



# Development and testing of homogenisation methods: moving parameter experiments with ACMANT

P. Domonkos and D. Efthymiadis

Centre for Climate Change, University Rovira i Virgili, Campus Terres de l'Ebre, Tortosa, Spain

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

Received: 21 November 2012 – Revised: 25 February 2013 – Accepted: 8 March 2013 – Published: 18 March 2013

**Abstract.** During the European project COST ES0601 (HOME) a new homogenisation method, ACMANT has been developed for the automatic homogenisation of monthly temperatures. ACMANT turned out to be one of the best performing methods during the blind test experiments of HOME. The methodological development of ACMANT has been continued since then, and nowadays ACMANT is likely the best homogenisation method for large and spatially dense temperature datasets. Ensemble moving parameter experiments have been done to obtain more information about the performance of ACMANT. The HOME Benchmark was used as test dataset, thus the results of the latest experiments with ACMANT are comparable with the performance of the other homogenisation methods participated in HOME. The results indicate that the performance of ACMANT is generally not sensitive to its parameterisation, i.e. the change of the performance is generally small for quite a wide range of each parameter. The presented methodology of moving parameter experiments provides results in a fast and easy to evaluate form.

## 1 Introduction

In the time series of climatic observations, temporal biases from the true macroclimatic characteristics often occur due to technical, personal or environmental changes. With the homogenisation of time series, the frequency and magnitudes of such biases can be reduced. Time series homogenisation can be done with the use of documentary information about the changes in the settings of the observation or with statistical procedures or with the combination of both (Aguilar et al., 2003; Auer et al., 2005, etc.). When the spatial density of observing sites and the spatial correlation of the observed values are high, data quality can be significantly improved by statistical homogenisation. The conditions mentioned are generally met for European and North American temperature data from the last century (Menne and Williams, 2009; Domonkos and Štěpánek, 2009).

A large number of statistical homogenisation methods have been developed in the last decades. Between 2007 and 2011 action COST ES0601 (HOME, [www.homogenisation.org](http://www.homogenisation.org)) was dedicated to test the efficiency of the existing methods and foster further methodological developments. Under HOME, an international blind test experiment was organised,

in which MASH (Szentimrey, 1999), PRODIGE (Causinus and Mestre, 2004), USHCN (Menne and Williams, 2009), Craddock-test (Craddock, 1979) and ACMANT (Domonkos, 2011a, hereafter D2011) produced the best results in homogenising monthly temperature data (Venema et al., 2012). The development of ACMANT has been continued since the blind tests, and as ACMANT is a fully automatic method, its performance can be objectively tested also in non-blind mode. The aim of the present study is to examine the sensitivity of the performance of ACMANT to its parameterisation, because low sensitivity could confirm the leading position of ACMANT, while high sensitivity would indicate the uncertainty in the rank order of method efficiencies. The efficiency tests were made with ensemble moving parameter experiments, which is a useful tool for testing automatic and semi-automatic methods (McCarthy et al., 2008; Tichner et al., 2009; Williams et al., 2012). The methodology that we present allows the fast and easy evaluation of the results, therefore we recommend its use also to the examination of other automatic algorithms. The paper allows insight also into the theoretical aspects of the latest developments and the performance of the newest version of ACMANT.

The organisation of the paper is as follows. In Sect. 2, the latest development of ACMANT is presented. In Sect. 3, the methodology and the results of the moving parameter experiment are presented. Finally, in Sect. 4, the results and the further tasks ahead the method developers are discussed and the conclusions are summarised.

## 2 ACMANT

The full description of the first version of ACMANT (Applied Caussinus-Mestre Algorithm for homogenising Networks of Temperature series) has already been published (D2011). That version was referred as ACMANT late by Venema et al. (2012) and ACMANTv1 in web. However, the method has been developed since then. Here, the structure of ACMANT will be briefly described first, then the recent changes will be presented in details.

### 2.1 Structure of ACMANT

The ACMANT was developed on the basis of detection and correction algorithms in PRODIGE. PRODIGE was selected, because its inhomogeneity detection part and correction method (Caussinus and Mestre, 2004) have been turned out to be highly effective in comparison with many other methods (Domonkos, 2011b; Domonkos et al., 2013). However, ACMANT differs from PRODIGE in many details, first of all ACMANT uses reference series instead of pairwise comparisons and ACMANT is fully automatic.

In ACMANT reference series (i.e. series with which the so-called candidate series is compared to find the inhomogeneities in the series of spatial differences) are pre-homogenised before the main detection part. During pre-homogenisation, the use of the candidate series of the Main Detection is excluded from the calculation of adjustment-terms, thus double use of the same spatial connection is not allowed in ACMANT. Both in the pre-homogenisation and in the Main Detection, the optimal step function fitting with the Caussinus–Lyazrhi criterion is applied (Caussinus and Mestre, 2004). However, a novelty of ACMANT is that change-points are searched with bivariate tests, namely joint statistics of the annual means and summer–winter differences are examined (D2011). Correction-terms are always calculated with ANOVA (Caussinus and Mestre, 2004). After the main detection, further inhomogeneities are searched on a monthly scale. Outlier-filtering and filling of gaps of time series are applied three times in ACMANT: first, before pre-homogenisation, second, after pre-homogenisation and third, together with the final adjustments; see further details in D2011.

### 2.2 Latest development of ACMANT

Two kinds of development have been applied to ACMANT since its publication (D2011). One of these changes is that

in the new ACMANT, ANOVA is applied also in the pre-homogenisation, while the other is the introduction of filtering of outlier periods. Together with these changes, the Secondary Detection is also modified.

#### 2.2.1 Calculation of adjustment-terms

In the new ACMANT, ANOVA is always applied and thus unified relative time series, as well as all that were written in Sects. 3.3.4 and 3.6.2. of D2011 are no longer included in ACMANT.

ANOVA is applied in three modes in ACMANT: (a) In the pre-homogenisation it is applied on annual values at an outlier filtering step, (b) In the pre-homogenisation it is applied on annual values and with the exclusion of the further candidate series for preparing reference series to the Main Detection and (c) After the Main Detection and after the Secondary Detection it is applied on monthly data. See the description of ANOVA model in Caussinus and Mestre (2004).

#### 2.2.2 Filtering of outlier-periods

Outlier-periods could be also referred to as short-term inhomogeneities, since their model is a short-term platform-like bias from the correct values. In such platforms the bias is constant for the outlier-period. In the real world the magnitude of the bias could vary within the period, so the platform is only a model. Note, however, that Domonkos (2011b) reported that observed temperature series of Hungary can be modelled well with the inclusion of a large number of short-term, platform-shaped inhomogeneities. Note also that short-term inhomogeneities can be detected only when the size of the bias is large, since with the shortening of the duration of biases the signal to noise ratio worsens. So that, the aim of the filtering of outlier-periods is to filter the short-term, large biases from time series. Both the detection and the correction of outlier-periods are more similar to the common outlier-filtering than to the inhomogeneity detection.

In ACMANT, filtering of outlier-periods is applied for 2–30 month long periods always after the common outlier filtering. The maximum length of such periods is controlled by a parameter ( $c1$ , Table 1), its value was between 18 and 30 in the latest examinations. The values of the detected outlier-periods are treated in the same way as data gaps or outliers, and the interpolation of Sect. 3.2 in D2011 is applied to them. However, for outlier-periods reaching a given threshold ( $c2$ ), the starting and ending dates of outlier-periods are considered as change-points in the final calculations of correction-terms by ANOVA.

In searching outlier-periods, relative time series are used on monthly scale. First the values are standardised, i.e. they converted to have 0 mean and 1 standard deviation. Note that seasonal cycles had been filtered earlier (cf. Eq. 4 in D2011). The standardised relative time series are denoted with  $\mathbf{E} = [e_1, e_2, \dots, e_{nm}]$ , the length of series with  $nm$  and

**Table 1.** Moving parameters and their values in the experiments. Six values are allowed for each parameter. In column “basic” the values in ACMANTv1 (D2011) are shown.

		basic	1	2	3	4	5	6
c1	outlier-p.: max. effective window (month)	–	18	20	22	24	27	30
c2	distance between change-points (month)	–	3	4	5	6	8	10
c3	significance threshold in outlier p. filtering	–	16.0	20.0	25.0	32.0	42.0	60.0
c4	outlier period: power of duration	–	0.5	0.6	0.7	0.8	0.9	1.0
c5	outlier p.: length of outer periods (month)	–	24	24	24	36	36	36
c6	outlier period: correction for seasonality	–	0.10	0.15	0.20	0.30	0.40	0.50
c7	outlier period: precision of timing (month)	–	6	8	10	12	15	18
c8	summer–winter difference	0.50	0.25	0.33	0.42	0.50	0.59	0.67
c9	pre-homogenisation, penalty-coefficient	1.0	1.0	1.2	1.4	1.6	1.8	2.0
c10	Secondary Detection, penalty coefficient	1.0	1.0	1.25	1.5	1.8	2.3	3.0
c11	overlap of relative time series (year)	9	5	7	8	9	10	12
c12	window width in Secondary Det. (month)	60	36	42	50	60	70	80
c13	MA5 threshold	2.0	1.6	1.8	2.0	2.2	2.4	2.6
c14	MA10 threshold	1.4	1.2	1.3	1.4	1.5	1.65	1.8
c15	min. length for harmonic functions (month)	10	8	9	10	12	14	16
c16	Main Detection, monthly precision (month)	25	17	21	25	29	33	37
c17	$r$ -threshold for 2 reference composites	0.5	0.7	0.8	0.8	0.85	1.0	1.0
c18	$r$ -threshold for 3 reference composites	0.5	0.5	0.6	0.7	0.7	0.7	0.8
c19	$r$ -threshold for 4 reference composites	0.5	0.4	0.48	0.6	0.6	0.5	0.6
c20	$r$ -threshold for 5 reference composites	0.5	0.4	0.4	0.5	0.5	0.4	0.5

section means with upper stroke. Further denotations:  $l$  – length of outlier period,  $i$  and  $j$  – starting and ending months (respectively) of the outlier-period in the first estimation,  $m_1$  – number of summer months (of June, July or August) in the outlier-period,  $m_2$  – number of winter months (of January, February or December) in the outlier-period,  $\text{int}$  – integer part,  $\text{sgn}$  – sign of expression,  $\text{mod}$  – module 12,  $c3$ ,  $c4$ , ... are parameters.

Only one outlier-period is identified and selected in a particular step, i.e. the one with the most significant statistic ( $c3$ ). Then its values are adjusted (temporal adjustments, which are valid only in the section of filtering the outlier-periods), and further outlier-periods are searched as long as at least one can be found with significant statistic. The identification of an outlier-period comprises two phases. In the first phase (A), the most significant outlier-period of the time series is selected and a first estimate is made for its position. In the second phase (B) the starting and ending months of the outlier-period are determined.

Phase (A): the outlier-period with the maximal  $c3$  is searched for each  $i, j$  pairs ( $1 \leq j - i < 30$ ) of standardised relative time series. If more than one relative time series are available for a given  $i, j$  pair, always the one with the highest sum of squared correlations of the reference composites is selected for the examination.

$$c3 = l'^{c4} d^2 \quad (1)$$

where  $d$  (magnitude-characteristic) and  $l'$  (duration-characteristic) are determined by Eqs. (2) and (3):

$$d = \overline{e_{i-c5}, e_{i-1}} + \overline{e_{j+1}, e_{j+c5}} - \overline{e_i, e_j} \quad (2)$$

$$l' = \text{int}(\max\{l - c6|m_1 - m_2|, 1 + (1 - c6)(l - 1)\}) \quad (3)$$

Further conditions are that

$$\text{sgn}(\overline{e_i, e_j} - \overline{e_{i-c5}, e_{i-1}}) = \text{sgn}(\overline{e_i, e_j} - \overline{e_{j+1}, e_{j+c5}}) \quad (4)$$

$$c5 \geq 24 \quad (5)$$

$$\text{mod}(c5) = 0 \quad (6)$$

Phase (B): the first and last months of the outlier-period are re-estimated with fitting optimal step-function in window  $[e_{i-c5}, e_{j+c5}]$ . This procedure is made in the same way as in the Secondary Detection (see Sect. 3.5.2. in D2011) with the exception that solutions with exactly two change-points are accepted only, and the first and second change-points are expected in the periods  $[i - c7, i - 1]$  and  $[j, j + c7 - 1]$ , respectively. So that, the final duration of an outlier-period is equal or greater than the pre-estimated duration. If the resultant duration is longer than  $c1$ , then the temporary adjustment is applied for allowing the search of further outlier-periods, but otherwise such outlier periods are left out of consideration.



## Notes:

- a. Outlier-periods can be detected also in the ends of time series, and slightly different rules are applied for them. At the ends of the series Eq. (4) cannot be expected,  $c5 \geq 36$ , in phase B the window around the pre-estimated position is not symmetric and exactly 1 change-point is searched in that phase.
- b.  $c5$  has a minimum threshold for providing a relatively large sample-size of the surrounding values in the evaluation of the deviation of a pre-assumed outlier-period. Naturally, too great values for  $c5$  are not advisable, either.
- c. Equation (6), as well as the consideration of  $m_1$  and  $m_2$  are necessary due to the seasonal behaviour of inhomogeneity-caused biases. Otherwise a long-term inhomogeneity with large bias in the seasonal cycle could be detected as short-term outlier-period when a seasonal peak of the long-term bias and random noise accidentally add up.
- d. The role of  $c1$  is to separate which biases are considered short-term outlier-periods and which are considered long enough to be treated by the main detection part of ACMANT.
- e. In Eq. (3) the coefficient of  $(l-1)$  is  $1-c6$ . It is an arbitrary choice, the coefficient could be independent from  $c6$ .

### 2.3 Modification in the Secondary Detection

Parameter  $c2$  (see Sect. 2.2) is used also in the Secondary Detection to separate outlier-periods from more persistent inhomogeneities. If the period between two adjacent change-points is shorter than  $c2$ , then that period is treated in the same way as outlier-periods detected by Eqs. (1)–(6).

## 3 Moving parameter experiments

Moving parameter experiments have been done to examine the sensitivity of the performance of ACMANT to changes in its parameterisation, as well as to find optimal parameter values if this sensitivity is significant. An objective was to vary each of the arbitrarily set parameters of ACMANT that might influence its performance.

### 3.1 Methods

#### 3.1.1 Moving parameter experiments

Twenty parameters of ACMANT were varied. Seven from the twenty are defined in Sect. 2.2. of this study, while we refer to D2011 presenting the definition of the others:  $c8$  is the coefficient of summer–winter difference ( $c_0^2$  in Eq. 25).

$c9$  and  $c10$  are coefficients for the penalty term of the Caussinus–Lyazrhi criterion in the Pre-homogenisation and the Secondary Detection, respectively (Eq. 26).  $c11$  is the length of overlap when more than one relative time series are used (Eq. 28).  $c12$  is the window width in the Secondary Detection (Sect. 3.5).  $c13$  ( $c14$ ) is the threshold for accumulated anomaly MA5 (MA10) (Sect. 3.5.1).  $c15$  is the minimum duration for fitting harmonic functions in the Secondary Detection (Sect. 3.5.2), as well as in the new version, also in determining the position of outlier period.  $c16$  is the length of the period in which the timing of change-point is expected in the monthly precision section (Sect. 4, Part III, point 4).  $c17$ ... $c20$  are minimum thresholds for the spatial correlation of increment-series ( $r$ ). In the previous versions of ACMANT two reference components with  $r \geq 0.5$  was satisfactory to derive relative time series. In the present experiments various thresholds of  $r$  are applied according to the number of reference composites.  $c17$ ,  $c18$ ,  $c19$  and  $c20$  are thresholds for the cases of 2, 3, 4 and 5 reference composites, respectively.  $c17$ ,  $c18$ ,  $c19$  and  $c20$  indicate four conditions for the reference composites, and reference series was built when at least 1 condition of them was met.

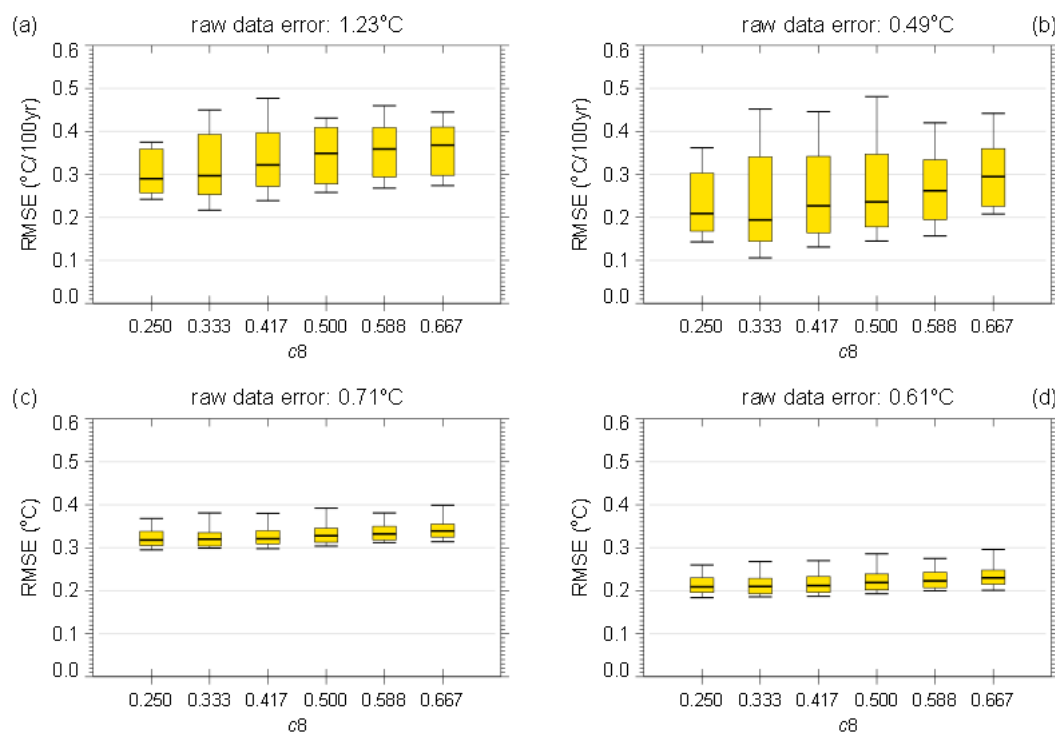
Two thousand experiments were made, each with the 20 networks of the surrogated temperature section of HOME benchmark. All the listed parameters were varied randomly, but only six values were allowed for each parameter (except for  $c5$ , for that parameter only 2 values seemed applicable). Hereafter we refer to the six allowed values as value “1”, value “2”, etc. The restriction to six values allows us to evaluate the impact of any of the twenty parameters with the comparison of six sub-samples, in which the examined parameter is constant.

The values of  $c17$ ,  $c18$ ,  $c19$  and  $c20$  were mutually dependent in the experiment in a way that if  $c17$  had value “ $a$ ”,  $c18$ ,  $c19$  and  $c20$  also had the values “ $a$ ”, so in reality only six sets of values were allowed for them. Due to this dependence, these parameters and their effects were jointly examined.

Table 1 presents the 20 parameters each together with a key-word referring to its role in ACMANT, with its value in ACMANTv1 (D2011) and with its values in the moving parameter experiments. It can be seen that each parameter is varied in a rather wide range.

#### 3.1.2 Efficiency measures

We apply four efficiency measures: (i) root mean squared error (RMSE) of monthly values ( $^{\circ}\text{C}$ ), (ii) RMSE of annual values ( $^{\circ}\text{C}$ ), (iii) RMSE of individual trend-slope biases ( $^{\circ}\text{C}/100\text{yr}$ ) and (iv) RMSE of network-mean trend-slope biases ( $^{\circ}\text{C}/100\text{yr}$ ). All these measures characterise the residual errors in homogenised series, the first three are for the entire period of time series, while type (iv) is for period 1925–1999. For visualising the improvement in the data accuracy due to the use of ACMANT, raw data errors, and the residual errors after the blind homogenisation of HOME with various



**Figure 1.** The sensitivity of residual errors to parameter  $c8$ . (a) Trend bias of station series, (b) trend bias of network-mean series, (c) RMSE of monthly values, (d) RMSE of annual values.

methods (Venema et al., 2012) are also presented in some figures. The involved homogenisation methods of the HOME experiments are PRODIGE, MASH, USHCN and C3-SNHT (Venema et al., 2012), since apart from ACMANT, they produced the highest efficiency for the entire surrogated dataset.

In HOME, only 15 networks from the 20 were evaluated formerly, and for monthly and annual residual errors Centred RMSE (CRMSE) were used (Venema et al., 2012). These differences did not result in visible deviations of the efficiencies presented in our study, relative to the HOME results.

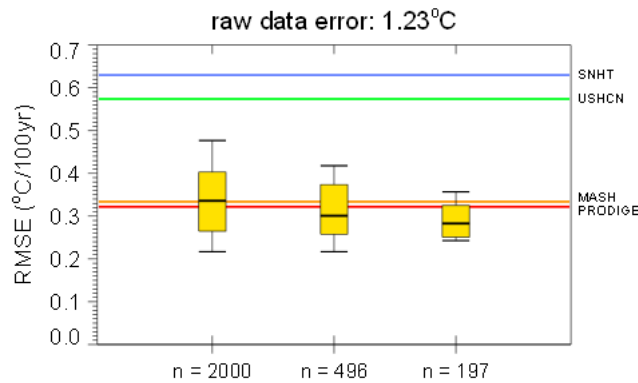
### 3.2 Results

The moving parameter experiments proved the generally low sensitivity of ACMANT to its parameterisation. For most parameters, there are no significant differences of performance with changes of the value within the range examined. For the other parameters slight but significant declines of performance can be observed in one or both tails of the examined parameter ranges. The relatively highest sensitivity of performance is to  $c8$  (Fig. 1), and for this parameter the basic value (ACMANTv1, Table 1) is clearly suboptimal. However, even in this case, the sensitivity is still moderate.

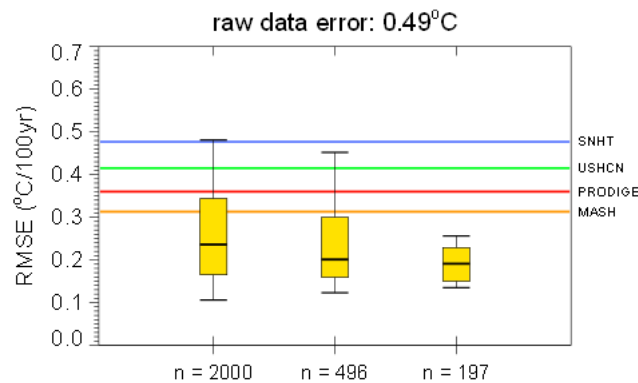
ACMANT has a generally high efficiency in homogenising temperature series and the moving parameter experiments described allows us to characterise the stability of the performance of ACMANT. Figures 2–5 present the results in

comparison with the raw data errors and with the remaining errors after the HOME experiments with four homogenisation methods. Each of Figs. 2–5 shows the results of ACMANT in three columns. The first column shows the results of all the 2000 experiments, thus it includes also the experiments with suboptimal parameter values. For the second column, experiments with seven suboptimal parameter values are excluded, but each basic parameter value (Table 1) is remained in the accepted parameter ranges. For the last column, further four parameter values are excluded and two of them are basic parameter values in ACMANTv1. As at least one of the excluded parameter values frequently occurred, only 496 (197) experiments remained suitable for column 2 (column 3) of Figs. 2–5. The excluded parameter values from columns 2 and 3 (only column 3) are as follows.  $c3$ : value “1”,  $c4$ : “6” (“5”),  $c8$ : “5”, “6” (“4”),  $c9$ : “5”, “6” (“1”),  $c12$ : “1”, ( $c17$ – $c20$ : “6”). Figures 2–5 show that ACMANT generally has the highest performance among the examined homogenisation methods. The advantage of ACMANT is even more enhanced when some suboptimal parameter values are excluded.

The advantage of ACMANT is not the same in the examined efficiency measures. In the residual errors of trends of station data (Fig. 2), PRODIGE and MASH have similar performance to that of ACMANT, although in the right column 95 % of the results with ACMANT are better than those with any other method. In estimating network-mean trends



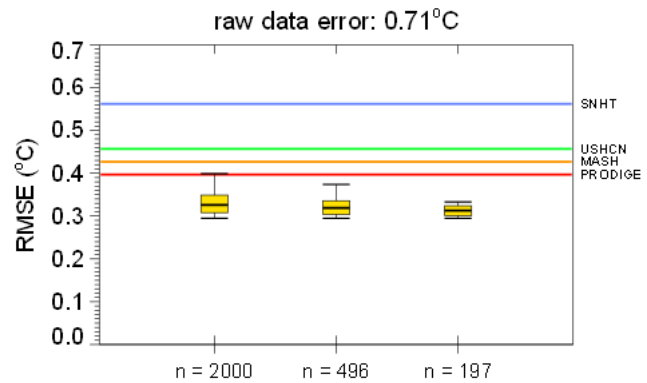
**Figure 2.** Boxplot of the residual errors in trend biases of station series after homogenising with ACMANT the HOME surrogated temperature data. Left column is for all the 2000 experiments, while for the middle and right columns some suboptimal parameters are excluded. Boxes include the values between 5 and 95 percentiles. Raw data errors and residual errors after blind homogenisation with various methods under HOME are also shown.



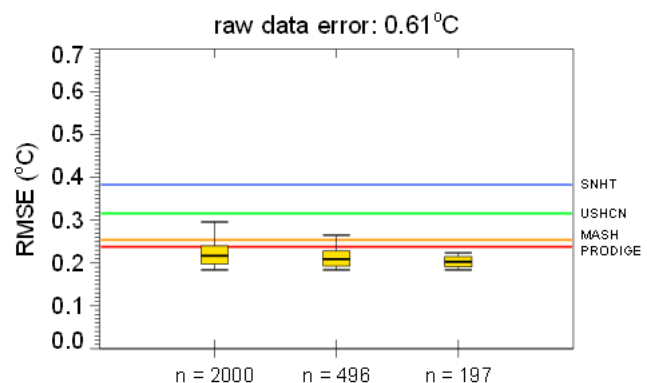
**Figure 3.** The same as Fig. 2, but for trend biases in network mean series.

(Fig. 3) the lead of ACMANT is even more dominant than in estimating trends of station data. On the other hand, the scattering of the results is great (except in the right column), and in the least successful experiments with ACMANT the residual error is as large as in the raw data.

The most important novelty of ACMANT is the harmonisation of the work between monthly and annual scales. As a consequence, the residual monthly RMSE (Fig. 4) is always the best with ACMANT, more precisely, the worst result among the 2000 experiments with ACMANT equals to the result of the second best method (PRODIGE). The residual annual RMSE (Fig. 5) has similar features than the previous efficiency measures, i.e. the results with ACMANT are the best except for a few experiments with suboptimal parameter values in ACMANT.



**Figure 4.** The same as Fig. 2, but for the RMSE of monthly values.



**Figure 5.** The same as Fig. 2, but for the RMSE of annual values.

#### 4 Discussion and conclusions

The results of the moving parameter experiments prove that the good performance of ACMANT late reported in Venema et al. (2012) is not for the overfitting of its parameters, since the performance has low sensitivity to the parameterisation. The high performance is a stable characteristic of ACMANT and it is a consequence of its good methodological properties.

The results indicate that after excluding some suboptimal parameter values, ACMANT always performs better than any other homogenisation methods in homogenising monthly temperatures. However, this result must be treated with some reservations, since (i) Twenty networks is a relatively small sample, because the efficiency of homogenising station series is strongly interdependent within networks; (ii) The HOME benchmark dataset captures well some characteristics of observed temperature data, but it does not mean that the frequency and magnitude distribution of inhomogeneities and some other characteristics are the same in each observed temperature dataset; (iii) Not all the known homogenisation methods are tested in HOME and even the tested methods could produce better results in repeated examinations either due to their new developments, or accidentally. Note here that relying on the experiences of HOME,



the new homogenisation method HOMER has been developed (Mestre et al., 2013). Moreover, a new, automatic version of MASH has been reported, as well as Climatol and RHTest also have newly developed versions (www.climatol.eu/DARE).

For the reasons listed above, further tests are needed to confirm or deny the leading position of ACMANT. However, some other open questions seem to be even more important than the rank order among the best homogenisation methods. The efficiency of any homogenisation method strongly depends on the number and spatial correlations of time series in networks and on the frequency and other characteristics of inhomogeneities as well. When one or more of these characteristics are unfavourable, homogenisation could even worsen the data quality. For instance, it happened to the monthly precipitation data in the HOME experiments. The lesson of that failure is not that monthly precipitation data cannot be homogenised, but it is that the successful homogenisation has also some conditions others than applying good statistical methods. Our present knowledge is yet limited in this field, what means that at present we cannot quantify the necessary conditions. Similarly, we cannot separate confidentially the tasks for automatic homogenisation from the ones for manual homogenisation with intensive metadata use. We have the general knowledge that while huge datasets can hardly be homogenised manually, in case of small number of time series the human control is always advisable. However, this qualitative knowledge is not always sufficient to make optimal decisions in the selection of methods, therefore further tests are necessary to widen our knowledge in this field. Tests with automatic homogenisation methods can be executed relatively easily and their results are partly applicable also to non-automatic methods. Domonkos (2013) describes the objectives and the expected benefits of various kinds of efficiency tests, while Guijarro (2012) shows examples of comparative efficiency tests for various automatic homogenisation methods.

Our main conclusions are as follows:

- The performance of ACMANT has low sensitivity to the parameterisation of the method.
- ACMANT is one of the best methods for homogenising monthly temperatures, and it is likely the best method for homogenising large and spatially dense networks of temperature data.
- Although ACMANT is only for monthly temperature data, its development will help in the future to improve the efficiency of homogenisation also in other climatic variables and in other time resolutions.
- Further efficiency tests are needed to quantify the connections between dataset properties and method performance. First the automatic methods (such as e.g. ACMANT) must be tested, and their results will be partly applicable also for non-automatic methods.

**Acknowledgements.** The research was supported by the EURO4M FP7-SPACE-2009-1/242093.

Edited by: M. Brunet-India

Reviewed by: two anonymous referees

## 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.
- 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, 2005.
- Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, *J. Roy. Stat. Soc., C53*, 405–425, doi:10.1007/s11200-009-0015-9, 2004.
- Craddock, J. M.: Methods of comparing annual rainfall records for climatic purposes, *Weather*, 34, 332–346, 1979.
- Domonkos, P.: Adapted Caussinus-Mestre Algorithm for Networks of Temperature series (ACMANT), *Int. J. Geosci.*, 2, 293–309, doi:10.4236/ijg.2011.23032, 2011a.
- Domonkos, P.: Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods, *Theor. Appl. Climatol.*, 105, 455–467, doi:10.1007/s00704-011-0399-7, 2011b.
- Domonkos, P.: Measuring performances of homogenisation methods, *Időjárás*, 117, 91–112, in press, 2013.
- Domonkos, P. and Štěpánek, P.: Statistical characteristics of detectable inhomogeneities in observed meteorological time series, *Stud. Geophys. Geod.*, 53, 239–260, doi:10.1111/j.1467-9876.2004.05155.x, 2009.
- Domonkos, P., Venema, V., Auer, I., Mestre, O., and Brunetti, M.: The historical pathway towards more accurate homogenisation, *Adv. Sci. Res.*, 8, 45–52, doi:10.5194/asr-8-45-2012, 2012.
- Guijarro, J. A.: After COST/ES0601: Automatic comparison of series homogenization methods, 8<sup>o</sup> Congreso de Asociación Española de Climatología, 25–28 September 2012, Salamanca, 2012.
- 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.: Homogenization of temperature series via pairwise comparisons, *J. Climate*, 22, 1700–1717, doi:10.1175/2008JCLI2263.1, 2009.
- Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertaenik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER: HOMogenisation softwarE in R – methods and applications, *Időjárás*, 117, 47–67, in press, 2013.
- 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 reassessing tropospheric temperature trends from radiosondes using realistic validation experiments, *J. Climate*, 22, 465–485, doi:10.1175/2008JCLI2419.1, 2009.
- Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnick, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquafredda, 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.
- 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, D5, doi:10.1029/2011JD016761, 2012.