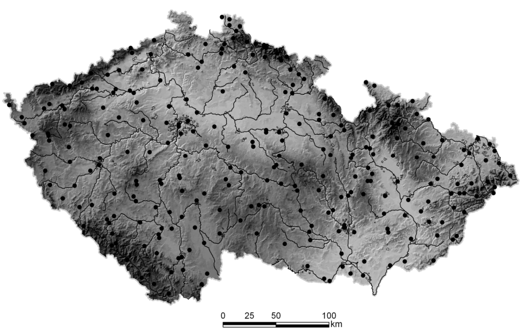

Figure 1. Spatial distribution of the so called technical series for climatological stations (268), which are quality controled, homogenized and with filled gaps in the period 1961-2010

Conclusions

The current work presents a methodology for outlier detection, series homogenization and interpolation technique for various meteorological elements in the area of the Czech Republic in the period 1961–2010. Thanks to the technical series we gained sufficiently large number of climatological series with equal spatial distribuatiton in the area of Czech Republic. These series are withnout outliers, homogeneous and gaps have been filled. Technical series are used for research in various projects in climatology and hydrology. From technical series we also calculated maps of various meteorological elements for each months and day in the period 1961-2010 (more than 130 000 maps). Further steps in creating technical series will lead to analysis of individual observation hours and also back to historical records.

Acknowledgement: The authors would like to acknowledge the financial support of the Grant Agency of the Czech Republic for project no.P209/10/0605 and P205/09/1297

References

Deque, M.: 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, 2007.

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

Štěpánek, P., Zahradníček, P., Skalák, P., Data quality control and homogenization of air temperature and precipitation series in the area of the Czech Republic in the period 1961–2007. Advances in Science and Research, 3, 23–26. 2009