Vol. 74: 31–42, 2017
https://doi.org/10.3354/cr01488

CLIMATE RESEARCH
Clim Res

Published online November 2

Time series homogenisation of large observational datasets: impact of the number of partner series on efficiency

Peter Domonkos[1,*], John Coll[2]

[1]Acces al Seminari 16, 43500 Tortosa, Spain
[2]Irish Climate Analysis and Research UnitS (ICARUS), Department of Geography, Maynooth University, Co Kildare, Ireland

ABSTRACT: Changes in climatic observations (such as station relocations and changes of instrumentation) often affect the spatial and temporal comparability of the data; therefore, an important part of improving the accuracy of observed climate variability is the time series homogenisation of the source data. In undertaking homogenisation, an essential step is the spatial comparison of the data within the same geographical region. To optimise the efficiency of homogenisation, we should know when and to what extent two series are of the same geographical origin from a climatic perspective, and how many partner series should be used. This study presents a number of novel experiments for obtaining objective answers to these questions. Monthly temperature test datasets were homogenised with ACMANT (Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series) by varying the number of partner series and their spatial correlations with the candidate series. First, a homogeneous benchmark is constructed from 2 regional subsets of a simulated surface air temperature dataset from earlier work. Various kinds of inhomogeneities are then inserted into the time series, producing 5 basic types of test datasets for each geographical region. Further variation is introduced by adding additional noise to some datasets, providing more diverse spatial correlations. The results indicate that for the identification and correction of long-lasting biases in the data, the optimal number of partner series is about 30. The optimum is largely independent from the frequency and intensity of inhomogeneities and from the spatial correlation between the candidate series and its partner series. This latter finding is unexpected; hence, its possible causes and the consequences are discussed and explored more fully here.

KEY WORDS: Time series · Homogenisation · Data quality · Efficiency test · ACMANT · Temperature

1. INTRODUCTION

Instrumental observations are an important source of knowledge in many areas of climatology; therefore, data quality and the spatio-temporal comparability of data are widely discussed. The homogenisation of observational time series aims to remove the effects of technical changes of the observations from the data, as technical changes, such as changes in the instrumentation, observation rules, station location, etc., may influence the apparent climate variability (e.g. Aguilar et al. 2003, Auer et al. 2005, Williams et al. 2012). Some of these factors can cause abrupt shifts; others can cause gradual changes over time, which can hamper identification of genuine climatic variations or lead to erroneous interpretations (Peterson et al. 1998). Since these shifts are often of the same magnitude as the climate signal (Auer et al. 2007, Menne et al. 2009), a direct analysis of the original

*Corresponding author: dpeterfree@gmail.com

data series can lead to incorrect conclusions about the evolution of climate.

The most frequently applied mode of time series homogenisation is the filtering of local, technical effects from the data via comparisons with neighbouring series and the use of some statistical tests. Such procedures are often referred to as relative homogenisation, since the spatial comparison of the data helps to separate the possible significant deviations in the apparent temporal evolution of local climate relative to the true and regionally common climate signal. Relative homogenisation (hereafter: homogenisation) can be applied with or without the use of documented information (metadata) about known technical changes in station history. In the present study, we focus on the case of large and dense datasets where the data availability favours the application of automated statistical procedures without metadata use.

The selection of partner series and how they are applied in the homogenisation process significantly impacts both the detection and the correction of inhomogeneities in the candidate series (Szentimrey 2010). There is no scientific consensus regarding the use of partner series, partly because its impact on the results also depends on other technical aspects of the homogenisation method. However, likely to be of greater importance is the lack of objective knowledge about the impact of partner series selection on the efficiency of homogenisation methods.

The most frequently applied modes in the use of partner series are the creation of only one reference series based on the weighted average of the partner series, and the deduction of the relevant adjustment terms from the difference between the candidate and reference series. These methods are included in e.g. the Standard Normal Homogeneity Test (Alexandersson & Moberg 1997), Multiple Linear Regression (Vincent 1998) and the RHTest (Wang et al. 2007). Other modes of the use of partner series include: (1) pairwise comparison in inhomogeneity detection, in PRODIGE (Caussinus & Mestre 2004), HOMER (Mestre et al. 2013) and the USHCN homogenisation (Menne & Williams 2009); (2) calculation of the minimum residual variance of time series for finding the best adjustment terms (ANOVA correction) in HOMER and ACMANT (Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series) (Domonkos & Coll 2017); (3) derivation of confidence intervals of adjustment terms from the use of multiple reference series (MASH; Szentimrey 1999); (4) optimal interpolation for the optimal weighting of partner series (Szentimrey 2010); and (5) interpolation with orthogonal regression (Climatol; Guijarro 2014).

When considering the best selection criteria for the most relevant set of partner series, views within the research community are even more diverse. Karl & Williams (1987) proposed to build a reference series using as many time series as available from the climatic region of the candidate series. By contrast, Peterson & Easterling (1994) recommend the use of 2 to 5 series only. The number of partner series is 10 in the work of Kuglitsch et al. (2009) and Manara et al. (2017), while it is 5 in the work of Begert et al. (2008) and only 1 in the derivation of daily adjustment terms in Della-Marta & Wanner (2006) and Mestre et al. (2011). High spatial correlations between the candidate series and the partner series are a general expectation, but views on what constitutes an acceptable minimum vary. Thus, for example, the suggested threshold is 0.8 by Peterson & Easterling (1994), Alexandersson & Moberg (1997) and Kuglitsch et al. (2009), whereas it is 0.7 by Auer et al. (2005), DeGaetano (2006) and Brunet et al. (2008), 0.6 by Vicente-Serrano et al. (2010) and Trewin (2013) and only 0.4 by Mitchell & Jones (2005) and Domonkos & Coll (2017). Geographical distance is also used instead of spatial correlation in some homogenisation methods, e.g. in Climatol, and in MASH optionally. The Climatol functionality incorporated within HOMER also allows the operator to select geographical distance as an option for reference network selection (e.g. Coll et al. 2014).

In this study, experiments with the monthly temperature homogenisation program of the automatic method ACMANT are presented. The principal aim is to determine the optimal number of partner series and to objectively determine the necessary spatial correlation for their selection. The first version of ACMANT (Domonkos 2011) was created during the European project COST ES0601 (referred to as HOME). The objective of its creation was to put together some of the best segments of existing homogenisation techniques with some novel and highly effective mathematical tools. In the experiments with the HOME benchmark dataset (Venema et al. 2012), ACMANT was one of the most effective homogenisation methods, and later tests (Killick 2016, Guijarro et al. 2017) confirmed its high performance. ACMANT is much faster than other available automatic methods, which is an additional favourable characteristic in the homogenisation of large datasets. Recently, version 3 of ACMANT was published (Domonkos & Coll 2017), and the experiments presented here were undertaken using this

version, with the exception of 1 small modification specified in Section 2.3.

2. METHODS

2.1. Test datasets

To ascertain the efficiency of homogenisation, artificially developed but realistic test datasets are needed with known inhomogeneity properties (Venema et al. 2012, Williams et al. 2012). As an initial step, the homogeneous sets of 2 segments of the recently developed US daily surface air temperature benchmark dataset were selected for use (Willett et al. 2014, www.metoffice.gov.uk/hadobs/benchmarks/). The 2 segments are 210 time series of the southeastern region (SE) and 158 time series of Wyoming, USA (WY). As monthly homogenisation is tested in the study, first the monthly means were derived from the daily data. The version that is used here does not contain missing data. The original 42 yr long series were lengthened, keeping climatic characteristics unchanged insofar as it is feasible (see the Appendix, Section 1). Climatic trends were then added and inhomogeneities inserted in various ways. Additional noise was also added to some of the datasets, in order to produce a few datasets with lower spatial correlations than those of the original data. Altogether, 14 test datasets were created, 7 each from the homogeneous bases SE and WY. The number of time series has always remained the same as that in the homogeneous base. A short description of the 7 datasets is provided below (more details are provided in the Appendix, Section 4).

(A) 60 yr long series. High spatial correlations, low frequency and small size of inhomogeneities. Typical of European and North American temperature data of the modern era.

(B) 60 yr long series. High spatial correlations, medium frequency and size of inhomogeneities. Significant systematic trend bias is included. Typical of the relatively less accurate data of mid-latitude countries.

(C) 60 yr long series. Moderate spatial correlations, high frequency and often large inhomogeneities. High frequency of short-term, platform-shaped inhomogeneities (15 per 100 yr). Typical of data from relatively poor countries.

(CR) 60 yr long series. Same as (C), but with high spatial correlations.

(D) 60 yr long series. Same as (C), but with much lower frequency of short-term, platform-shaped inhomogeneities (3 per 100 yr).

(E) 100 yr long series. High spatial correlations. The first 50 yr have inhomogeneities as in (B), while the second 50 yr have inhomogeneities as in (A).

(ER) Same as (E), but with moderate spatial correlations.

Note that (CR) and (ER) differ from (C) and (E) only in spatial correlations, but while for (C) the default version has relatively low spatial correlations, for (E) the default version has high spatial correlations.

2.2. Calculation of spatial correlations

One purpose of the study is to reveal the influence of spatial correlations on the role of partner series; therefore, the way in which spatial correlations are calculated is important. Here the correlations are calculated from the monthly resolution data. First the mean annual cycle of data is removed, then the first difference (increment series) is created according to Peterson & Easterling (1994) and subsequent studies applying their methods. The correlations are then calculated for these time series. The use of increment series for the estimation of the spatial correlation is justified by the fact that the increment series are less affected by inhomogeneities than the raw series.

2.3. Homogenisation method

The monthly temperature homogenisation program of ACMANT3 software (Domonkos & Coll 2017) was applied with 1 modification described here.

ACMANT has 2 kinds of break detection routines (where a 'break' is defined as a shift in the means). One is for long-term biases, and for these the smallest distance between 2 adjacent breaks is 3 yr (although when the first detection results are refined, distances may be shortened). The other kind of detection is for the identification of large-size, short-term biases, also termed 'filtering of outlier-periods' in ACMANT.

ACMANT generally uses each partner series that has a common time period and at least a 0.4 correlation with the candidate series. However, in the version applied here, the maximum number of partner series (N) is 10 in the outlier filtering routines, while it remains unlimited in the routine of break detection with low time resolution. The reason for this modification is that while the difference between local climate and regional climate is expected to be the same or almost the same for the means of long periods, the natural spatial differences in climate might be more

varied for relatively short periods, spanning from 1 mo to 1–2 yr. Therefore, only the most highly correlated 10 partner series are used here in the detection of short-term inhomogeneities.

2.4. Networking

Although ACMANT is a fully automatic method, it requires the preparation of networks that include time series of the same climatic area. In the experiments presented here, different size networks are formed, spanning from 8 to 51 time series network^{-1}. For simplicity, individual networks are prepared for each candidate series; this means that e.g. if a dataset has 210 series, then 210 networks are formed, and this procedure is repeated for each of the network sizes examined. In all cases, the series most correlated with the candidate series are selected as partner series, hence it was easy to automate the procedure. Automated networking with ACMANT was presented first by Domonkos (2015), although the core idea is older (Begert et al. 2008). The main advantage of forming a network for each candidate series is to avoid having candidate series near the borders of geographical clusters, which could easily introduce biased climatic properties of the partner series relative to those of the candidate series. The only known drawback of the method described here is the elevated computational time demand, which is 10 – 30 times higher than that associated with more conservative clustering techniques.

Beyond the above basic mode of networking, another mode is also applied for some datasets. In this mode, the first 10 partner series are selected in the same way as in the basic mode, but in this variation further partner series are accepted only where the correlation with the candidate series is <0.6 (but still ≥0.4). The aim of this experiment is to examine the impact of the reduction in the associated spatial correlations, and this was performed with datasets of medium mean spatial correlations where the amount of correlations falling into the range 0.4–0.6 is sufficient for the experiment.

2.5. Efficiency measures

The best way to characterise efficiency is the presentation of the residual errors of the homogenised data, where 'error' means the deviation from the real climate (Domonkos 2013). Three kinds of residual errors are presented in the study: (1) monthly root mean squared error (RMSEm), (2) annual root mean squared error (RMSEy) and (3) trend bias, i.e. the deviation in the mean linear trend for the entire period of the time series. For each efficiency measure, the arithmetical mean and the threshold for the upper 5% (empirical cumulative probability distribution function [CDF] = 0.95) values are calculated.

3. RESULTS

3.1. Residual homogenisation errors as a function of the number of partner series

All variations of the datasets (A, B, C, D and E) are used with their default mean spatial correlations and their versions for both geographical regions, so that results of 10 individual experiments are averaged for the calculation of the residual errors shown in Fig. 1. As might be expected, the homogenisation errors decrease with increasing number of partner series (N), until it reaches sufficiently high values. RMSEm and RMSEy decline until the number of partner series reaches values between 15 and 20, and any further increase in the number of partner series has no effect on the residual error. However, the picture is slightly different with the residual trend bias; in this case, the tendency to decline remains for sections of N > 20, although the rate of decrease slows gradually and becomes insignificant for the sections associated with the highest number of partner series.

3.2. Differences in the dependence on the number of partner series according to dataset properties

Fig. 2 shows the curves of mean residual errors as a function of increasing number of partner series for each dataset with default mean spatial correlation. Each curve is the average of 2 datasets: one is with data from SE and the other is with data from WY.

It is apparent that the mean size of the residual error depends strongly on the frequency and magnitude of inhomogeneities of the datasets, and that changes in the number of partner series have little effect on these differences. For datasets with small inhomogeneities (dataset A), the residual error starts to increase with increasing number of partner series for N > 20, while the opposite tendency can be seen for the dataset with the most frequent and largest inhomogeneities (dataset C). However, the observed relationships are so weak that it appears the dependence of residual errors on the number of partner

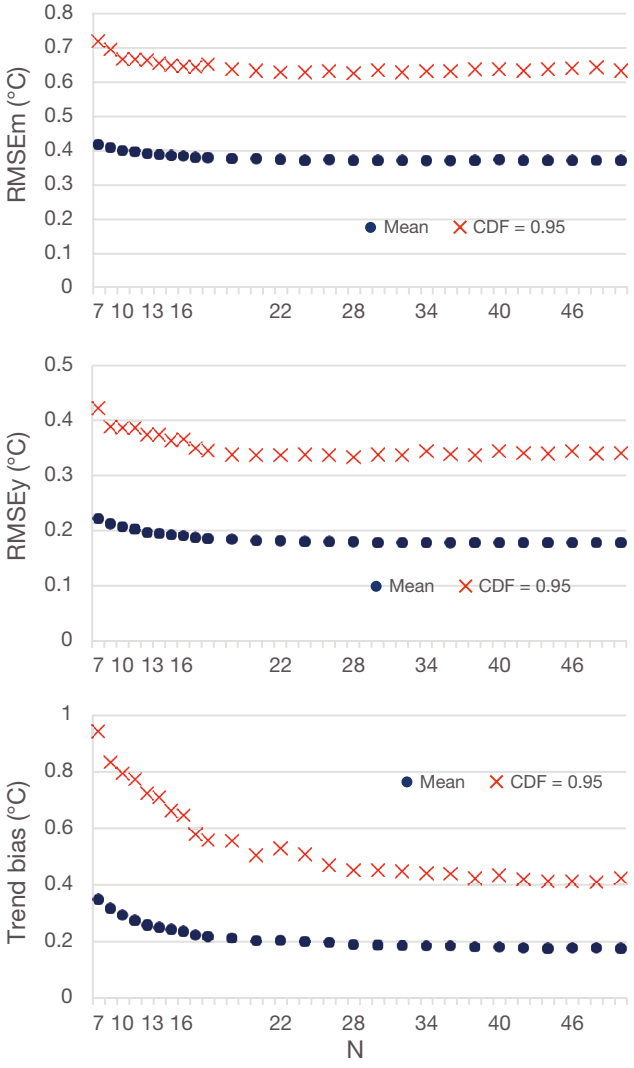

Fig. 1. Residual errors as a function of the number of partner series (N), after homogenisation with ACMANT. Averages for all test datasets with their default spatial correlations and both geographical regions (Wyoming and the southeastern region of the USA). RMSEm (RMSEy): monthly (annual) root mean squared error; CDF: cumulative probability distribution function

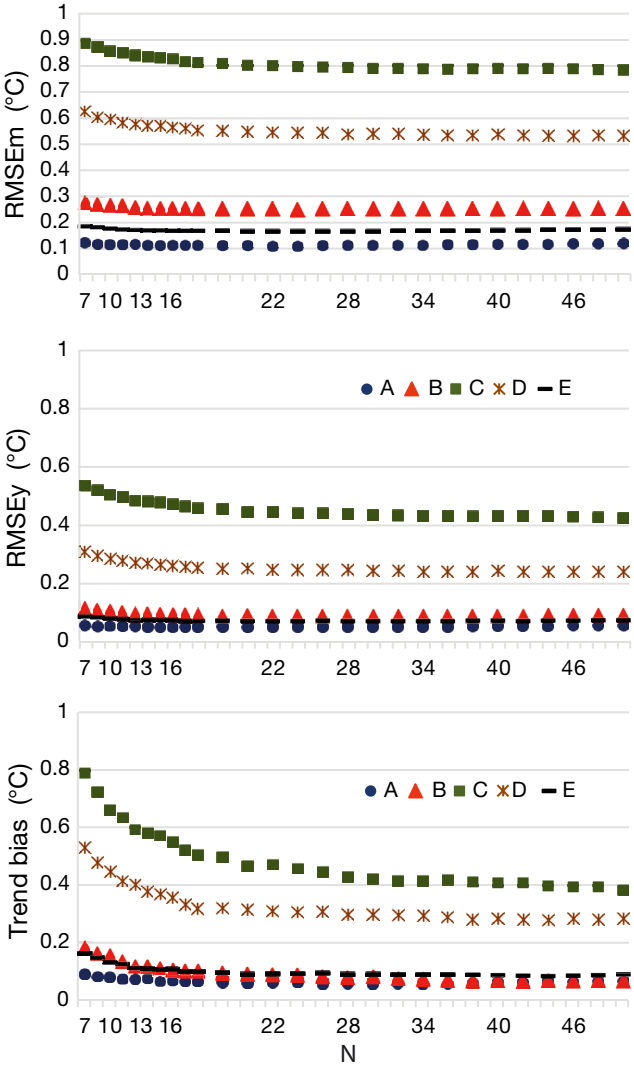

Fig. 2. As in Fig. 1, but for each test dataset (A, B, C, D and E; see Section 2.1 for details) with its default spatial correlation. Each curve is the average of 2 datasets: data from the southeastern USA and that from Wyoming

series does not have a significant connection with the inhomogeneity properties of the datasets.

3.3. Differences in the dependence on the number of partner series according to geographical regions

All variations of the datasets are again used with the default spatial correlations, and the relevant sections of the results are averaged for each of the 2 geographical regions. The differences in the mean errors between WY and SE are then expressed as the percentage of the mean error for the SE region (Fig. 3).

It can be seen that the residual errors are always larger for WY than for SE, except for the trend bias in a few experiments with a small number of partner series. The difference between the efficiency for WY and SE increases rapidly initially with a growing number of partner series, but the increase becomes much slower for N > 20. However, this moderate increase in the difference associated with the growing number of partner series implies that the dependence on the number of partner series for the individual regions is not as flat as it appears from their average (Fig. 1). For instance, for RMSEm and RMSEy, the residual error increases in WY and declines in SE with a growing number of partner series for N > 20, although these trends are weak.

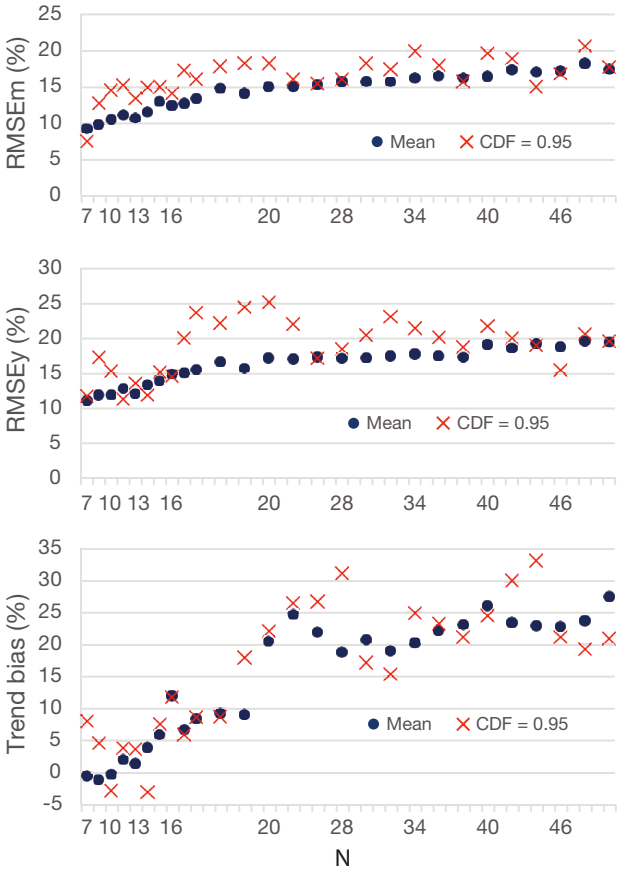

Fig. 3. Exceedance of residual errors for the Wyoming region relative to those of the southeastern (SE) US region in the percentage of the residual error for the SE region. Data are mean for all test datasets with their default spatial correlations. Abbreviations as in Fig. 1

Observations in Fig. 3 also imply that the results of trend bias are strongly scattered, despite each point in Fig. 3 representing the average of 5 experiments with a large number of time series in each.

3.4. Differences in the dependence on the number of partner series according to the spatial correlations

Two kinds of experiments were done for the examination of the dependence on spatial correlations.

In the first experiment, the dependence on the number of partner series for datasets C and E is examined with both kinds of mean spatial correlations. The relevant sections of the results are averaged for both datasets and for both geographical regions, and the differences in the mean residual errors associated with moderate spatial correlations are then compared with those for high spatial correlations and expressed as the percentage of the residual errors with high spatial correlations (Fig. 4). As might be expected, the residual errors associated with moderate correlation are always higher than those associated with high correlations. The difference increases with increasing number of partner series until about N = 25 and remains near constant thereafter. The increasing difference with an increasing number of partner series for low N demonstrates that the increase in the number of partner series is more beneficial for efficiency when the correlations are high than in the opposite case. On the other hand, the nearly constant difference of residual errors for the 2 kinds of mean spatial correlations for N > 25 implies that the correlation does not significantly influence the residual error trends with increasing number of partner series for high N.

In the other experiment, 2 different network configurations are assembled for datasets C, D and ER (see Section 2.4). In 1 of the network formations, the spatial correlations are <0.6 for all but 10 of the partner series, and as a consequence of this, the mean correlation in networks sharply declines with a growing number of partner series (Fig. 5). Fig. 6 shows the difference in residual errors resulting from the networking with limited spatial correlation compared to those from basic networking, in the percentage of the error with basic networking. Examining each error type one by one, it is apparent that the differences are very small for RMSEm, where an increasing trend in the difference with increasing number of partner series appears only for N > 35, but the difference never reaches 10% even in this situation. Regarding RMSEy, the differences are even smaller, i.e. they never reach 5%. The trend in the difference of trend biases with increasing number of partner series shows a somewhat different picture; here, a clear increasing trend appears until about N = 15, and the difference slightly exceeds 10% at that point. However, thereafter the difference declines, while for the cases where the number of partner series is >35, the difference in the mean trend bias is consistently near zero.

4. DISCUSSION

4.1. Optimal number of partner series in homogenising with ACMANT

Lindau & Venema (2016) showed that the performance of any homogenisation strongly depends on the signal-to-noise ratio of the time series analysed, where the signal is equal to the inhomogeneities

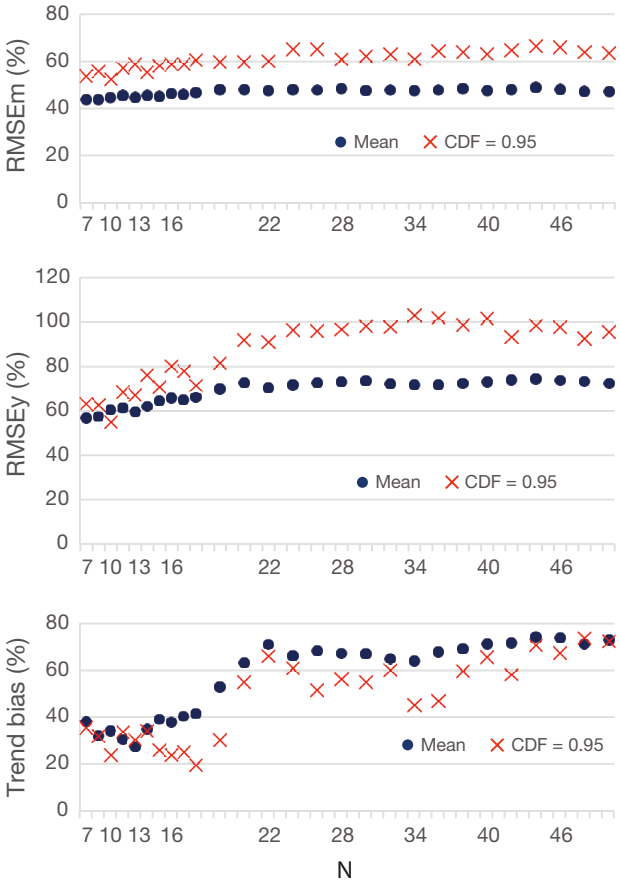

Fig. 4. Exceedance of residual errors for datasets with medium spatial correlations relative to those with high spatial correlations in the percentage of the residual error for high spatial correlations. Data are mean for the datasets of C and E and for both regions (Wyoming and southeastern USA). Abbreviations as in Fig. 1

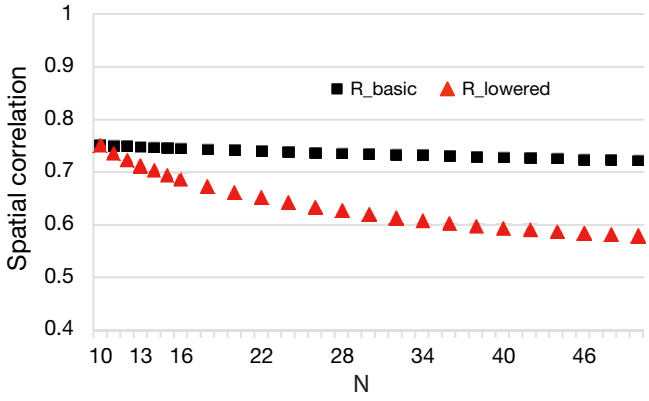

Fig. 5. Mean spatial correlation (R) of the partner series with the candidate series as a function of the number of partner series (N), for the basic networking and for the networking with lowered spatial correlations (i.e. R < 0.6 for each newly selected partner series when N > 10)

included, while the noise is equal to the deviation of the homogeneous series from the true climate signal. In relative time series, the averaged inhomogeneities of the partner series act as additional noise. As the inhomogeneities of individual partner series are usually independent, the noise of the relative time series is expected to decrease with averaging a higher number of partner series. However, incorporating partner series at a greater distance from the candidate series might cause the incorporation of additional noise due to the dissimilarity of climate. The examinations presented here characterise quantitatively the effect of the number of partner series on the efficiency of the homogenisation with ACMANT.

The results suggest that the optimal network size in dense datasets is generally about 30 time series, and the number of partner series should be lower for optimising residual RMSE than for optimising residual trend bias. The optimum network size does not show any significant connection with the characteristic inhomogeneity size in raw data, nor with the frequency of inhomogeneities. Surprisingly, spatial correlations do not have a significant impact on the optimum network size either. While the correlations of the 10–15 series most highly correlated with the candidate series has a strong influence on the residual errors, the correlations of other partner series seem to be near neutral.

The incorporation of additional partner series with at least 0.4 correlation is therefore always beneficial or at least not unfavourable for the efficiency, at least in situations where the network is smaller than the optimal size. In fact, a clearly increasing residual error with an increasing number of partner series was found only in a few experiments, in RMSEm when N > 30. Even when the correlations are <0.6 for the newly selected partner series, and where the candidate series have several partner series with much higher correlations, no increase in residual errors with increasing number of partner series (other than RMSEm with N > 30) appears. This implies that the inclusion of partner series from adjacent climatic areas is often favourable for the efficiency of homogenisation, even when the candidate series has several partner series in the same climatic area. Two interpretations may explain this result. (1) The detection of inhomogeneities is safer with a large number of partner series than in small networks, while climatic differences between adjacent regions tend to have smaller impact on the accuracy of the results than inhomogeneity detection errors. (2) The other possible explanation is that the low frequency climate variability has higher spatial similarity than month-to-month changes, with which the spatial cor-

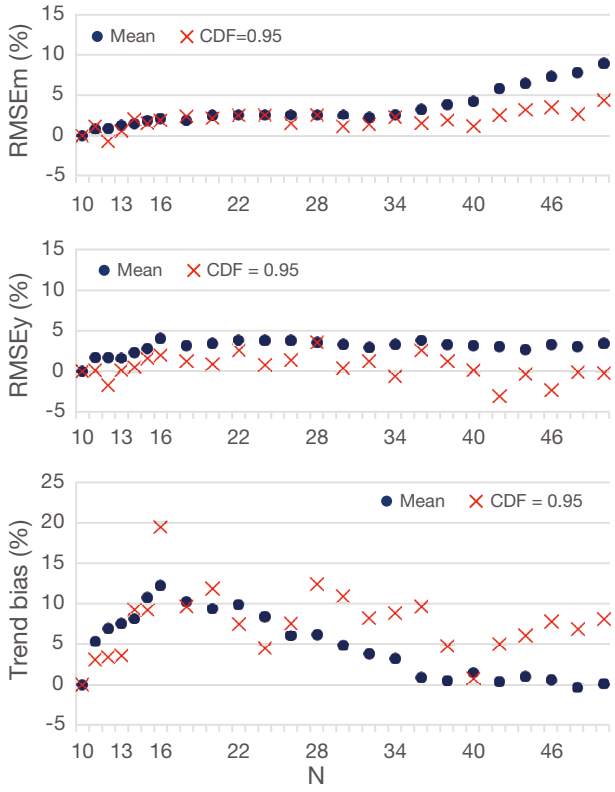

Fig. 6. Exceedance of residual errors for networking with lowered spatial correlations relative to those with basic networking in the percentage of the residual error for basic networking. Data are mean for the datasets of C, D and ER and for both regions (Wyoming and the southeastern USA). Abbreviations as in Fig. 1

relation is characterised in this study. The results suggest that the best minimum threshold of spatial correlation could be even lower than 0.4, but the further reduction of this threshold is not recommended, as the spatial similarity of low frequency variability over partly varied climatic areas might vary in space and time. Therefore, a further reduction in spatial correlations would include some risk to reliability, and would represent a kind of risk which is difficult to quantify.

It is further recommended that no more than 10 partner series should be used for outlier filtering and for the identification of other short-term biases. There is no accurate information available about the short-term variability of spatial climatic gradients, but the results of this study show that the potential decrease in residual RMSE is small when raising the number of partner series above 10, while the long-term trend bias is much less affected by short-term quality problems than RMSE. Taken together, these all indicate that even if the optimal number of partner series for short-term quality problems was somewhat higher than 10, any additional error arising from this is likely to be small.

In this study, we used complete and synchronous time series, and we note that this is not always the case in real homogenisation tasks. If the time series of the dataset do not cover the same period, one must consider — in the decision about the number of partner series — the amount of truly comparable data for all parts of the candidate series. It is recommended that on average ≥15–20 spatially comparable data for each decadal section of the candidate series should be included within the network where possible, even in situations where this choice elevates the number of partner series considerably.

It is a general expectation in time series homogenisation that partner series with high correlation with the candidate series should better favour the accuracy of the homogenisation than those with lower correlation, as higher spatial correlation is generally interpreted to mean more, and more accurate, added information about the climate at the site of the candidate station. However, we did not find this relationship for partner series of >20 rank order positions where the partner series are ordered by their correlation with the candidate series. In addition, in one kind of experiment, the inclusion of partner series of <0.6 correlation reduced the residual trend bias more than the inclusion of partner series of higher correlation for serial numbers between 15 and 35. At present, which specific feature of ACMANT or the experimental setup causes this paradox cannot be adequately explained, and hence requires further examination.

The dependence of the residual errors on the number of partner series is different for the two geographical regions included in the analysis; in turn, this indicates that it would be useful to incorporate more geographical regions in further such analyses. Similarly, the inclusion of other homogenisation methods for future experimentation would be beneficial. However, the computational time demand for these kinds of experiments is high, firstly due to the very high number of homogenisations performed (e.g. some 96 000 network homogenisations are involved in the present study), and secondly, for the enhanced time demand of homogenising large networks (N > 30), even with a relatively fast automated method such as ACMANT. Therefore the extension of this work in the near future is unlikely.

The results of residual trend bias show a wide scatter, indicating a large random component, despite there being a large number of time series involved in the calculations. For instance, and to give this con-

text, 1050 time series of SE and 790 time series of WY were used to produce the results presented in Fig. 3. The experimental trend bias has a much larger random component than RMSE has; as for trend bias calculations, there is only one piece of results for each series, and these results are statistically dependent on the series being homogenised together. It is useful to note here that in the widely referenced HOME benchmark tests, only 111 temperature series and 111 precipitation series were used (Venema et al. 2012). Although there are some newer efficiency tests, for instance under the Spanish project MULTITEST (Guijarro et al. 2017), in general the overall quantity and variety of test experiments is not sufficient. Consequently, a targeted international effort is needed in this specific area of the discipline in order to base time series homogenisation on objective knowledge rather than on poorly tested hypotheses.

4.2. Comparison with other homogenisation methods

For the homogenisation of large datasets of spatially dense observations, the use of automatic methods is the most practical. MULTITEST tests 13 versions of 6 automatic homogenisation methods (only automatic methods are tested in this project). The participating methods are Climatol, ACMANT, MASH, RHTest, USHCN and HOMER. In an experiment of MULTITEST, a comparison was made in homogenising the same kind of test dataset with 9, 19, 39 and 79 available partner series (Guijarro et al. 2017). Its results are broadly similar to those of the present study, except that a very small, insignificant decrease in the residual errors also appears above 39 partner series both for RMSEm and trend bias. One possible explanation for the difference in the MULTITEST results in comparison with the results of our study is that the spatial differences of the climate in MULTITEST are modelled with white noise, while their structure is more realistic in the datasets of the present study.

According to which homogenisation method is used, differences in the speed of decrease in residual errors with increasing numbers of available partner series are generally very small, with some exceptions. The residual errors of HOMER fall spectacularly with an increase in the number of partner series, which is partly due to the low efficiency when only 9 partner series are used. It should be noted that HOMER in automatic mode is used only for tests; the operational mode of HOMER is interactive, and using it in the prescribed way, the efficiency of HOMER is higher than in the MULTITEST experiments. With an increasing number of partner series, RMSEm declines faster with Climatol than with other methods, and while in most of the experiments the residual errors of ACMANT are the smallest, in cases that use 39 or 79 partner series, the RMSEm of Climatol is the smallest. Considering all the examinations of MULTITEST performed so far, ACMANT showed the highest efficiency, followed closely by Climatol. The residual errors of the other methods are substantially larger, although each method markedly reduced the raw data errors. Note here that in the international tests of daily temperature homogenisation methods (Killick 2016), ACMANT and Climatol also produced the highest efficiencies, and they were tied for first place.

Although Climatol uses composite reference series in a fashion similar to ACMANT, the Climatol approach to find the homogenised series is specific. It applies a large number of iterations using a redefined set of partner series in each step. Therefore, in spite of Climatol using only 10 partner series in a particular step, it can take advantage of the availability of a large number of partner series. Consequently, the same kind of optimisation that is presented here for ACMANT is not needed for Climatol.

5. CONCLUSIONS

Artificially developed but realistic monthly temperature datasets were homogenised with varied networking combinations in order to find the optimal number of partner series (N) and identify the requisite spatial correlations associated with the candidate series. The main findings are:

(1) In dense datasets, the generally optimal number of partner series is about 30, and this optimum has low dependence on the frequency and characteristic size of inhomogeneities. The optimal number of partner series is lower for improving residual RMSE than for optimising trend bias. However, for the detection of outlying values and other short-term quality problems, a much smaller number of partner series is recommended.

(2) While the mean correlation with the candidate series has a strong impact on residual errors, it has no significant connection with the optimum number of partner series. The inclusion of partner series of ≥0.4 correlation with the candidate series is always beneficial (or at least not unfavourable), in cases where the network size is smaller than the optimum size.

(3) The impact of spatial correlations on the efficiency of homogenisation is more pronounced for the 10–15 partner series most highly correlated with the candidate series than for the other partner series in the network.

(4) When the time series included in the network cover different time periods, it has to be considered (in the selection of the number of partner series) that the amount of data truly available for spatial comparison may be much less than for complete and synchronous time series.

The results presented here characterise first of all the optimum number of partner series and necessary spatial correlation, achieved through a number of experiments with ACMANT, and using the temperature climate associated with the source data. As the efficiency of ACMANT is often the highest among the efficiencies observed in international method comparisons, and the use of ACMANT is easy and fast, we recommend the use of ACMANT for the homogenisation of large temperature datasets.

Acknowledgements. Thanks to Kate Willett and her colleagues for providing open access to the temperature database they developed. J.C. acknowledges funding provided by the Irish Environmental Protection Agency under project 2012-CCRP-FS.11

LITERATURE CITED

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J (2003) Guidelines on climate metadata and homogenization. WMO-TD No. 1186, WCDMP No. 53. World Meteorological Organization, Geneva

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

Auer I, Böhm R, Jurkovic A, Orlik A and others (2005) A new instrumental precipitation dataset for the Greater Alpine Region for the period 1800–2002. Int J Climatol 25: 139–166

Auer I, Böhm R, Jurkovic A, Lipa W and others (2007) HISTALP – historical instrumental climatological surface time series of the Greater Alpine Region. Int J Climatol 27:17–46

Begert M, Zenklusen E, Häberli C, Appenzeller C, Klok L (2008) An automated procedure to detect discontinuities; performance assessment and application to a large European climate data set. Meteorol Z 17:663–672

Brunet M, Saladié O, Jones P, Sigró J and others (2008) A case-study/guidance on the development of long-term daily adjusted temperature datasets. WMO-TD No. 1425, WCDMP No. 66. World Meteorological Organization, Geneva

Caussinus H, Mestre O (2004) Detection and correction of artificial shifts in climate series. J R Stat Soc Ser C 53: 405–425

Coll J, Curley C, Walsh S, Sweeney J (2014) Ireland with HOMER. In: Lakatos M, Szentimrey T, Marton A (eds) Proc 8th Seminar for Homogenisation and Quality Control in Climatological Databases and 3rd Conference on Spatial Interpolation in Climatology and Meteorology. WMO WCDMP No. 84. World Meteorological Organization, Geneva, p 23–45

DeGaetano AT (2006) Attributes of several methods for detecting discontinuities in mean temperature series. J Clim 19:838–853

Della-Marta PM, Wanner H (2006) A method of homogenizing the extremes and mean of daily temperature measurements. J Clim 19:4179–4197

Domonkos P (2011) Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT). Int J Geosci 2:293–309

Domonkos P (2013) Measuring performances of homogenization methods. Idojaras 117:91–112

Domonkos P (2015) Automatic networking for the homogenization of large climatic datasets. In: 15th Annual Meeting of the European Meteorological Society, Sofia, EMS2015-84

Domonkos P, Coll J (2017) Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. Int J Climatol 37: 1910–1921

Guijarro JA (2014) User's guide to Climatol. www.climatol.eu/climatol-guide.pdf

Guijarro JA, López JA, Aguilar E, Domonkos P, Venema V, Sigró J, Brunet M (2017) Comparison of homogenization packages applied to monthly series of temperature and precipitation: the MULTITEST project. Ninth Seminar for Homogenization and Quality Control in Climatological Databases (in press)

Karl TR, Williams CN Jr (1987) An approach to adjusting climatological time series for discontinuous inhomogeneities. J Clim Appl Meteorol 26:1744–1763

Killick RE (2016) Benchmarking the performance of homogenisation algorithms on daily temperature data. PhD thesis, University of Exeter

Kuglitsch FG, Toreti A, Xoplaki E, Della-Marta PM, Luterbacher J, Wanner H (2009) Homogenization of daily maximum temperature series in the Mediterranean. J Geophys Res 114:D15108

Lindau R, Venema V (2016) The uncertainty of break positions detected by homogenization algorithms in climate records. Int J Climatol 36:576–589

Manara V, Brunetti M, Maugeri M, Sanchez-Lorenzo A, Wild M (2017) Homogenization of a surface solar radiation dataset over Italy. AIP Conf Proc 1810:090004

Menne MJ, Williams CN (2009) Homogenization of temperature series via pairwise comparisons. J Clim 22: 1700–1717

Menne MJ, Williams CN, Vose RS (2009) The US Historical Climatology Network monthly temperature data, version 2. Bull Am Meteorol Soc 90:993–1007

Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S (2011) SPLIDHOM: a method for homogenization of daily temperature observations. J Appl Meteorol Climatol 50: 2343–2358

Mestre O, Domonkos P, Picard F, Auer I and others (2013) HOMER: homogenization software in R—methods and applications. Idojaras 117:47–67

Mitchell TD, Jones PD (2005) An improved method of constructing a database of monthly climate observations and associated high-resolution grids. Int J Climatol 25:

693–712
Peterson TC, Easterling DR (1994) Creation of homogeneous composite climatological reference series. Int J Climatol 14:671–679
Peterson TC, Easterling DR, Karl TR, Groisman P and others (1998) Homogeneity adjustments of in situ atmospheric climate data: a review. Int J Climatol 18:1493–1517
Szentimrey T (1999) Multiple Analysis of Series for Homogenization (MASH). In: Szalai S, Szentimrey T, Szinell Cs (eds) Proc 2nd Seminar for Homogenization of Surface Climatological Data. WMO WCDMP-41, Geneva, p 27–46
Szentimrey T (2010) Methodological questions of series comparison. In: Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds) 6th Seminar for Homogenization and Quality Control in Climatological Databases. WMO WCDMP-76, Geneva, p 1–7
Trewin B (2013) A daily homogenized temperature data set for Australia. Int J Climatol 33:1510–1529
Venema V, Mestre O, Aguilar E, Auer I and others (2012) Benchmarking monthly homogenization algorithms. Clim Past 8:89–115
Vicente-Serrano SM, Beguería S, López-Moreno JI, García-Vera MA, Štěpánek P (2010) A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. Int J Climatol 30:1146–1163
Vincent LA (1998) A technique for the identification of inhomogeneities in Canadian temperature series. J Clim 11: 1094–1104
Wang XL, Wen QH, Wu Y (2007) Penalized maximal *t* test for detecting undocumented mean change in climate data series. J Appl Meteorol Climatol 46:916–931
Willett KM, Williams CN, Jolliffe I, Lund R and others (2014) A framework for benchmarking of homogenisation algorithm performance on the global scale. Geosci Instrum Method Data Sys 3:187–200
Williams CN, Menne MJ, Thorne P (2012) Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. J Geophys Res 117: D05116

Appendix. Methodological details

1. Homogeneous base

The homogeneous base is taken from the benchmark for testing daily temperature homogenisation algorithms, developed by Willett et al. (2014; www.metoffice.gov.uk/hadobs/benchmarks/). Two segments were selected: one is the whole set of the US southeastern region (SE, 210 series) and the other is the whole set of Wyoming (WY, 158 series). First, the daily data are aggregated to deduce monthly mean values, then the series are lengthened with the repeated use of data in order to have 60 and 100 yr long series. In the original set, time series are 42 yr long, and from Year 43 of the lengthened series, the data of the original set are repeated using the data in the reverse order of years. Thus, Year 43 is the same as Year 41, Year 44 is the same as Year 40, etc. until Year 83, which is the same as Year 1. Then, keeping with the same logic, the order of years reverses again and Year 84 is the same as Year 2, Year 85 is the same as Year 3, etc., until Year 100. By means of this manipulation of the long dataset, the spatio-temporal structures of low-frequency oscillations remain similar to those in the original set, as data of the subsequent years in the new set are also of subsequent years in the original set. Finally, climatic trends are added, which halt the temporal repetitions of data. These climatic trends are entirely arbitrary but spatially uniform, and thus they are expected to be neutral for relative homogenization methods.

2. Spatial correlations

Spatial correlations of the first difference (increment) series are controlled. For the homogeneous base of the deseasonalised monthly mean values, the mean spatial correlation is 0.88 for the SE region and 0.83 for WY. In several experiments, additional noise terms are included in the homogeneous base in order to examine cases with reduced spatial correlations. These additional noise terms are generated by monthly series of normally distributed red noise with 0 mean and 0.15 autocorrelation. The variance of the noise is a function of the correlation with the 'central' series, where central series means the series with the highest mean correlation with all the other series in the regional set. The parameters of the noise generator are determined empirically to obtain homogeneous sets with 0.65 mean spatial correlation both for SE and WY.

3. Inserted outliers and inhomogeneities: common characteristics for each test dataset

Monthly outlier values and 3 types of inhomogeneities are inserted into the test datasets. All types of inhomogeneities are changes in the mean, and their forms are: (1) sudden shift, or (2) gradually changing bias (linear trend shaped inhomogeneity), or (3) platform shaped inhomogeneity with a pair of shifts of the same magnitude, but opposite direction.

The positions and properties of inhomogeneities are determined via the use of a random number generator. Consequently—and although in a given dataset all types of inhomogeneities occur with equal probability in time series—the number and sizes of inhomogeneities randomly vary between time series.

The lengths of the trend inhomogeneities are taken from the even distribution between 5 and 100 yr, but it should be noted that the effective length of trend inhomogeneities, falling between the start and end points of time series, is often shorter, whereas the length of platform inhomogeneities varies between 1 and 120 mo, and the frequency quadratically decreases with increasing length. Thus, most platform shaped biases are very short lived, and in the special case of 1 mo duration, an outlier value is generated. These outliers are beyond those which are directly generated and inserted by the outlier inserting section of the dataset development.

The size of outliers (for those that are not degenerated platforms) is taken from an even distribution with predefined thresholds.

The size of inhomogeneities is taken from a normal distribution with zero mean, and the standard deviation of the distribution varies according to datasets. The size distribution is the same for trend inhomogeneities as for single shifts, but when part of the trend inhomogeneity falls out of an end point of the time series, the effective size of the trend inhomogeneity is smaller. Sizes are often elevated for platform inhomogeneities, thus it is a dataset-dependent characteristic.

The sequence of inhomogeneities can be characterised as a limited random walk. The kind and size of an inhomogeneity does not depend on the previous inhomogeneities, but the accumulated bias size is limited. When the accumulated bias size would exceed a predefined threshold when inserting the last generated inhomogeneity, the generation of inhomogeneity size is repeated to avoid the exceedance. In some test datasets, bias limits are asymmetric to zero, resulting in a systematic trend bias.

Synchronous inhomogeneities in >1 time series might occur accidentally, but such events are not generated intentionally.

In all test datasets and for all kinds of inhomogeneities, 75% of the inhomogeneities introduce biases with a seasonal cycle, while in the other cases the introduced bias is constant throughout the year. Two kinds of seasonality of bias are applied: one is semi-sinusoid with modes in July and December, typical for the radiation cycle in mid-latitudes, while the other is more irregularly shaped, with modes in April and August–September, and hence more typical of a monsoon climate. The size of the seasonal cycle of bias is taken from an even distribution, and distinct limit bias values are applied for the accumulated size of such cycles.

4. Dataset properties for each test dataset

The properties of datasets A, B, C and E are summarised in Table A1. The other test datasets are derived from one of these datasets by adding slight modifications.

CR: same as (C), except the mean spatial correlation is 0.85.

D: same as (C), except the frequency of platform inhomogeneities is 3 per 100 yr.

ER: same as (E), except the mean spatial correlation is 0.65.

Table A1. Dataset properties for test datasets A, B, C and E. Frequencies are denoted by the number of occurrences per 100 yr unit; IH: inhomogeneity, σ: empirical standard deviation of the deseasonalised values of the homogeneous dataset. SD: standard deviation

	A	B	C	E	
				First half	Second half
Length (yr)	60	60	60	100	
Spatial correlation	0.85	0.85	0.65	0.85	
Frequency of outliers	2	5	10	5	2
Size of outliers (1 = σ)	0…5	0…5	0…6	0…5	0…5
Frequency of shifts	4	4	6	4	4
Frequency of trends	1	1	1	1	1
Frequency of platforms	2	3	15	3	3
SD of IH size (°C)	0.5	0.8	1.5	0.8	0.5
Size elevation for platform IHs (%)	50	30	0	30	30
Limit bias (°C)	−1.25; 1.25	−1.0; 3.0	−4.0; 6.0	−1.0; 3.0	−1.25; 1.25
Type of bias cycle	Mid-latitude	Mid-latitude	Monsoon	Mid-latitude	
Size of bias cycle (°C)	−1.5; 1.5	−1.0; 4.0	−3.0; 3.0	−1.0; 4.0	−1.5; 1.5
Limit size of bias cycle	−2.0; 2.0	−1.0; 4.0	−6.0; 8.0	−1.0; 4.0	−2.0; 2.0

Editorial responsibility: Toshichika Iizumi, Tsukuba, Japan

Submitted: March 15, 2017; Accepted: August 28, 2017
Proofs received from author(s): October 15, 2017