Advances in
Science & Research
Open Access Proceedings

The historical pathway towards more accurate homogenisation

P. Domonkos[1], V. Venema[2], I. Auer[3], O. Mestre[4], and M. Brunetti[5]

[1]Center for Climate Change, Univ. Rovira i Virgili, Campus Terres de l'Ebre, Tortosa, Spain
[2]Meteorological institute of the University of Bonn, Germany
[3]Zentralanstalt für Meteorologie und Geodynamik, Wien, Austria
[4]Meteo France, Ecole Nationale de la Meteorologie, Toulouse, France
[5]Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy

Correspondence to: P. Domonkos (peter.domonkos@urv.cat)

Abstract. In recent years increasing effort has been devoted to objectively evaluate the efficiency of homogenisation methods for climate data; an important effort was the blind benchmarking performed in the COST Action HOME (ES0601). The statistical characteristics of the examined series have significant impact on the measured efficiencies, thus it is difficult to obtain an unambiguous picture of the efficiencies, relying only on numerical tests. In this study the historical methodological development with focus on the homogenisation of surface temperature observations is presented in order to view the progress from the side of the development of statistical tools. The main stages of this methodological progress, such as for instance the fitting optimal step-functions when the number of change-points is known (1972), cutting algorithm (1995), Caussinus – Lyazrhi criterion (1997), are recalled and their effects on the quality-improvement of homogenisation is briefly discussed. This analysis of the theoretical properties together with the recently published numerical results jointly indicate that, MASH, PRODIGE, ACMANT and USHCN are the best statistical tools for homogenising climatic time series, since they provide the reconstruction and preservation of true climatic variability in observational time series with the highest reliability. On the other hand, skilled homogenizers may achieve outstanding reliability also with the combination of simple statistical methods such as the Craddock-test and visual expert decisions. A few efficiency results of the COST HOME experiments are presented to demonstrate the performance of the best homogenisation methods.

1 Introduction

Recently the COST Action HOME (ES0601) (hereafter: HOME) has evaluated the methodological progress in homogenisation of climatic time series. The project, which terminated in October 2011, fostered the development of new statistical tools for homogenising time series, as well as the objective assessment of the effectiveness of traditional and newly-developed methods. After this fruitful period it is timely to examine the recent methodological advances and new scientific results in a wider historical context. In this study a brief overview of the historical development of homogenisation methods is presented, focusing on those steps that turned out to have lasting influence on the practical solutions of the homogenisation task. Most findings are valid for the homogenisation of annual and monthly surface tempera-

ture datasets of high spatial correlations in networks, or for the homogenisation of other climatic variables when some basic statistical properties (length of time series, spatial correlations, etc.) are comparable with those of the surface temperature datasets. The review is supplied with some illustrations about the efficiencies of the best methods relying on recent test results of HOME.

2 The birth of statistical homogenisation

The idea that technical-originated biases should be eliminated from climatic time series through spatial comparisons of the data is as old as the existence of professional meteorological observing networks, and documents prove that homogenisation, i.e. correction of local biases for sections of observed time series took place as early as in the 19th

century (Kreil, 1854). The earliest methods were the subjective estimation of biases based on visual inspection of the observed data and the use of experimental measurements. For example, surface temperature measurements of Milan (Italy) between 1763 and 1834 were adjusted (Kreil, 1854) due to changes in the daily routine of observations. That time daily 7 observations were done in Milan (at 00:00, 03:00, 06:00, 09:00, 12:00, 18:00 and 21:00 LT), but Kreil and his colleagues wanted to produce more accurate daily mean temperatures than the average of such seven data daily. For this reason, a 16 month long experiment of hourly observation performed in Padua, and the bias of the average of the seven observations from the true daily mean was revealed, then the daily mean temperatures of Milan were corrected.

The first documented application of statistical homogenisation is known from 1925. The Austrian climatologist Victor Conrad searched possible change-points by splitting the series of annual precipitation totals into two parts at each year and assuming that the break is in the year for which the statistic shows the most significant difference between the two parts (Conrad, 1925). In the referred study the ratio of the upper and lower quartiles for two time series were compared, but Conrad also proposed two further statistical methods for detecting change-points. The scale of the applicable statistical tools is wide, since a change-point may cause significant difference in the means, accumulated anomalies, mean rank-order values, etc. between two parts of a time series. If the shift in the change-point is large compared to the noise level, these differences can even easily be identified visually (Fig. 1). When the signal-to-noise ratio is lower, statistical significance examination may help to make the distinction between random fluctuations and true inhomogeneities. Since the appearance of the first statistical homogenisation method, a large number of types and versions of homogenisation methods have been developed. Note that in spite of the great development of statistical tools, subjective decisions have not disappeared completely from the homogenisation of climatic time series.

3 Until 1990: single change-point models

In the 20th century, a large number of studies dealt with the problem of time series homogenisation, but before 1990 only the single change-point problem was analysed intensively. Around and after the mid 20th century the double mass analysis became a popular method (Kohler, 1949). In this old method the accumulated anomalies (i.e. progressive sums of anomalies from the beginning of the examined period) are visually compared between the tested series (candidate series) and another series (reference series). Among the later developed methods, the Craddock-test (Craddock, 1979) and the Buishand tests (Buishand, 1982) are also based on the examination of accumulated anomalies. Buishand (1982) provided empirically-developed significance thresholds to his
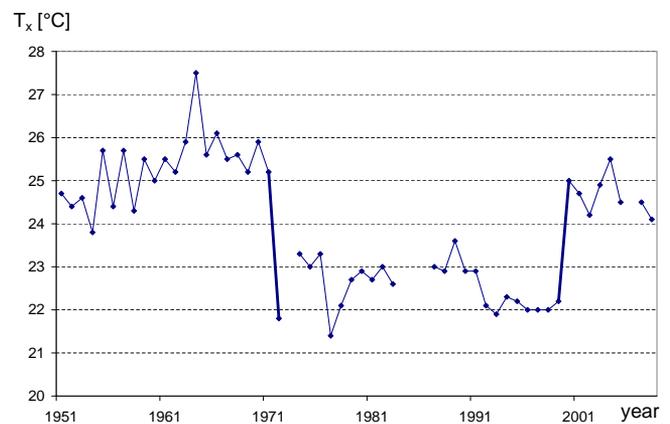
Figure 1. Summer daily maximum temperatures (T_x) of Tarifa (Spain) between 1951–2008. The two spectacularly large shifts are marked with thick lines.

methods, eliminating the need of subjective decisions about the significance of possible change-points. By contrast, the Craddock-test is essentially subjective.

Beyond the examination of accumulated anomalies, several other kinds of tests were developed for solving the single change-point problem. A few of the most used methods are mentioned here. Wilcoxon Rank Sum Test (Wilcoxon, 1945, also known as Mann-Whitney-Wilcoxon test) is a non-parametric method. It is based on the calculation of rank sum statistics before and after the potential change-point. The Maronna-Yohai test (Maronna and Yohai, 1978) is based on maximum likelihood estimations. In Standard Normal Homogeneity Test (SNHT, Alexandersson, 1986) the section-means before and after the potential change-point are compared. Solow (1987) searched for the change-points by fitting two-phase linear regression to the data points. The mentioned four kinds of approach (i.e. non-parametric method, maximum likelihood estimation, comparison of section-means, fitting linear regressions) have several other representatives. Among the contemporary methods the maximum likelihood methods were found to be the best theoretically and they show the highest performance in efficiency test examinations (Domonkos, 2008). However, none of the contemporary statistical methods treats the complex interactions of multiple change-points on the examined test-statistics or their effects on the calculation of correction-terms.

A seeming change-point in a climatic time series may have a macro-climatic or a local, technical origin, and it was known from the beginning that one must distinguish these two during the homogenisation (Conrad and Pollak, 1950). Therefore homogenisation is applied on differences or ratios of two series (relative homogenisation) instead of on one time series (absolute homogenisation), with very few exceptions. In relative homogenisation one of the crucial problems is that the detected change-points often could be originated

from more than one inhomogeneous time series and it is not easy to find the true "culprit". An important step forward was the creation of reference series from composite series of the nearby stations (WMO, 1966).

The best-known and most widely applied homogenisation method is SNHT. Alexandersson (1986) did not construct better method than his contemporaries, but his merits are that he set a good example of practical application from the selecting of time series until the final interpretation of detection results and provided a user-friendly description of his statistical method.

There was one additional line of methodological development at that time. Hawkins (1972) constructed the method of optimal segmentation of step-functions for the case of known number of steps. The presented method was not only optimal in statistical sense, but also economical in the use of computer-time, therefore this method is known also as dynamic programming algorithm. With that step research came close to the solution of the multiple inhomogeneities problem. However, at that time the importance of the multiple breaks problem in climatic time series was not widely recognized. Instead of following this line of research, in the following years Hawkins turned back to the examination of single change-point model (Hawkins, 1977).

4 Fast development from the 90's

Around 1990 global climate change started to be seen as a potential serious problem and consequently the homogeneity of the observed climatic data became more important. In 1994 the first seminar on data quality control and homogenisation was held, and since then this series of seminars, which is supported by the World Meteorological Organisation (WMO), is held regularly in Budapest. New research lines started in the framework of these seminars, but also independently from them. In the first homogenisation seminar appeared the initial versions of some excellent statistical tools. That time the first attempts to detect multiple change-points as a coherent structure (joint detection of multiple change-points) was published (Szentimrey, 1996; Caussinus and Mestre, 1996, as well as the stepwise comparison of time series, Caussinus and Mestre, 1996) and the use of multiple reference series (Szentimrey, 1996). In the same year as the first homogenization seminar, Peterson and Easterling (1994) published a new method to create reference series from the weighted averages of values in neighbouring stations, where the weights are the squared correlations of the increment series. Since then the latter three methods can be considered recommended ways for spatial comparisons of time series.

Easterling and Peterson (1995) published a hierarchic method for identifying multiple change-points. Later this method became popular, and it is often referred as cutting algorithm. Modern efficiency examinations show

(e.g. Domonkos, 2011a) that apart from the joint detection of multiple change-points the cutting algorithm is the best available tool. A complete new homogenisation method was also published in the same study (the combination of two phase regression method with multiple permutation procedure), but that has no better performance than the other contemporary methods.

In 1996 the first attempt to detect both change-points and trend-like inhomogeneities was published, Lanzante (1996) applied the modified version of Wilcoxon Rank Sum Test. However, we do not know of any later applications of that test. Later studies that also considered trend-like inhomogeneities, always applied the late version of SNHT (Alexandersson and Moberg, 1997) or Multiple Linear Regression (MLR, Vincent, 1998). However, efficiency-tests show that these detection methods have often poorer performance than the best of the other methods (Domonkos, 2008, 2011a). The performance of the newer version of SNHT is poorer than that of the earlier version (Alexandersson, 1986). On the other hand, Moberg and Alexandersson (1997) set a good example how to practically apply SNHT or other homogenisation methods that do not include joint detection of multiple change-points. They applied the reference creation by Peterson and Easterling (1994) and also the cutting algorithm. They developed a seemingly appealing semi-hierarchic algorithm for detecting multiple change-points, but its effect is neutral on the efficiency of detection results (Domonkos, 2011a).

In the second part of the 90's the development of methods with joint detection of multiple change-points was the most important research-line. Multiple Analysis of Series for Homogenisation (MASH, Szentimrey, 1999) counts all the possible combination of change-point positions and selects the most likely one based on hypothesis tests. The method uses multiple reference series to reduce the impact of inhomogeneities in reference series on the detection of inhomogeneities in the candidate series, and it has a specific philosophy for keeping the false alarm rate low, i.e. the lower limits of confidence intervals are used as adjustment-factors. MASH is rather complicated and it is more time consuming than any other homogenisation method, but at the end of the 90's it was likely the most effective method. Considering the reliability and preciseness of change-point detection, the Caussinus-Mestre method (its modern version is known as PRODIGE, Caussinus and Mestre, 1996; Mestre, 1999), could be its only competitor, but at that time the Caussinus-Mestre method had not reached its final form yet. PRODIGE applies multiple stepwise comparisons instead of reference series. Its detection part uses the maximum likelihood principle with the Caussinus-Lyazrhi criterion (Caussinus and Lyazrhi, 1997) for finding the optimal number of change-points in best fitting step-functions. This criterion combined with the dynamic programming algorithm to optimize the break positions (Hawkins, 1972) makes PRODIGE a powerful tool. A few years later the detection algorithm of

PRODIGE was refined and the ANOVA correction method was introduced (Caussinus and Mestre, 2004). Since that development the efficiency of PRODIGE is similar to the efficiency of MASH.

5   The 21th century: intercomparisons

Around 2000 the first review studies about homogenisation methods appeared (Peterson et al., 1998; Aguilar et al., 2003; Auer et al., 2005, etc.) and there are some studies about the efficiency of change-point detection by different methods (Easterling and Peterson, 1995; Lanzante, 1996; Ducré-Robitaille et al., 2003, etc.). However, the first experiments for characterising objectively the efficiencies of homogenisation methods suffered from substantial shortcomings: (i) simple models were used to generate test-datasets, thus the resemblance between test datasets and true observational datasets was generally low; (ii) detection skill (i.e. the skill in finding the positions of breaks in time series, see e.g. Menne and Williams Jr., 2005) was calculated only, and this measure does not always characterise well the skill in reconstructing and preserving the true characteristics of climatic variability in homogenised time series, (iii) small number of arbitrarily selected methods were tested only.

From 2004 our general knowledge about the statistical properties of inhomogeneities in observational datasets widened. In experiments with a Hungarian observed temperature dataset, statistical characteristics of inhomogeneity-detection results were compared for true and simulated datasets. The empirical results showed that high similarity between the characteristics for the observed and artificial datasets can be obtained only when large number of short-term, platform-like inhomogeneities are included in simulated time series and most inhomogeneities have small magnitude. These characteristics could remain hidden in direct examination of observed time series, because short-term inhomogeneities and small-size inhomogeneities often cannot be detected at all because of the noise. The newly discovered feature of observed time series is important, because small-size inhomogeneities have impact on the performance in detecting and correcting large-size inhomogeneities (Domonkos, 2004, 2011a).

Menne and Williams Jr. (2005) examined the detection skills of homogenisation methods using a simulated dataset that included change-point sizes determined by a normal distribution. Domonkos (2008) examined the efficiency of detection parts of fifteen homogenisation methods with a test dataset whose properties were similar to the observed temperature dataset of Hungary. Beyond detection skill, other efficiency-measures, such as RMSE and the accuracy of linear trend-slopes in homogenised time series were calculated. The influence of parameter-choices on the detection results was analysed in that study as well.

6   The HOME period

Between 2007 and 2011 HOME provided favourable conditions for the further developments. A benchmark dataset was created (Venema et al., 2010), which contains surrogate climate networks mimicking the statistical properties of monthly temperature and precipitation time series in European observational networks well. In the benchmark the frequency of inhomogeneities is diverse, and the size-distribution of inhomogeneities approaches well the true characteristics (much more small inhomogeneities than large ones). In the benchmark the resemblance of simulated networks to the real world is demonstrably high which was not common in earlier validation studies of homogenisation methods. HOME and the benchmark provided an excellent opportunity to evaluate the performances of a large number of whole homogenisation procedures with the participation of homogenisers from different countries (Venema et al., 2012). During the benchmark homogenization the true positions of the breaks remained unknown for the homogenizers. The blind test results were evaluated calculating RMSE of monthly and annual values, RMSE of trend-slope estimatons, detection power, false alarm rate, and some other efficiency measures. These tests confirmed that PRODIGE and MASH are indeed among the best homogenisation tools, but several other results may have been less expected. One main finding of the experiments was that the difference between the efficiency of PRODIGE and MASH on the one hand, and the other known methods such as SNHT, Penalised Maximum t-test (PMT, Wang et al., 2007), etc. is substantially larger than the differences between detection performances according to Domonkos (2008, 2011a). The found larger differences for performances of whole procedures are likely due to the over-simplified correction algorithms of several methods.

During HOME a new homogenisation method, Applied Caussinus-Mestre Detection Algorithm for homogenising Networks of Temperature series (ACMANT, Domonkos, 2011b) was developed for homogenising monthly temperature series. The ACMANT includes the step-function fitting and ANOVA correction segments of PRODIGE, but applies a bivariate-test for detecting change-points. Two annual variables are used in the detection, one is the annual mean, and the other is the summer-winter difference. Due to radiation-connected biases, joint inhomogeneities of these variables are frequent in mid-latitude temperatures (Drougue et al., 2005; Domonkos and Štěpánek, 2009; Brunet et al., 2011, etc.), thus the detection model of ACMANT is a powerful tool. The benchmark results show that ACMANT is the most effective homogenisation method for temperature, where it should be noted that the presented contribution by ACMANT was submitted after the deadline and the test was thus not blind. ACMANT applies the method of Peterson and Easterling (1994) to build the reference series and with its inclusion the procedure could be made fully automatic. ACMANT is able to homogenise time series with data-gaps

and networks of different-length time series in an automatic way. It applies sophisticated tools for treating the connections between annual- and monthly-scale examinations and corrections. Note that the use of bivariate search for change-point positions is not limited to ACMANT; it has recently been applied in another study to homogenise solar radiation and sunshine duration (Guijarro, 2011).

The methodological development has also been continued overseas. A new automatic homogenisation method, USHCN (Menne and Williams Jr., 2009), was published for homogenising huge temperature datasets, as found in the United States. The detection part of USHCN includes the early version of SNHT, cutting algorithm, Bayesian-based decisions about the form of inhomogeneities, i.e. trend-like inhomogeneities can also be detected, and a special purpose significance test. Important novelties of USHCN are that it applies pairwise comparisons in automated way and automatically uses metadata. USHCN applies homogeneity-adjustments only when the individual estimates from pairwise comparisons concordantly indicate the need for same-sign adjustment. USHCN was tested in the HOME experiment. Its general efficiency turned out to be slightly poorer than that of PRODIGE, MASH and ACMANT, but USHCN only performed annual correction and has the lowest false alarm rate. This latter positive feature of USHCN might become crucially important when datasets with low frequency of large-size inhomogeneities are examined.

Another important step in the methodological development is the introduction of perturbed parameter experiments to the test-process of automatic homogenisation methods (McCarthy et al., 2008; Domonkos, 2008; Titchner et al., 2009; Williams et al., 2012). In these examinations some parameters of the homogenisation methods are varied randomly (in ensemble tests) or systematically (in sensitivity tests) and from the synthesis of the results a more complete picture can be obtained about the connections between the dataset-properties and the performance of the applied homogenisation method.

Although the main line of achieving more precise homogenisation methods is the development of more powerful statistical tools, an old and very simple method, namely the Craddock-test was proven to be a very effective homogenisation method in the blind test experiments of HOME. In case of the Craddock-test, the main protagonist is the homogeniser. Skilled homogenisers may assess the timing and size of change-points well by examining the time series subjectively with the help of some simple statistical characteristics such as the series of accumulated anomalies for the differences of the compared time series. The relation between a skilled homogeniser and an automatic homogenisation method is similar to that between a chess-master and a chess-automat. In the HOME experiments, the results of Gregor Vertacnik (Slovenia) and Michele Brunetti (Italy) were as good as that of the best objective and semi-objective homogenisation methods. Note that the Craddock-
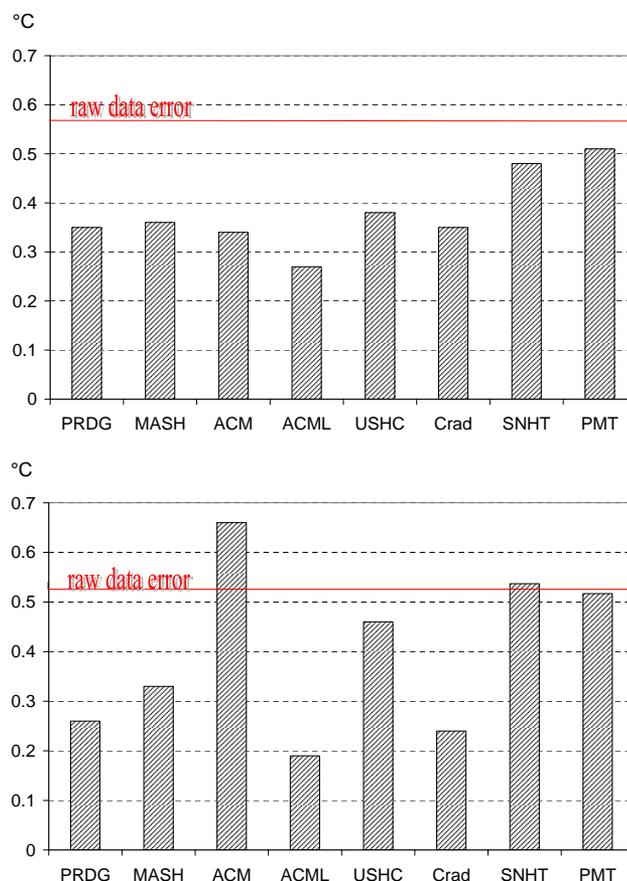
Figure 2. Efficiencies of monthly temperature homogenisation for selected homogenisation methods. PRDG = PRODIGE, ACM = early ACMANT, ACML = late ACMANT, USHC = USHCN, Crad = Craddock-test. **(a)** (upper) RMSE of monthly values. **(b)** (bottom) RMSE of biases in network-mean linear trends.

homogenisers used some selected parts of the benchmark only, thus their results are not fully comparable with the results of complete experiments.

The HOME experiments show that the best homogenisation methods are PRODIGE, MASH, ACMANT, USHCN and the Craddock-test. Note that these results refer primarily to the homogenisation of surface temperature time series of relatively dense European or North American observing networks. In Fig. 2 some efficiency results are illustrated for the seven best methods participated in the Benchmark homogenisation. In the construction of these figures full contributions (that used all the 15 surrogated temperature networks; Venema et al., 2012) were taken into account, except for the Craddock-test as Vertacnik's partial contribution contains only 7 networks. When authors produced more than one full experiment with different versions of their methods, the average error of these versions is shown. Most of the efficiencies shown in Fig. 2 are identical to the equivalent results of Venema et al. (2012), except the results for ACMANT

late which method was finalised later than the HOME experiments. One can see from the figures that the RMSE error of homogenised time series is lower than that of the raw data for all the methods shown. On the other hand, the efficiency can be negative in such an important climatic characteristic as the accuracy of the network-wide mean linear trend. For this characteristic, significant positive efficiencies are produced only by PRODIGE, MASH, Craddock-test and ACMANT late. Note that after the HOME experiments ACMANT late was subjected to further tests, both to blind tests and perturbed parameter tests, and that results (Domonkos, 2012) confirmed the good performance of ACMANT late (Fig. 2).

Based these theoretical considerations and the experiments with the benchmark, the HOME team developed a software-package for homogenising temperature and precipitation time series (www.homogenisation.org). This package incorporates segments from the best homogenisation methods examined, thus its efficiency is likely similar or better than the other best methods.

Finally, it must be noted that progress in homogenisation methods is never limited to the development and application of statistical tools. A recent example is the Spanish screen bias experiment in which the inhomogeneities caused by the change of the thermometer-screens in the early 20th century have been quantitatively assessed relying on experimental parallel measurements from the recent years (Brunet et al., 2011).

7 Concluding remarks

These conclusions refer particularly to the homogenisation of annual and monthly surface temperature time series of sufficiently dense observing networks, although some are more generally valid.

According to our present knowledge six homogenisation methods can be recommended. They are PRODIGE, MASH, ACMANT, USHCN, the Craddock-test and the HOME-software.

The appropriateness of the six methods listed above is often markedly different in solving particular tasks. For instance ACMANT is a highly efficient tool for homogenising temperature datasets of mid-latitudes, but is not tailored to other variables. For homogenising huge datasets USHCN or ACMANT are recommendable, because these methods are fully automatic. The HOME-software, PRODIGE and MASH are usable in wide range of tasks, but certain expertise is needed for their use. The Craddock-test is subjective and is inappropriate for homogenising large datasets.

Further tests are needed to understand the performance of homogenisation methods better. The characteristics of climatic time series are diverse, thus a large number of experiments with varied dataset properties is needed.

The general advantages of fully automatic methods are that they easily can be tested in multiple experiments; their test results are objective and can be reconstructed at any time.

The use of automatic homogenisation methods still needs some expertise in the time series preparation and in the interpretation of the homogenisation results.

The publication of this article is sponsored by the European Meteorological Society.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J.: WMO Guidelines on climate metadata and homogenization, WCDMP-No. 53, WMO-TD No 1186, WMO, Geneva, 2003.

Alexandersson, H.: A homogeneity test applied to precipitation data, J. Climatol., 6, 661–675, doi:10.1002/joc.3370060607, 1986.

Alexandersson, H. and Moberg, A.: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends, Int. J. Climatol., 17, 25–34, 1997.

Auer, I., Böhm, R., Jurković, A., Orlik, A., Potzmann, R., Schöner, W., Ungersböck, M., Brunetti, M., Nanni, T., Maugeri, M., Briffa, K., Jones, P., Efthymiadis, D., Mestre, O., Moisselin, J.-M., Begert, M., Brazdil, R., Bochnicek, O., Cegnar, T., Gajić-Čapka, M., Zaninović, K., Majstorović, Ž., Szalai, S., Szentimrey, T., and Mercalli, L.: A new instrumental precipitation dataset for the greater Alpine region for the period 1800–2002, Int. J. Climatol., 25, 139–166, doi:10.1002/joc.1135, 2005.

Brunet, M., Asin, J., Sigró, J., Bañon, M., García, F., Aguilar, E., Palenzuela, J. E., Peterson, T. C., and Jones, P.: The minimization of the screen bias from ancient Western Mediterranean air temperature records: an exploratory statistical analysis, Int. J. Climatol., 31, 1879–1895, doi:10.1002/joc.2192, 2011.

Buishand, T. A.: Some methods for testing the homogeneity of rainfall records, J. Hydrol., 58, 11–27, 1982.

Caussinus, H. and Lyazrhi, F.: Choosing a linear model with a random number of change-points and outliers, Ann. Inst. Statist. Math., 49, 761–775, doi:10.1023/A:1003230713770, 1997.

Caussinus, H. and Mestre, O.: New mathematical tools and methodologies for relative homogeneity testing, in: Proceedings of the First Seminar for Homogenization of Surface Climatological Data, edited by: Szalai, S., Hungarian Meteorlogical Service, Budapest, Hungary, 63–82, 1996.

Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, J. Roy. Stat. Soc., Series C53, 405–425, doi:10.1111/j.1467-9876.2004.05155.x, 2004.

Conrad, V.: Homogenitätsbestimmung meteorologischer Beobachtungsreihen, Meteorol. Z., 482–485, 1925.

Conrad, V. and Pollak, C.: Methods in Climatology, Harvard University Press, Cambridge, MA, p. 459, 1950.

Craddock, J. M.: Methods of comparing annual rainfall records for climatic purposes, Weather, 34, 332–346, 1979.

Domonkos, P.: Testing of homogenisation methods, in: 5th European Conference of Applied Climatology, 26–30 September, Nice, France, EMS04-A-00098, 2004.

Domonkos, P.: Testing of homogenisation methods: purposes, tools and problems of implementation, Proceedings of the 5th Seminar and Quality Control in Climatological Databases, WCDMP-No. 71, WMO-TD 1493, WMO, Geneva, 126–145, 2008.

Domonkos, P.: Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods, Theor. Appl. Climatol., 105, 455–467, doi:10.1007/s00704-011-0399-7, 2011a.

Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), Int. J. Geosci., 2, 293–309, doi:10.4236/ijg.2011.23032, 2011b.

Domonkos, P.: ACMANT: Why is it efficient? Proceedings of the 7th Seminar and Quality Control in Climatological Databases, WMO-HMS, www.c3.urv.cat/publicacions/publicacions2012.html, in press, 2012.

Domonkos, P. and Štěpánek, P.: Statistical characteristics of detectable inhomogeneities in observed meteorological time series, Studia Geoph. et Geod., 53, 239–260, doi:10.1007/s11200-009-0015-9, 2009.

Drogue, G., Mestre, O., Hoffmann, L., Iffly, J.-F., and Pfister, L.: Recent warming in a Small region with semi-oceanic climate, 1949–1998: what is the ground truth?, Theor. Appl. Climatol., 81, 1–10, doi:10.1007/s00704-004-0088-x, 2005.

Ducré-Robitaille, J.-F., Vincent, L. A., and Boulet, G.: Comparison of techniques for detection of discontinuities in temperature series, Int. J. Climatol., 23, 1087–1101, doi:10.1002/joc.924, 2003.

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

Guijarro, J. A.: Joint homogenization of sunshine duration and global solar radiation Spanish series. 10th European Conference on Applied Meteorology, 12–16 September, Berlin, Germany, EMS2011-679, 2011.

Hawkins, D. M.: On the choice of segments in piecewise approximation, J. Inst. Math. Appl., 9, 250–256, doi:10.1093/imamat/9.2.250, 1972.

Hawkins, D. M.: Testing a sequence of observations for a shift in location, J. Amer. Stat. Assoc., 72, 180–186, 1977.

Kohler, M. A.: Double-mass analysis for testing the consistency of records and for making adjustments, B. Am. Meteorol. Soc., 30, 188–189, 1949.

Kreil, K.: Mehrjährige Beobachtungen in Wien vom Jahre 1775 bis 1850, Jahrbücher der k.k. Central-Anstalt für Meteorologie und Erdmagnetismus, I. Band – Jg. 1848 und 1849, 35–74, 1854.

Lanzante, J. R.: Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, Int. J. Climatol., 16, 1197–1226, 1996.

Maronna, R. and Yohai, V. J.: A bivariate test for the detection of a systematic change in mean, J. Amer. Stat. Assoc., 73, 640–645, 1978.

McCarthy, M. P., Titchner, H. A., Thorne, P. W., Tett, S. F. B., Haimberger, L., and Parker, D. E.: Assessing bias and uncertainty in the HadAT adjusted radiosonde climate record, J. Climate, 21, 817–832, 2008.

Menne, M. J. and Williams Jr., C. N.: Detection of undocumented changepoints using multiple test statistics and composite reference series, J. Climate, 18, 4271–4286, doi:10.1175/JCLI3524.1, 2005.

Menne, M. J. and Williams Jr., C. N.: Homogenization of temperature series via pairwise comparisons, J. Climate, 22, 1700–1717, doi:10.1175/2008JCLI2263.1, 2009.

Mestre, O.: Step-by-step procedures for choosing a model with change-points, Second Seminar for Homogenization of Surface Climatological Data, edited by: Szalai, S., Szentimrey, T., and Szinell, Cs., WCDMP 41, WMO-TD 962, WMO, Geneva, 15–25, 1999.

Moberg, A. and Alexandersson, H.: Homogenization of Swedish temperature data. Part II: Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861, Int. J. Climatol., 17, 35–54, 1997.

Peterson, T. C. and Easterling, D. R.: Creation of homogeneous composite climatological reference series, Int. J. Climatol., 14, 671–679, doi:10.1002/joc.3370140606, 1994.

Peterson, T. C., Easterling, D. R., Karl, T. R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Böhm, 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.: Homogeneity adjustments of in situ atmospheric climate data: a review, Int. J. Climatol., 18, 1493–1517, 1998.

Solow, A. R.: Testing for climate change: an application of the two-phase regression model, J. Clim. Appl. Meteor., 26, 1401–1405, 1987.

Szentimrey, T.: Statistical procedure for joint homogenization of climatic time series, in: Proceedings of the First Seminar for Homogenization of Surface Climatological Data, edited by: Szalai, S., Hungarian Meteorlogical Service, Budapest, Hungary, 47–62, 1996.

Szentimrey, T.: Multiple Analysis of Series for Homogenization (MASH), in: Second Seminar for Homogenization of Surface Climatological Data, edited by: Szalai, S., Szentimrey, T., and Szinell, Cs., WCDMP 41, WMO-TD 962, WMO, Geneva, 27–46, 1999.

Titchner, H. A., Thorne, P. W., McCarthy, M. P., Tett, S. F. B., Haimberger, L., and Parker, D. E.: Critically re-assessing tropospheric temperature trends from radiosondes using realistic validation experiments, J. Climate, 22, 465–485, doi:10.1175/2008JCLI2419.1, 2009.

Venema, V., Mestre, O., and the COST HOME Team: Benchmark database, EGU General Assembly, Vienna, Austria, 3–7 May 2010, EGU CL4.6-13357, 2010.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Mller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89–115, doi:10.5194/cp-8-89-2012, 2012.

Vincent, L. A.: A technique for the identification of inhomogeneities in Canadian temperature series, J. Climate, 11, 1094–1104, 1998.

Wang, X. L., Wen, Q. H., and Wu, Y.: Penalized maximal t test for detecting undocumented mean change in climate data series, J. Appl. Meteor. Climatol., 46, 916–931, doi:10.1175/JAM2504.1, 2007.

Wilcoxon, F.: Individual comparisons by ranking methods, Biometrics Bull., 1, 80–83, 1945.

Williams, C. N., Menne, M. J., and Thorne, P. W.: Benchmarking the performance of pairwise homogenization of surface temperatures in the United States, J. Geophys. Res.-Atmos., 117, D05116, doi:10.1029/2011JD016761, 2012.

WMO: Climatic Change, Report of a working group of the Commission for Climatology, Technical Note 79, WMO – No. 195. TP. 100, pp. 79, 1966.