ACMANT: WHY IS IT EFFICIENT?

Peter Domonkos
Center for Climate Change, Univ. Rovira i Virgili, Campus Terres de l'Ebre, Tortosa, Spain, e-mail: peter.domonkos@urv.cat

Abstract

In the COST ES0601 project (COST HOME) a benchmark temperature and precipitation dataset (Benchmark) was developed for assessing the efficiencies of homogenisation methods via blind test experiments. In homogenising temperature data, the best performance was achieved with ACMANT. However, ACMANT was developed with the use of the Benchmark, thus its observed performance is not fully comparable with the blind tests results of other methods. Consequently, our knowledge about the true performance of ACMANT is limited. This study includes a brief analysis of the theoretical properties of ACMANT, as well as presents new experimental results. The aim of the study is to provide more information for the objective evaluation of ACMANT.

1. INTRODUCTION

ACMANT (Adapted Caussinus-Mestre Algorithm for homogenising Networks of Temperature series) is a recently developed homogenisation method. It is applicable for monthly temperature series. In the blind test experiments of COST HOME an early version of ACMANT was tested. The early version produced outstanding results regarding the RMSE of monthly values, but it was rather poor in reconstructing true climatic trends (Venema et al. 2012). Later the ACMANTv1.2 (http://www.c3.urv.cat/members/pdomonkos.html) was developed, it is referred as ACMANT late in Venema et al. (2012) and referred as ACMANT in this study.

The efficiency results that are obtained with homogenising Benchmark could be affected by the fact that Benchmark was used in the development of the new method. The treatment of seasonal changes in inhomogeneity caused biases (hereafter: station effects) is particularly criticised from the point of view that the properties of true observed data might considerably differ from the model that applied both in Benchmark and by ACMANT. Unfortunately, at present there is no opportunity to perform new blind tests for ACMANT that are comparative with other homogenisation methods. In this paper the arguments and evidences are intended to be collected in order to obtain the evaluation of ACMANT as objective as possible. The organisation of the paper deviates somewhat from the traditional form. In the next section the theoretical properties of ACMANT will be discussed, then a distinct section will be devoted to analyse the seasonal changes of station effects in temperature series. Studies about real data homogenisation, as well as a test experiment with data of uni-seasonal station effects are discussed there. In section 4 more test results with ACMANT will be presented and analysed, while the last (fifth) section is for synthesising the results and drawing the conclusions.

2. DESCRIPTION OF ACMANT

2.1. General characterisation

In this section a brief description of ACMANT is provided, the full description can be found in Domonkos (2011a). ACMANT is a fully automatic homogenisation method what means that after inputting the raw data and some characteristics of the network (number of stations, length of time series, etc.) the execution does not need human assistance as far as the final homogenised data are produced by the software.

The most important characteristics of ACMANT are a) The harmonisation of examinations in different time-scales (i.e. in annual and monthly scales), b) The use of the optimal segmentation and Caussinus-Lyazrhi criterion in the detection of inhomogeneities, c) The use of ANOVA for the final corrections of inhomogeneities.

ACMANT has four main parts.

I) Preparation. This part contains initial calculations (anomalies, spatial correlations, etc.), outlier filtering, as well as filling the data gaps caused by missing data and outliers.

II) Pre-homogenisation. The purpose of pre-homogenisation is to filter the largest errors from the reference-composites of the final homogenisation. In the pre-homogenisation temporary adjustments are applied to reduce the station effects. Then outlier-filtering and interpolations are performed again using the improved data.

III) Homogenisation. First the long-term biases are searched in annual scale (Main Detection), then with further calculations the timings of the change-points are determined in monthly scale. The remaining station effects are checked on monthly scale with Secondary Detection. Once the detection is finished, ANOVA is applied for calculating the correction terms. ANOVA also provides the final calculations for filling the gaps caused by missing data and outliers.

IV) Final adjustments. In this step change-points with insignificant shift-sizes are excluded from the list of change-points, and ANOVA is applied again with the reduced set of change-points.

2.2. Selected segments of ACMANT

a) Building reference series from composites.
ACMANT performs relative homogenisation. It means that before homogenisation relative time series are derived from the original series. A relative time series (**T**) is the arithmetic difference between the candidate series (**A**) and a reference series (**F**). The traditional way of creating reference series (Peterson and Easterling, 1994) is applied with a specific parameterisation in ACMANT. For a candidate series (g) all the other series in network ($s = 1 \ldots S$) are used as reference composites when the spatial correlation (r) with g exceeds a preset threshold. More precisely, r stands for the spatial correlation between the first difference (increment) series. The reference-composites are weighted according to the squared correlations (r^2).

$$\mathbf{F}_{\mathbf{g}[\mathbf{j_1},\mathbf{j_2}]} = \frac{\sum_{s=1}^{S} r_{g,s}^2 \mathbf{A}_{\mathbf{s}[\mathbf{j_1},\mathbf{j_2}]}}{\sum_{s=1}^{S} r_{g,s}^2} \tag{1}$$

In eq. 1 [j1,j2] represents a section between years $j1$ and $j2$, i.e. eq. 1 can be applied to any period of the time series. In ACMANT the threshold r is 0.4, but for at least two composites $r \geq 0.5$ is expected. Note that these thresholds are low relative to some recommendations (Alexandersson and Moberg, 1997, Auer et al. 2005, etc.).

b) Creation of multiple relative time series
ACMANT uses multiple relative time series, because the number of available reference composites often varies according to the sections of the time series. The minimum length of **T** series is 30 years. Three **T** series are always constructed: (i) with the highest possible Σr^2, (ii) with the earliest starting year, (iii) with the latest ending year. However, as criterions (i), (ii) and (iii) might be satisfied by the same series, the true number of **T** series can be lower than 3. On the other hand, more than three **T** series can be involved according to the changes in Σr^2 for different sections of the candidate series (Domonkos, 2011a).

c) Main Detection:
Detection of inhomogeneities is always performed one-by-one for different candidate series in ACMANT. Optimal step-function is fitted to two annual variables, namely to the annual mean (**TM**) and amplitude of seasonal cycle (**TD**), and the common change-points of them are searched. The minimum distance between two change-points is 3 years. In other respects Main Detection is the same as the detection of PRODIGE (Caussinus and Mestre, 2004).

Eqs. 2 and 3 show the calculation of tm and td for year j.

$$tm_j = \frac{\sum_{m=1}^{12} t_{j,m}}{12}, \quad td_j = \frac{t_{j,5} + t_{j,6} + t_{j,7} + 0.5t_{j,8} - t_{j,11} - t_{j,12} - t_{j,1} - 0.5t_{j,2}}{3.5} \tag{2), (3}$$

Then the optimal segmentation of an L year long period into $K + 1$ segments is given by eq. 4.

$$\min_{[j_1, j_2, \ldots j_K]} \left\{ \sum_{k=0}^{K} \sum_{i=j_k+1}^{j_{k+1}} (tm_i - \overline{\mathbf{TM_k}})^2 + c_0^2 (td_i - \overline{\mathbf{TD_k}})^2 \right\} \tag{4}$$
$$j_0 = 0, \quad j_{K+1} = L$$

Upper stroke denotes the time average for segment k, c_0 is constant, its value in ACMANTv1.2 is 1.414.

The number of segments is optimised by the Caussinus-Lyazrhi criterion (eq. 5, Caussinus and Lyazrhi, 1997).

$$\ln\left\{1-\frac{\sum_{k=0}^{K}(j_{k+1}-j_k)\cdot\left[(\overline{\mathbf{TM_k}}-\overline{\mathbf{TM}})^2+c_0^2(\overline{\mathbf{TD_k}}-\overline{\mathbf{TD}})^2\right]}{\sum_{i=1}^{L}(tm_i-\overline{\mathbf{TM}})^2+c_0^2(td_i-\overline{\mathbf{TD}})^2}\right\}+\frac{2K}{L-1}\ln(L) \qquad (5)$$

d) Precision of the timings of detected IHs:
Main Detection works on relatively long time-scale. After Main Detection more precise timings of change-points are searched applying 48-month symmetric windows around the pre-estimated timings of the change-points. Note that within such a window only 1 change-point was detected, since the minimum length of segments is 3 years in Main Detection. In a window, two-phase harmonic functions are fitted to the values, and the optimum fitting is searched. Phase-change in one of the 25 central months is accepted only. The timing of the phase-change in the optimum fitting is the final timing of the detected change-point.

e) Secondary Detection:
When after the adjustments that applied according to Main Detection results, accumulated anomalies still exceed some predefined thresholds, Secondary Detection is applied.

In Secondary Detection 60-month long sub-series of monthly values around the maximum of accumulated anomalies are examined. At this step the optimal segmentation is applied for the time averages of monthly values, but the section-means are substituted with harmonic functions of annual cycle for sections of minimum 10 months, and the number of segments is maximised by 3 for a sub-series.

f) Correction: ANOVA
In ACMANT, ANOVA is applied for calculating the final correction-terms. This procedure minimises the standard deviation of the homogenised data. It can be proved that ANOVA provides the optimal estimation of correction terms when the following two conditions exist: i) the climate signal is the same in the network, ii) the station effect is constant between two adjacent known change-points (Caussinus and Mestre, 2004).

g) Pre-homogenisation:
In ACMANT each time series are pre-homogenised in a way that in the calculation of the adjustment-terms for series s, series g is excluded from the process, when s is prepared to be a reference-composite in the final segmentation of series g. As s usually takes part in the homogenisation of all the other time series of the network, usually N-1 different pre-homogenisations are performed for an individual s in a network of N stations.

First the order of the candidate series is set, it is from the series of estimated poorest quality to the series of estimated best homogeneity. The determination of the order is based on the estimation of maximal station effects. This step necessarily contains some arbitrary criterions (Domonkos, 2011a).

During the pre-homogenisation ANOVA is not applied, because the repetition of ANOVA would overuse the spatial connections among data. In the pre-homogenisation temporary adjustment terms are applied. These terms are calculated with the help of unified relative time series.

h) Temporary adjustments: Unified relative time series
A unified relative time series has the following properties: i) It covers the whole period for which the homogenisation of the candidate series can be performed, ii) It includes the relative time series unchanged for which Σr^2 is the highest (it is often shorter than the whole period), it is the principal section, iii) The principal section is completed with other relative time series to cover the whole period defined by i), the latter series are complementary series, iv) The complementary series are adjusted before completion in a way that systematic biases due to differences of the spatial means of station effects for differing subsets of stations are intended to be elaborated. With other words, as the regional average of all stations (N) does not equal to the average of some subsets of Q stations ($Q < N$), this source of bias has to be treated in the creation of unified relative time series. In the early version of ACMANT that was tested in the blind test experiments with the Benchmark (Venema et al. 2012), the main source of trend-errors with ACMANT was the lack of harmonisation among relative time series.

2.3. Evaluation of the theoretical properties of ACMANT

I) Factors explaining the high efficiency

- Optimal segmentation with the Caussinus-Lyazrhi criterion. Earlier efficiency tests of detection methods proved that this method performs best among the inhomogeneity-detection methods used in climatology (Domonkos, 2008, 2011b). The joint segmentation (Picard et al., 2011) could have similar or even better performance, but it has not been proved yet with tests.
- Harmonisation of the examinations of different time-scales in the detection process. The signal to noise ratio is higher in low time resolution than in high resolution, but to find the precise timings and the occurrences of short-term biases the examinations in high (monthly) resolution is also necessary. This kind of harmonisation is unique in ACMANT. A particularly valuable novelty is the bivariate detection that yields results in monthly scale in spite of the examinations are made in annual scale. Note that the latter is also criticised sometimes, because the true annual cycles of station effects are unknown.
- The application of ANOVA makes the correction terms to be as accurate as possible. See also Domonkos et al. (2012).
- Pre-homogenisation is applied in a way that multiple use of the same spatial connection is not allowed, thus the accumulation of errors due to the repeated inclusion of a noise-term or a non-revealed inhomogeneity is excluded.
- Outlier filtering and gap filling are repeated using data of higher and higher quality during the homogenisation procedure.

II) Weak points and doubts around ACMANT

- ACMANT was tested with Benchmark, and in Benchmark the station effects have harmonic annual cycles with maximum biases in winter and summer, thus this feature of Benchmark favours to ACMANT. It could be questioned how the Benchmark-model is applicable for real data with respect to the annual cycle of station effects. See its analysis in Sect. 3.
- ACMANT uses some arbitrary parameters. These parameters were set with the help of Benchmark-experiments, thus the performance of ACMANT on datasets of markedly different properties from Benchmark is unclear. See more discussion about it in Sect. 4.2.
- The use of unified relative time series for calculating adjustment-factors may be suboptimal, since the use of homogeneous sections in pairwise comparisons could likely produce

more accurate results. However, it is a challenge to find a good automatic subprogram of pairwise comparisons (although such subprogram has already been created by Menne and Williams, 2009). Note that the unified relative time series are applied only for temporary adjustments, thus the undesired effect to the final efficiency is supposed to be little if any.

- The use of multiple comparison might be better in the detection part than the use of one reference series from composites. It is a question that cannot be decided in a theoretical way, and the results of the Benchmark experiment are also inadequate to clarify this point, since the effect of the chosen method in time series comparison cannot be examined separately from the other characteristics of homogenisation methods.

3. THE SEASONAL CYCLE OF STATION EFFECTS AND ITS IMPACT ON THE PERFORMANCE OF ACMANT

ACMANT presumes harmonic annual cycle of station effects with extremes of biases in mid-summer and mid-winter, therefore the adequacy of Benchmark to this respect has to be evaluated.

In Benchmark the station effects have annual cycles of definitely harmonic shape and modes in winter and summer. The size of the maximum deviation has a standard normal distribution with 0 expected value and 0.4 °C standard deviation. Although the modes are always in winter and summer, the phases still have substantial variation: the modes can be in any month of summer and winter, and they more often occur in the beginning or ending months of seasons than in the middle month (Venema et al. 2012).

In studies of the homogenisation of true observational temperature series more inhomogeneities were reported for summer series than for winter series (Moberg and Alexandersson, 1997; Drogue et al. 2005; Domonkos, 2006, Domonkos and Štěpánek 2009). Moberg and Alexandersson (1997) pointed on the main cause of this seasonal difference: changes in radiation-effects due to technical changes of the temperature observations are larger in summer than in winter. Czech and Hungarian temperature series were homogenized with 16 homogenization methods (Domonkos and Štěpánek 2009), and the results show that both the frequency and the magnitude of station effects are the smallest in winter and the largest in summer. On the other hand, Moberg and Alexandersson (1997) found the largest temperature-shifts in winter (but the highest shift-frequency in summer), Štěpánek reported the largest station effect for September (personal information, 2010), and according to some assessments, the mean magnitude of seasonal cycles is larger in Benchmark than in true observational series (Venema et al. 2012). Note that synoptic climatological factors might cause annual cycle of station effects in other way than with summer and winter modes.

Considering that the detection results provide only estimations of the true properties, and the seasonal cycles of Benchmark have some non-natural irregularity, it is hard to estimate if Benchmark or the true observational data favours more ACMANT regarding to the seasonal cycle of station effects included in them. Anyhow, it is useful to know the performance of ACMANT when there is no seasonal cycle of station effects at all. For this reason a special experiment was made with Benchmark: The monthly anomalies (from the monthly mean of a given series) of a given year were randomly re-ordered. The way of reordering was the same for each station-series and for both the homogeneous and inhomogeneous ("raw") data. The re-ordering was changed randomly year-by-year. In this way a dataset was created in which all the annual values are the same as in Benchmark, but the seasonal cycle of station effects is ceased. The performance of ACMANT ("ACMANTx" in Table 1) was tested in this dataset.

Method	CRMSE(m)	CRMSE(a)	RMSE(t)	RMSE(nt1)	RMSE(nt2)
ACMANTv1.2	0.563	0.728	0.768	0.382	0.544
ACMANTx	0.412	0.662	0.739	0.272	0.634
PRODIGE main	0.416	0.676	0.724	0.042	0.436
PRODIGE trendy	0.413	0.679	0.728	0.031	0.435
PRODIGE monthly	0.431	0.693	0.735	0.093	0.426
MASH main	0.398	0.667	0.706	0.315	0.441
Craddock Vertacnik	0.461	0.724	0.770	0.278	0.403
USHCN 52x	0.382	0.586	0.490	-0.124	0.102

Table 1. Efficiencies in reducing the RMSE / CRMSE errors of the row data applying various homogenisation methods (1 = perfect homogenisation). CRMSE = centered RMSE (Venema et al, 2012), m = monthly, a = annual, t = trend slope, nt1 = network mean trend slope for 1900-1999, nt2 = network mean trend slope 1950-1999, the description of the homogenisation methods is in Venema et al. (2011).

In accordance with the expectations, the efficiencies are lower than with the true Benchmark, but the decline turned out to be not too large. The observed efficiencies are in the range of the efficiencies of PRODIGE and MASH with the true Benchmark, and the estimation of network-mean trend bias is even better. The efficiency for reconstructing the network-mean trends of 1950-1999 has increased relative to the original experiment with ACMANT, but it is likely a random effect due to the small sample size. The highest decline of efficiency occurred with the CRMSE of monthly values (from 0.56 to 0.41) but the 0.41 is still not lower than the efficiency of the other best methods, except Craddock-test. Note that the observed efficiencies of the Craddock-test are based on a partial contribution (i.e. 7 networks were homogenised from the available 15), therefore they are not fully appropriate for making comparisons with the results of full Benchmark experiments.

When comparisons are made between the performance of ACMANT and the performance of other best homogenisation methods in this study, the “other best methods” are composed by PRODIGE main, PRODIGE trendy, PRODIGE monthly and MASH main. Although the group of the best methods is wider, it includes also Craddock-test, USHCN and the newly developed HOMER of the COST HOME team, good test results of full Benchmark experiments are available only for PRODIGE and MASH. There are full contributions also with USHCN, but the USHCN belongs to the best methods for other reason than the observed efficiencies (that are not too high), i.e. the USHCN has a stably low false alarm rate, which is an important positive feature, but out of the analyses of this study.

4. FURTHER TEST EXPERIMENTS WITH ACMANT

4.1. Tests with the “Big Benchmark”

Big Benchmark is another (bigger) test dataset than the official Benchmark, and its creator is also Victor Venema. The original reason of its creation was to test the fully automatic USHCN software with a test dataset that is similar to the true observational temperature dataset in the United States. The statistical properties of Big Benchmark are similar to those of Benchmark with two important exceptions: (i) The Big Benchmark is much bigger than Benchmark, it contains 200 surrogated and 200 synthetic network, (ii) In Big Benchmark networks may contain 9 or 15 time series, but never only 5 time series. Note that the Benchmark surrogated temperature dataset consists of 15 networks, and 9 of them contains only 5 time series. For this reason, unfortunately, test results with Big Benchmark are not directly comparable with the Benchmark result, although they still can be interesting.

ACMANT was subdued to a blind test with the first 100 surrogated networks of Big Benchmark. The efficiency turned out to be higher than with the official Benchmark (Table 2), likely due to the denser networks of Big Benchmark. For making fairer comparisons, characteristics for subsets of fixed network-sizes are also shown. Note however, that the number of networks in the official Benchmark with 9 (15) time series is only 4 (2), therefore the opportunity to draw profound conclusions from these results is very limited. Due to the small sample-size in Benchmark, network-mean errors have not been calculated for subsets of fixed network-sizes.

Test-data	CRMSE(m)	CRMSE(a)	RMSE(t)	RMSE(nt1)	RMSE(nt2)
OB, 9 series / nt	56.5	72.1	74.2		
BB, 9 series / nt	56.2	74.1	71.6		
OB, 15 series /nt	58.4	78.0	84.0		
BB, 15 series /nt	62.2	80.9	85.0		
OB, all networks	56.3	72.8	76.8	38.2	54.4
BB, all networks	60.1	78.5	79.6	40.5	55.5

Table 2. Efficiency characteristics with ACMANT for the official Benchmark (OB) and Big Benchmark (BB). nt = network, denotations in the headline are the same as in Table 1.

The results with the Big Benchmark experiment indicate that the performance of ACMANT is stably high for dense networks of data with high spatial correlations. The observable efficiency characteristics here are higher than those were achieved with the other best homogenisation methods during the Benchmark experiment, but for making correct comparisons with them, further common blind test experiments would be needed.

4.2. ACMANT with changing parameters

The second unit of the Caussinus-Lyazrhi criterion (sect. 2.2, eq. 5) is a penalty-term. Its effect is that higher number of change-points (K) is allowed only when the fitting of the step function to the data becomes substantially better with the increase of K. In the present experiment this penalty-term is supplied with coefficient p (eq. 6), and it is varied between 1 and 7 in the pre-homogenisation part of the procedure. (In the final homogenisation p always equals 1.)

$$\ln\left\{1-\frac{\sum_{k=0}^{K}(j_{k+1}-j_k)\cdot\left[(\overline{\mathbf{TM_k}}-\overline{\mathbf{TM}})^2+c_0^2(\overline{\mathbf{TD_k}}-\overline{\mathbf{TD}})^2\right]}{\sum_{i=1}^{L}(tm_i-\overline{\mathbf{TM}})^2+c_0^2(td_i-\overline{\mathbf{TD}})^2}\right\}+p\frac{2K}{L-1}\ln(L) \qquad (6)$$

The effect of increasing p is that only the most significant inhomogeneities can be filtered out during the pre-homogenisation. As the aim of the pre-homogenisation is just to eliminate the largest biases of the raw data, the application of $p > 1$ in the pre-homogenisation might result in higher efficiency than the basic version. The results of the experiment are shown in Fig. 1. In the figure the mean efficiencies for the other best methods are also shown for making comparisons.

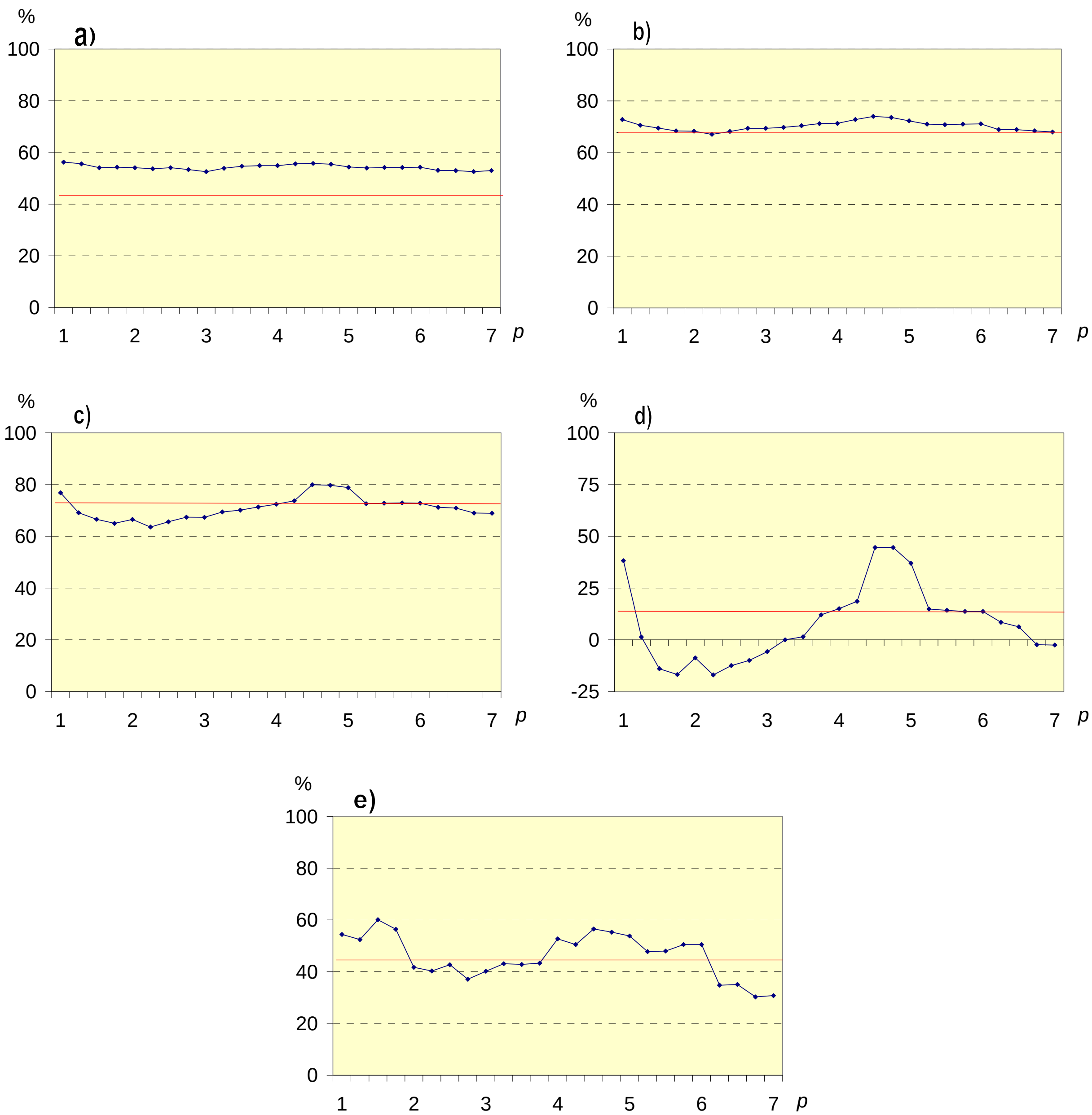

Fig. 1. Efficiency in reducing the RMSE / CRMSE error of raw data in function of *p*. a) CRMSE monthly, b) CRMSE annual, c) RMSE of individual trends, d) RMSE of network-mean trends for 1900-1999, e) RMSE of network-mean trends for 1950-1999. Red horizontal lines represent the mean efficiency of full experiments with PRODIGE and MASH during the Benchmark homogenisation.

Two maximums appear in function of *p*, the first is at *p* of 1 – 1.5, while the other is at *p* of 4.5 – 5.0. I cannot explain this double optimum, even do not know whether it would be similar for other datasets, or it is a peculiarity of Benchmark, more precisely the examined 15 networks of the surrogated temperature dataset of Benchmark. The absolute maximum of observed efficiency is $p = 1$ for the monthly CRMSE, $p = 1.5$ for the network-mean bias for 1950-1999, while for the other efficiency measures the absolute maximums are with $p = 4.5$. The variation of efficiency is small in Fig. 1a and 1b, moderately large in Fig. 1c, while quite large for network-mean biases

(Fig. 1d and 1e). The efficiency of ACMANT is slightly higher than that of PRODIGE and MASH in annual CRMSE, substantially higher than that of PRODIGE and MASH in monthly CRMSE, while for efficiencies in trend-bias reduction this relation is parameter-dependent.

Finally I note that I made another experiment varying another parameter of ACMANT, i.e. the exponent of the denominator in eq. 42 of Domonkos (2011a), but in that case the observed variation of efficiency was much less than in the presented case.

5. DISCUSSION AND CONCLUSIONS

The study of Venema et al (2012) contains several statements about the performances of the methods participated in the Benchmark experiment. Now two of them about the late ACMANT are quoted here, for discussing if we have better understanding after the examinations presented. The two statements are (i) “ACMANT late contribution suggests that ACMANT is currently the most accurate method available”, and (ii) “ACMANT late is optimized based on the benchmark data itself. It is thus not clear how much of this performance would be realised in an application to a real dataset.”

First it has to be made clear that the projection of the observed efficiency results to the application to real data has limitation due to the differences between the surrogated data and real data that obviously exist in spite of the effort has been made to have the surrogate data similar to the real data. However, this limitation is not specific for ACMANT. An exception could be the harmonic annual cycle of station effects, which is exploited more intensively by ACMANT than by any other homogenisation method. However, the analyses of Sect. 3 proved that its effect is minor if any, in raising artificially the performance of ACMANT.

A more serious problem is that 15 networks of data is not very much either to find an optimum parameterisation or to achieve an accurate validation. It is because the within network errors are often interdependent and the distribution of the degree of errors is non-normal, but rather exponential. For illustrating the latter, a brief statistic of the network-mean biases for 1900-1999 in the Big Benchmark experiment is presented here. The absolute value of the bias was below 0.5°C in 94 cases (from the examined 100), in four cases the bias was between 0.51 and 0.63 (°C), but the highest two biases were 0.83°C and 1.20°C. The likely explanation is that rare unfavourable coincidence of change-points in different time series, missing data, as well as unfavourable interference with unusually high noise may result in large homogenisation errors even with such a sophisticated method as ACMANT. This fact limits the opportunity to draw final conclusions from the experiments with the 15 surrogated temperature networks, and Big Benchmark was examined only with ACMANT.

Both the Big Benchmark experiment and the parameterised examination showed that the performance of ACMANT is more stable with respect to the reduction of monthly and annual CRMSE errors than in the reduction of trend-biases. The decrease of monthly CRMSE is spectacularly greater with ACMANT than with any other homogenisation method, and this difference is not sensitive to the chosen set of parameters of ACMANT. The performance of ACMANT in this characteristic may fall onto the level of the other best homogenisation methods only when the annual cycle of biases is entirely removed from the test datasets, which is, however, an unrealistic condition for the observed temperature data of mid- or high geographical latitudes. This good result with ACMANT is a consequence of the sophisticated treatment of different time-scales (i.e. multi-annual, annual and monthly) from which only one piece is the bivariate detection with two annual variables in the Main Detection segment.

The main conclusions are as follows:

- All the examinations confirm that ACMANT belongs to the family of the best homogenisation methods (i. e. PRODIGE, MASH, Craddock-test, HOMER and USHCN).
- ACMANT is particularly effective in reducing the RMSE of monthly temperature data. When dense networks with high spatial correlations are treated, this favourable characteristic of ACMANT is very stable. This characteristic of ACMANT might have importance in the future in providing input data for daily data homogenisation.
- The performance in trend-bias reduction is more parameter-dependent than in the reduction of RMSE errors. The examinations presented do not seem to be sufficient to find the optimal parameterisation of ACMANT, thus further analyses are needed.

Acknowledgements: The research was supported by the European projects COST ES0601 and EURO4M FP7-SPACE-2009-1/242093 and by the Spanish project "Tourism, Politics and Environment" ECO 2010-18158.

REFERENCES

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

Auer, I., Böhm, R. and coauthors, 2005: 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.

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

Caussinus, H. and Mestre, O. 2004: Detection and correction of artificial shifts in climate series. J. Roy. Stat. Soc. Series C 53, 405-425.

Domonkos, P. 2006: Application of objective homogenization methods: Inhomogeneities in time series of temperature and precipitation. Időjárás, 110, 63-87.

Domonkos, P. 2008: Testing of homogenisation methods: purposes, tools and problems of implementation. Proceedings of the 5th Seminar and Quality Control in Climatological Databases. (Ed. Lakatos, M., Szentimrey, T., Bihari, Z. and Szalai, S.),WCDMP-No. 71, WMO/TD-NO. 1493, 126-145.

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

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

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

Domonkos, P., Venema, V. and Mestre, O. 2012: Efficiencies of homogenisation methods: Our present knowledge and its limitation. In: Proceedings of the 7th Seminar and Quality Control in Climatological Databases.

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

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

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

Peterson, T.C. and Easterling, D.R. 1994: Creation of homogeneous composite climatological reference series. Int. J. Climatol. 14, 671-679.

Picard, F., Lebarbier, E., Budinska, E. and Robin, S. 2011: Joint segmentation of multivariate Gaussian processes using mixed linear models, Computational Statistics and Data Analysis, 55, 1160-1170.

Venema, V. K. C., Mestre, O. and the COST HOME team, 2011: Description of the COST-HOME monthly benchmark dataset and the submitted homogenized contributions. Report, Meteorological Institute, University of Bonn, Germany, available at: http://www2.meteo.uni-bonn.de/venema/articles/2011/report home.pdf, 2011.

Venema, V., Mestre, O. and the COST HOME team, 2012: Benchmarking monthly homogenization algorithms. Climate of the Past, 8, 89-115, doi:10.5194/cp-8-89-2012, 2012.