Advances in
Science & Research
Open Access Proceedings

Newest developments of ACMANT

P. Domonkos, R. Poza, and D. Efthymiadis

Centre for Climate Change (C3), Geography Dept., University Rovira i Virgili,
Campus Terres de l'Ebre, C/Betánia 5 Tortosa, 43500, Spain

Abstract. The seasonal cycle of radiation intensity often causes a marked seasonal cycle in the inhomogeneities (IHs) of observed temperature time series, since a substantial portion of them have direct or indirect connection to radiation changes in the micro-environment of the thermometer. Therefore the magnitudes of temperature IHs tend to be larger in summer than in winter. A new homogenisation method, the Adapted Caussinus – Mestre Algorithm for Networks of Temperature series (ACMANT) has recently been developed which treats appropriately the seasonal changes of IH-sizes in temperature time series. The performance of ACMANT was proved to be among the best methods (together with PRODIGE and MASH) in the efficiency test procedure of COST ES0601 project. A further improved version of the ACMANT is described in this paper. In the new version the ANOVA procedure is applied for correcting inhomogeneities, and with this change the iterations applied in the earlier version have become unnecessary. Some other modifications have also been made, from which the most important one is the new way for estimating the timings of IHs. With these modifications the efficiency of the ACMANT has become even higher, therefore its use is strongly recommended when networks of monthly temperature series from mid- or high geographical latitudes are subjected to homogenisation. The paper presents the main properties and the operation of the new ACMANT.

1 Introduction

The investigation of climate change and climate variability needs a large amount of observed data of high quality. In the last decades a new branch of quality control and quality improvement for observed data has been developed, it is the so-called homogenisation. The purpose of homogenisation is to filter out the effects of technical imperfectness, i.e. that of methodological or environmental changes from the results of observations (Aguilar et al., 2003; Auer et al., 2005; Gérard-Marchant et al., 2008, etc.). The most frequent form of IHs is a sudden shift (change-point) in the series of the data. The documents of technical changes and the statistical properties of datasets can be used for homogenising time series. During statistical homogenisation the spatial redundancy can be utilised for identifying and adjusting local biases. The spatial redundancy means that the same climatic sign often appears in more than one time series. Hereafter homogenisation means statistical homogenisation in this study. In spite of the spatial redundancies, observed data contains real site-specific differences, therefore the perfect homogenisation is usually impossible, but homogenisation procedures have stochastic behaviour (Sherwood, 2007; Menne and Williams, 2009; Titchner et al., 2009, etc.). The success of homogenisation depends on (i) the number and completeness of time series and their spatial correlations, (ii) the signal/noise rate which is influenced by the spatial correlations, the standard deviation of the data, and the size of IHs, (iii) the ability of homogenisation methods to identify the timings of change-points, (iv) the ability of homogenisation methods to treat with multiple structures of IHs, (v) the way of adjusting time series for detected IHs. In this study the characteristics of a new homogenisation method, the Applied Caussinus-Mestre Algorithm for homogenising Networks of Temperature series (ACMANT), are discussed. At the time of the submission of this paper, the ACMANT is still in progress. In this paper only some important parts of the method are shown in detail. The full description of the method will be published once the method is finished. However, a tested preliminary version of the method is already available for interested users. For more information one can contact the main author.

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

The main novelty of the ACMANT is that the seasonal changes of IHs are modelled by harmonic functions, because IHs in temperature data are often related to radiation changes, and the seasonal curve of irradiation in mid- and high latitudes can be approached well with harmonic functions. This thesis is supported also by experimental results (Domonkos and Štěpánek, 2009; Brunet et al., 2011).

2   Main properties of ACMANT

In the ACMANT the core of the detection and adjustment methods is the same as in the PRODIGE (the method of Caussinus and Mestre, 2004). The PRODIGE has a strong mathematical background, and earlier studies (Domonkos, 2006), as well as the homogenisation of COST HOME benchmark dataset (hereafter: benchmark; Venema et al., 2010) have proved that the PRODIGE is one of the most effective homogenisation methods. With the recent developments the highest efficiencies reached earlier are now significantly exceeded. The operational features and conditions for the ACMANT are as follows:

  i. The use of the ACMANT is recommended especially for temperature series from mid- and high-latitudes, since its algorithm supposes quasi-harmonic annual cycle of considerable amplitude in IH-sizes

 ii. A fully automated method.

iii. A relative homogenisation method, thus it can be used only for networks, and not for single time series. Reference series are always built from a minimum of two component series. Lengths of time series in a network can be different, and in that case different reference series may be used for different sections of the same candidate series.

 iv. Occurrences of data-gaps are allowed up to 83% for any 30-year inside section of time series, and unlimitedly in the tails of the series (inside section means that there are no data gaps at the edges of the section).

  v. The ACMANT contains separate segments for filling data-gaps and substituting outlier values. Missing data are never filled before (after) the first (last) observed value of time series.

 vi. The input data-field for the ACMANT: Monthly temperature characteristics with monthly time resolution. The lengths of the original time series may be different, but the data-fields of each series are required to be converted into a common format (which format includes the same number of data for each temperature series) in a way that missing values are filled with −999.9. After preparation only 4 parameters have to be introduced before application: (a) length of time series, (b) first year

of time series, (c) number of time series in the network, (d) identifier of network.

vii. The result of homogenisation is (a) timings and sizes of IHs for each series, (b) timings of outliers, (c) filled data-gaps caused by missing values or outliers inside the series, (d) homogenised time series. Sizes of IHs are characterised with two variables: (a) shift in annual means, (b) shift in the amplitude of seasonal cycle.

The operation of ACMANT is illustrated in Fig. 1.

3   Description of selected segments of ACMANT

A concise description of the method is provided here, and only segments that are primarily important in achieving higher efficiency are shown in detail.

3.1   Constructing relative time series

The spatial comparison of time series relies on the rules introduced by Peterson and Easterling (1994), with some modifications in the parameterisation.

Anomaly series (A_j) are created first, by subtracting monthly means from the raw values. j is a station-identifier ($j = 1,2,...J$ where J stands for the total number of time series in the network). For the candidate series A_j, a relative time series (T_j) is the arithmetical difference of the candidate series and the so-called reference series (F_j) (Eq. 1).

$$T_j = A_j - F_j \tag{1}$$

Note that in accordance with Sect. 2, (point iii.), more than one relative time series are often created for the same candidate series (this is not shown in Eq. (1) to keep the description brief). Reference series are the weighted averages of neighbouring anomaly series (A_i) around the candidate series where the weights are the squared spatial correlations (r) with the candidate series (Eq. 2). Following the recommendations of Peterson and Easterling (1994), the first difference (increment) series are applied for estimating spatial correlations, since in this way the estimations are less affected by the inhomogeneities in time series.

$$F_j = \frac{\sum_{i=1}^{J^*} r_{j,i}^2 \cdot A_i}{\sum_{i=1}^{J^*} r_{j,i}^2} \quad i = 1,2,...J^*, J^* \leq J \tag{2}$$

In the ACMANT, every A_i with $r_{j,i} \geq 0.4$ is considered for building reference, but a minimum condition for applying the homogenisation is that at least two reference composites have to exist with $r_{j,i} \geq 0.5$.
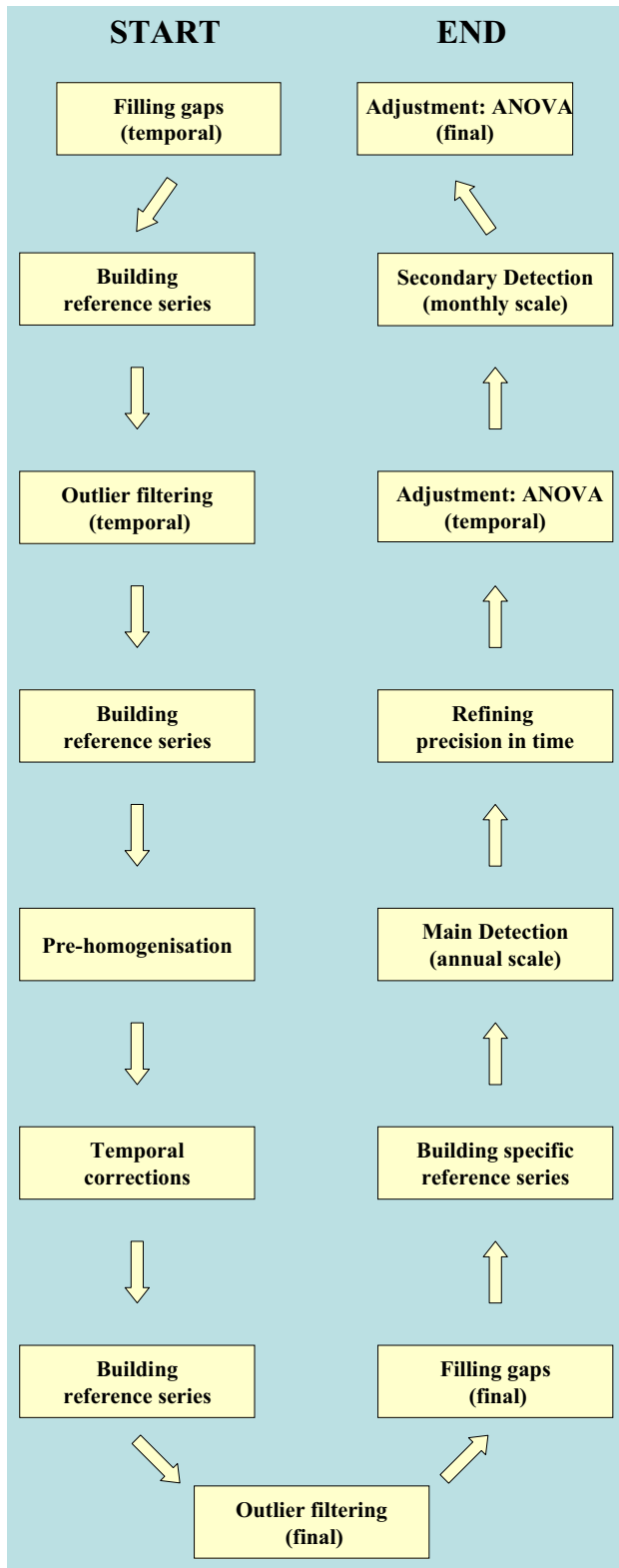
START **END**

Filling gaps
(temporal)

Adjustment: ANOVA
(final)

Building
reference series

Secondary Detection
(monthly scale)

Outlier filtering
(temporal)

Adjustment: ANOVA
(temporal)

Building
reference series

Refining
precision in time

Pre-homogenisation

Main Detection
(annual scale)

Temporal
corrections

Building specific
reference series

Building
reference series

Filling gaps
(final)

Outlier filtering
(final)

Figure 1. The operation of ACMANT.

3.2 Detecting IHs with Main Detection

Timings and sizes of IHs are searched by fitting step-functions to two annual characteristics, i.e. to annual means (TM) and to the range of the seasonal cycle (TD) in relative time series. Solutions with common timings of change-points are considered only, and the minimum sum of squared errors is searched for with the so-called dynamic programming algorithm, described by Hawkins (1972) first. Note that in the ACMANT, similarly to a large number of other methods, gradually changing biases (i.e. trend-like IHs) are represented as a series of change-points.

Let the length of time series be denoted by L, the number of change-points by K, their serial numbers by k ($k = 1,2,...K$), and their timings by y_k. Note that k also shows the serial number of the section between adjacent change-points (Eqs. 3 and 4).

$$\min_{[y_1, y_2, ... y_K]} \left\{ \sum_{k=0}^{K} \sum_{i=y_k+1}^{y_{k+1}} (tm_i - \overline{TM_k})^2 + c_0 (td_i - \overline{TD_k})^2 \right\} \quad (3)$$

$$y_0 = 0, \quad y_{K+1} = L \quad (4)$$

In Eq. (3) upper stroke marks time-average, and $c_0 = 0.5$. The minimum distance between two change-points is set to be 3 years.

For using Eq. (3) K has to be set. For selecting the most appropriate K, the Caussinus – Lyazrhi criterion (Caussinus and Lyazrhi, 1997) is applied (Eqs. 5 and 6).

$$\ln \left\{ 1 - \frac{\sum_{k=1}^{K+1} (y_k - y_{k-1}) \cdot \left[(\overline{TM_k} - \overline{TM})^2 + c_0 (\overline{TD_k} - \overline{TD})^2 \right]}{\sum_{i=1}^{L} (tm_i - \overline{TM})^2 + c_0 (td_i - \overline{TD})^2} \right\} + w \quad (5)$$

$$w = p \frac{2K}{L-1} \ln(L) \quad (6)$$

Equations (5) and (6) are calculated with each possible K, and the K providing the minimum of term (5) is retained. The Main Detection differs in three points from the classic Caussinus-Mestre detection method: (i) Step functions are fitted to two variables, (ii) the minimum distance between two change-points is 3 time-units, (iii) an extra parameter (p) is included in the penalty-term, and its value depends on J. If $J = 2$, then $p = 1.5$, if $J = 3$, then $p = 1$, and if $J > 3$, then $p = 0.75$.

The parameterisation relies on semi-empirical experiences about the changes of signal/noise rate in functions of Δy and J.

3.3 Calculation of timings of change-points with monthly preciseness

48-month wide windows are symmetrically set around the pre-estimated timings of change-points (from the Main Detection). Two-phase functions (U) are fitted to the monthly

values of relative time series within each window. The functions are harmonic functions of 12-month cycle in both phases. The timing of change-point is searched in a narrower, 24-month wide window. Equation (7) shows the calculation for calendar month m.

$$u_m = \alpha + \beta \left(c_A \sin \frac{2\pi m}{12} + c_B \cos \frac{2\pi m}{12} \right) \tag{7}$$

Constants of c_A and c_B are determined in a way to have the modus of the annual cycle in the solstices. In the optimum fitting the sum of squared errors in the 48-month window is minimal. The optimum values for α, β and the timing of change-point are estimated through iterative tests.

3.4 Homogenisation-adjustment with ANOVA

In the detection process an individual IH usually causes biases in more than one time series, since during the spatial comparison each time series is used as reference composite, and homogenizers cannot pre-assume which series are homogeneous (if any). Therefore the calculation of cumulated effects of IHs needs the use of some equation system. In the ANOVA the observed values are considered to be a composition of climate-effect, station-effect and noise, and an equation-system is built and solved for the case of zero noise. Caussinus and Mestre (2004) proved that the equation system of the ANOVA provides the best estimation of IH-sizes. In that paper the full description of the ANOVA is provided.

In the ACMANT the ANOVA is applied to the two annual variables (***TM*** and ***TD***) separately, and thereafter monthly adjustments are calculated as a composition of the shift in annual mean and the relevant value of the seasonal cycle.

4 Some further details of ACMANT

The main characteristics of two further segments are presented here, these are the Pre-homogenisation and the Secondary Detection.

The aim of Pre-homogenisation is to reduce the impact of possible large IHs in the composites of reference series. The detection process is the same as in the Main Detection, but the way of the application is different: (i) Candidate series are ordered according to a pre-estimation of the severity of inhomogeneities. This ordering is based on the maximal absolute values of 5-year mean anomalies in relative time series. (ii) The Main Detection is applied starting from the least homogeneous series and proceeding always towards the more homogeneous series; (iii) Simplified adjustments (e.g. Alexandersson, 1986) are applied just after the detection of IHs in one candidate series; (iv) In building reference series for pre-homogenisation, adjusted versions of composites are used when they are available; (v) In building reference series for Pre-homogenisation, one time series is excluded from taking into account as reference-composite,

namely the one for which the pre-homogenised time series will be used as reference-composite in the Main Detection. As one time series might be used as reference-composite for $J-1$ different candidate series, $J-1$ pre-homogenisations are accomplished for each time series through pre-homogenizing J sub-networks each containing $J-1$ time series. In the step of "Building specific reference series" (Fig. 1) these pre-homogenised series are used.

The aim of the Secondary Detection is to find and correct large-size but short-term biases caused by IHs. Series of accumulated anomalies are examined in time series adjusted by the ANOVA after the Main Detection. If the absolute values of accumulated anomalies exceed some arbitrarily given thresholds, IHs are searched for in monthly series, within a 60-month wide window symmetrically located around the anomaly-peak. The detection method is similar to the procedure described in Sect. 3.3, with two main differences (i) The number of change-points can be 0, 1 or 2 within a 60-month section, and the optimal choice from these three possibilities is determined by the Caussinus – Lyazrhi criterion; (ii) The length of the section between two adjacent change-points is three months minimum, and when it is shorter than 10 months, a constant function is fitted (with zero amplitude of seasonal cycle). After the Secondary Detection the application of the ANOVA is repeated both for ***TM*** and ***TD***.

5 The efficiency of ACMANT

The efficiency of homogenisation procedures can be characterised with the improvement in root mean squared errors (RMSE). We calculated the RMSE characteristics of monthly biases, annual biases and biases of linear trend-slopes for the time series of the benchmark. The efficiency is characterised by the rate of the decrease of RMSE due to the homogenisation, proportioned to the RMSE of raw time series (Eq. 8).

$$\text{Eff} = \frac{W_R - W_H}{W_R} \tag{8}$$

In Eq. (8) W_R (W_H) denotes the RMSE in raw (homogenised) time series. If a homogenisation is perfect, then the efficiency equals to 1, while in case of no change in the RMSE the efficiency is zero. Efficiencies for the ACMANT were calculated separately for the monthly, annual and trend-slope biases. All the 40 simulated temperature networks of the benchmark were used for the calculations. The results show that the efficiency of the ACMANT is 0.545 in monthly RMSE, 0.666 in annual RMSE and 0.753 in trend-slope RMSE. These values are significantly higher than those were achieved by any other method during the COST HOME experiments. The superior performance of ACMANT stems from three main properties: (i) The ACMANT adopts the best detection and adjustment segments of earlier methods. (ii) The ACMANT applies appropriate time-scales during the detection; considering that the standard deviation is smaller

for annual variables than for monthly variables, the application of a relatively coarse time-resolution often yields the best results; regarding this point, a special characteristic of the ACMANT is that the seasonal changes of IH-sizes are estimated with the use of two annual characteristics only. (iii) The ACMANT has a pre-homogenisation part in which the reference composites for a particular candidate series are adjusted in a way that the spatial correlation between the candidate series and reference composites are not utilised at all.

Further verifications are still needed, because in the last phase of the development the IHs of the benchmark were known to the authors. Nevertheless, considering that the verified dataset contains 340 time series with more than one thousand IHs, the impact of previous knowledge of the time series characteristics could cause minor bias in the efficiencies calculated.

6   Conclusions

A new homogenisation method, the ACMANT has been developed. It is applicable for homogenising observational networks of monthly temperature series. While the ACMANT is based on one of the best methods that existed earlier (i.e. on the PRODIGE), it also includes a new treatment of the seasonal cycle of inhomogeneity-sizes and some other modifications. The new version of the ACMANT shows a favourably high efficiency when it is used for homogenising the COST HOME benchmark dataset. The properties described and the efficiency results obtained indicate that the ACMANT is an excellent tool for homogenising networks of temperature datasets from mid- or high latitudes.

Edited by: M. Brunet-India
Reviewed by: K. Willett and another anonymous referee

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.

Auer, I. and 24 coauthors: 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., doi:10.1002/joc.2192, 2011.

Caussinus, H. and Lyazrhi, F.: Choosing a linear model with a random number of change-points and outliers, Ann. Inst. Statist. Math., 49/4, 761–775, doi:10.1023/A:1003230713770, 1997.

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.

Domonkos, P.: Testing of homogenisation methods: purposes, tools and problems of implementation, Proceedings of the 5th Seminar and Quality Control in Climatological Databases, CD, 126–145, 2006.

Domonkos, P. and Štěpánek, P.: Statistical characteristics of detectable inhomogeneities in observed meteorological time series, Studia Geoph. et Geod., 53, 239–260, doi:10.007/ s11200-009-0015-9, 2009.

Gérard-Marchant, P. G. F., Stooksbury, D. E., and Seymour, L.: Methods for starting the detection of undocumented multiple changepoints, J. Climate, 21, 4887–4899, doi:10.1175/2008JCLI1956.1, 2008.

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.

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.

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.

Sherwood, S. C.: Simultaneous detection of climate change and observing biases in a network with incomplete sampling, J. Climate, 20, 4047–4062, doi:10.1175/JCLI4215.1, 2007.

Titchner, H. A., Thorne, P. W., McCarthy, M. P., Tett, S. F. B., Haimberger, L., and Parker, D. E.: Critically reassessing 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.