



Comparison of HOMER and ACMANT homogenization methods using a central Pyrenees temperature dataset

N. Pérez-Zanón, J. Sigró, P. Domonkos, and L. Ashcroft

Center for Climate Change (C3), Campus Terres de l'Ebre, Universitat Rovira i Virgili, Tortosa, Spain

Correspondence to: N. Pérez-Zanón (nuria.perez@urv.cat)

Received: 13 January 2015 – Revised: 19 May 2015 – Accepted: 27 May 2015 – Published: 9 June 2015

Abstract. The aim of this research is to compare the results of two modern multiple break point homogenization methods, namely ACMANT and HOMER, over a Pyrenees temperature dataset in order to detect differences between their outputs which can affect future studies. Both methods are applied to a dataset of 44 monthly maximum and minimum temperature series placed around central Pyrenees and covering the 1910–2013 period. The results indicate that the automatic method ACMANT produces credible results. While HOMER detects more breaks supported by metadata, this method is also more dependent on the user skill and thus sensitive to subjective errors.

1 Introduction

The latest report of the Intergovernmental Panel on Climate Change indicates that the Mediterranean region is one of the most vulnerable areas of the Earth to global warming (Barros et al., 2014).

The Pyrenees in southwestern Europe is a particularly valuable mountain range because of its biodiversity and water resources which allow population development through agriculture, hydropower energy and tourism (López-Moreno et al., 2008). Future climate scenarios indicate that a decrease in snow cover in this region is likely during the next century (López-Moreno et al., 2009). Therefore it is essentially important to learn and understand more precisely the climate and climate change of this important area.

For any type of climate analysis, including paleoclimatology studies, which use climate series for calibration of the proxies (Bradley, 1999), the use of a high-quality observed dataset is essential. There are several reasons why inhomogeneities (changes in the meteorological records due to non-climatic factors) occur in observational series, such as stations relocation, changes in the environment around the station, changes in the observing time (Aguilar et al., 2003; Brunet et al., 2008).

To detect and correct inhomogeneities, a large number of methods have been developed (Venema et al., 2012). In the present study, two modern multiple break point ho-

mogenization methods developed during the Action COST-ES0601 (HOME), namely HOMER (HOMogenization software in R, Mestre et al., 2013) and ACMANT (Adapted Caussinus-Mestre Algorithm for Networks of Temperature series, Domonkos, 2011b), are used. We considered it important to select methods that treat multiple occurrences of inhomogeneities with adequate statistical tools, as observed temperature series usually contain 5 or more inhomogeneities per 100 years on average (Domonkos, 2011a; Venema et al., 2012; Willett et al., 2014). While ACMANT is fully automatic, and thus convenient to use for large datasets, the use of HOMER can include the consideration of metadata (document information about the geographical and technical evolution of the observations) because HOMER is an interactive method.

Previous studies in the Pyrenees have developed homogenized series for different spatial and temporal coverage: Bücher and Dessens (1991) examined one station using a bivariate test to detect systematic changes in the mean, while Esteban et al. (2012) used HOMER to homogenize three stations in Andorra. Cuadrat et al. (2013) have recently produced a homogenized dataset for the whole Pyrenees but only for the 1950–2010 period.

For this study, 123 series of either automatic or manual observations were gathered around the Pyrenees covering 1910–2013. Section 2 describes the data in more detail, while

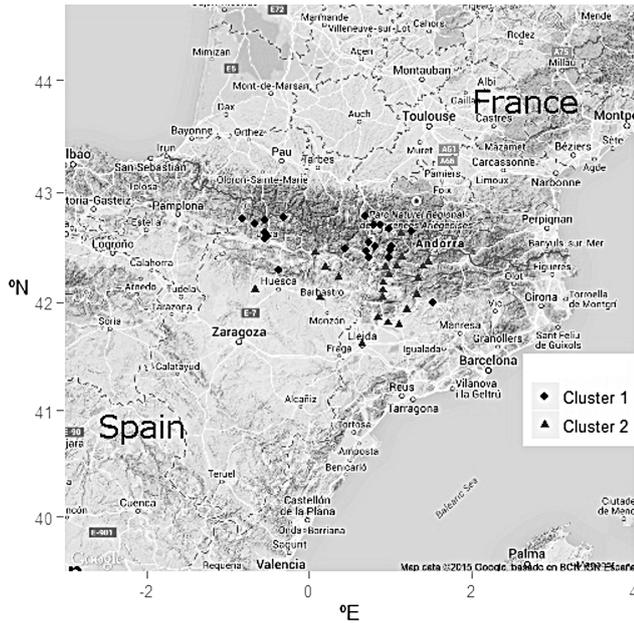


Figure 1. Location of the study area and stations in cluster 1 (black circles) and cluster 2 (grey triangles) using RGoogleMaps package (Kilibarda and Bajat, 2012).

Sect. 3 presents the homogenization methods employed. In Sect. 4, we show the comparison between the two homogenization techniques before drawing conclusions in Sect. 5.

2 Data

The Pyrenees is a mountainous region located in the southwest of Europe and influenced by Atlantic and Mediterranean climate features (e.g. López-Moreno et al., 2008). For this region, 123 series were obtained from the Spanish National Meteorology Agency (AEMET) and the Catalonia Meteorological Service (SMC). The series are of various lengths and cover the period 1910–2013. To develop long-term series, short periods of neighbouring observations were merged and the date of the combination was stored as metadata. The maximum distance accepted in combinations was 7.7 km (in Boí), while the mean difference of altitude was 147 m due to the complexity of the terrain.

Internal consistencies and temporal coherency quality control tests were run on the daily data, following Brunet et al. (2008). More than 834 000 values for each variable (maximum and minimum daily temperature) were examined, with 61 values corrected and 573 set to missing.

Monthly averages of maximum (T_{max}), minimum (T_{min}) and mean temperature (T_{mean}), calculated as the average of T_{max} and T_{min} , were calculated with the missing data tolerance that no more than 7 non-consecutive or 5 consecutive missing daily values for a month were allowed.

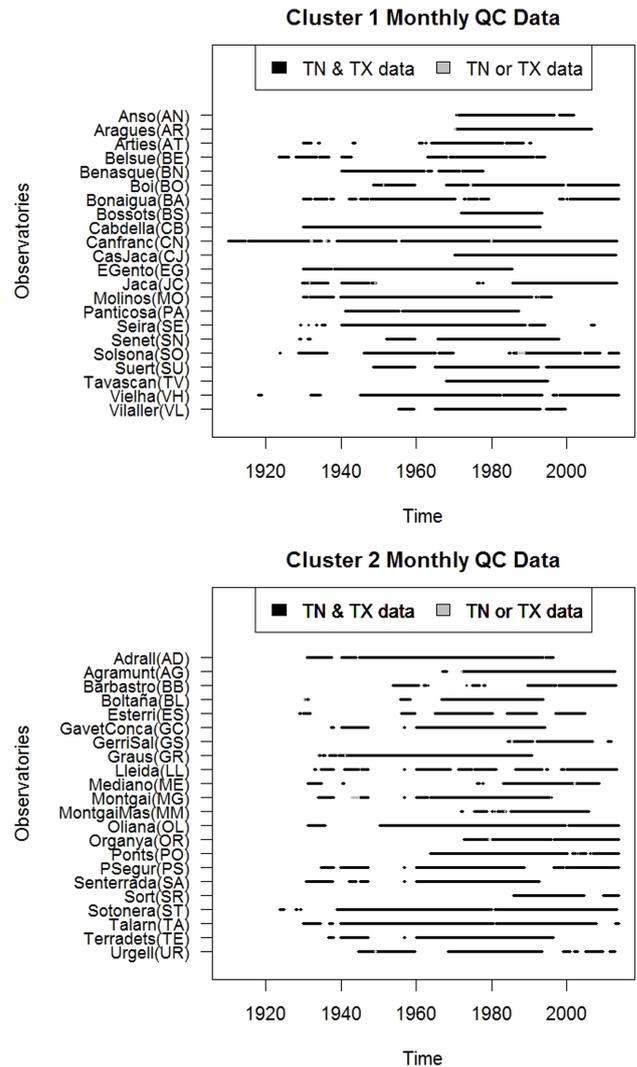


Figure 2. The temporal coverage of monthly maximum (T_X) and minimum (T_N) temperature data for each climatological cluster.

Monthly data must satisfy minimum requirements in terms of the length and completeness to run both HOMER and ACMANT. Finally, 44 T_{min} and 44 T_{max} series are included in the dataset prepared for the homogenization.

Next, the time series were sorted into two climatological clusters, applying a cluster analysis on monthly T_{mean} (see Sect. 3.1). Each cluster contained 22 series. For each cluster, monthly quality control was applied to detect additional outliers using the Fast QC routine included in HOMER. A total of 13 monthly values were removed from Cluster 1 and 34 from Cluster 2. The obtained series will be referred as QC data and the place and period of data coverage for each station is shown in Figs. 1 and 2.

To perform the comparison between the homogenization methods, only the periods homogenized by ACMANT (varies according to series) are taken into account (see

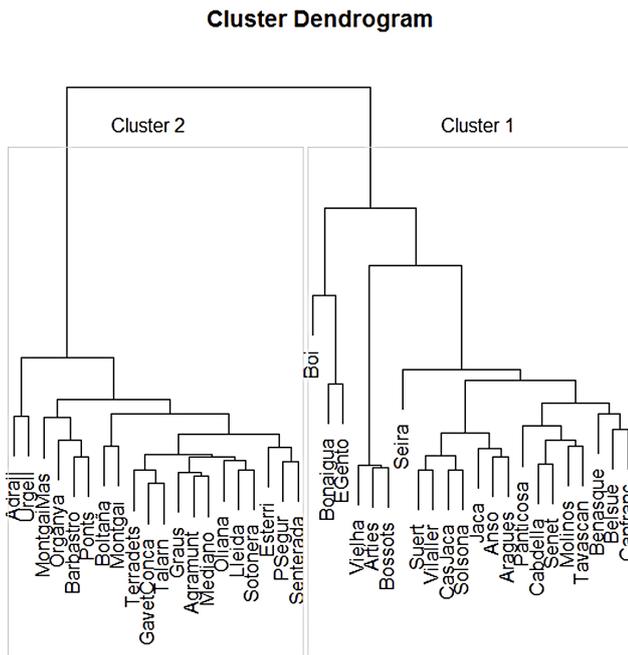


Figure 3. Result of the hierarchical cluster analysis on the first step of computation (see Sect. 3.1).

Sect. 3.3). For these periods, the percentages of monthly missing values ranges between 0.5 and 44 % and 31 series have less than 20 % of missing values.

3 Methodology

3.1 Cluster analysis

As HOMER is interactive software, in order to make the homogenization process more easily manageable, the series were split into two groups. Cluster analysis was applied on T_{mean} data, and not T_{max} or T_{min} , to reduce the likelihood of simultaneous breaks being clustered together. Euclidean distance was calculated between T_{mean} series after normalization (mean = 0; standard deviation = 1). A hierarchical cluster analysis was applied using Ward agglomeration method, where the analysis starts with as many clusters as the number of time series and the clusters are built by adding T_{mean} series of the least distance in each step (Wilks, 2011).

The Pyrenees series cover different sections of the period 1910–2013, and two pairs of series (EGento and Benasque, with Sort and GerriSal) do not have overlapping sections (see Fig. 2) making cluster analysis impossible. For this reason, the cluster analysis had to be performed twice: first, without the two non-overlapping series to define the two clusters, and second, changing the pairs of series, to include the series that were excluded in the first step to determine to which cluster they belong. The result of cluster analysis for the first step is shown in Fig. 3, and the excluded series were assigned to cluster 2.

3.2 Homogenization methods

Two modern homogenization methods, HOMER (Mestre et al., 2013) and ACMANT (Domonkos, 2011b, 2014) were applied. Both methods were developed during the HOME Cost Action (Venema et al., 2012). HOMER was designed to include the best segments and features of some other state-of-the-art methods: PRODIGE (Caussinus and Mestre, 2004), ACMANT and Joint Detection (Picard et al., 2011).

ACMANT can be applied on both daily and monthly data, but as HOMER works only with monthly data, we ran both methods on monthly data. The two methods have several similarities: both methods are based on the optimal step function fitting (Hawkins, 1972) with the Caussinus Lyazrhi criterion (Caussinus and Lyazrhi, 1997) for optimizing the number of steps (also referred to as “breaks”). Both methods also include the bivariate detection for shifts in the annual means and the summer–winter differences (Domonkos, 2011b), and the minimization of the residual variance (ANOVA, Caussinus and Mestre, 2004) in finding the optimal adjustment terms.

On the other hand, the two methods differ in several other aspects: while HOMER implements a pairwise comparison and a network-wide harmonization in the break detection, ACMANT uses weighted reference time series. A new feature of the most recent ACMANT version (Domonkos, 2014) is that it can also detect relatively short-term inhomogeneities, which are known to be important for long-term data quality (Domonkos, 2011a).

As ACMANT is fully automatic, it can easily be applied to large datasets, while the interactive HOMER allows human intervention to the homogenization procedure and it is possible to decide about the significance of indicated breaks, based on metadata or research experience (Mestre et al., 2013).

HOMER was run comparing all the stations from each cluster separately with annual and seasonal detection, while using ACMANT, again all the time series within clusters were used together and the outlier filtering “off” option was selected, as the input dataset had been quality controlled earlier. HOMER is a user dependent method, and the main way of running the program can be summarised in three steps. First, big break points are identified and corrected. The second step is to repeat the detection in order to evaluate which break points are identified by metadata, and detect those breaks which have smaller amplitude than previously corrected. Finally, annual series are compared by plotting QC data with the homogenized series output. These plots (not shown) allow the user to understand and review the corrections applied. For detection, the three available methods (pairwise detection, joint-segmentation method and ACMANT detection) were considered.

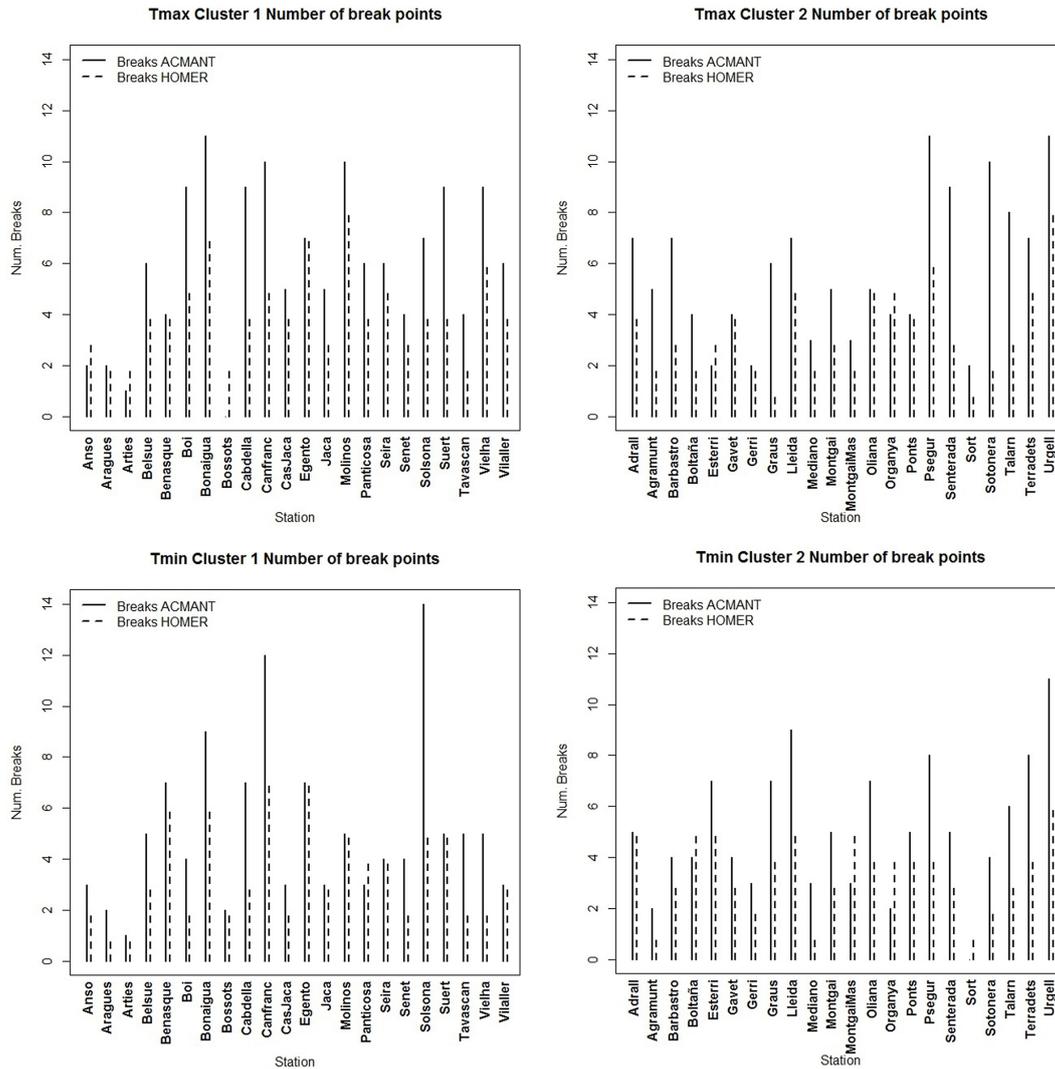


Figure 4. Number of break points detected by ACMANT (continuous line) and HOMER (dashed line) for each station for T_{max} (top panels) and T_{min} (bottom panels) in cluster 1 (left panels) and cluster 2 (right panels).

3.3 Comparison methods

Only periods with data in the QC dataset are considered to compare the output of the two homogenization methods. The period of examination was determined by ACMANT, because ACMANT needs at least 4 spatially comparable series for each section of the homogenization period and this minimum condition is stricter than that of HOMER.

The number of breaks detected by HOMER and ACMANT, spatial connections of homogenized data, as well as trend slopes of homogenized series were analysed. Spatial connections were examined using Spearman Correlation Coefficients (SCC) calculated between all pairs of monthly series for QC data, data homogenized by HOMER and data homogenized by ACMANT in each cluster. The trend analysis was calculated for the period 1961–1990 in all those stations with more than 80 % of the monthly data available during

this period. 12 series met with this condition in each cluster. This analysis was performed by linear regression on the annual series including years in which all monthly values were available. Significance of trends was evaluated using the Student’s t test ($p < 0.05$) (Wilks, 2011).

4 Results

4.1 Break point analysis

HOMER detected at least 1 break in all series, while ACMANT did not detect any break in two of the 88. However, the maximum number of breaks detected with ACMANT (14) was much higher than that with HOMER (8) as shown in Fig. 4. In 90% of the series ACMANT detected more breaks than HOMER. The average difference was 2 more breaks per series using ACMANT, which de-

Table 1. Slope trends in °C decade⁻¹ for the period 1961–1990 for each station with more than 80 % of monthly data in this period. Significant trends (evaluated using the Student's *t* test with a significance level of $p < 0.05$) are shown in bold.

Cluster	Station	QC- T_{\max} (°C dec ⁻¹)	HO- T_{\max} (°C dec ⁻¹)	AC- T_{\max} (°C dec ⁻¹)	QC- T_{\min} (°C dec ⁻¹)	HO- T_{\min} (°C dec ⁻¹)	AC- T_{\min} (°C dec ⁻¹)
1	Arties	-0.20	0.01	0.27	-0.25	-0.22	0.00
1	Belsue	0.94	0.21	0.68	0.64	0.00	0.16
1	Cabdella	-0.17	0.06	0.52	0.00	-0.09	0.09
1	Canfranc	0.82	0.04	0.59	0.88	-0.03	0.15
1	Egento	0.68	-0.24	0.18	1.75	-0.27	0.03
1	Molinos	0.31	0.05	0.30	-0.10	-0.01	0.20
1	Panticosa	-0.16	0.05	0.57	-0.41	-0.13	0.15
1	Seira	0.17	-0.27	0.44	0.12	-0.25	0.00
1	Senet	0.24	0.15	0.61	-0.25	0.04	0.45
1	Suert	0.70	0.19	0.75	-0.21	0.06	0.20
1	Vielha	0.17	0.40	0.63	0.26	0.25	0.26
1	Vilaller	1.21	0.39	0.77	-0.01	0.06	0.16
2	Adrall	1.60	0.18	0.32	1.25	-0.01	-0.08
2	Boltaña	1.04	0.21	0.19	-0.27	0.24	-0.10
2	Gavet	0.65	0.23	0.32	-0.27	0.08	-0.12
2	Graus	-0.05	-0.05	0.28	0.29	0.12	-0.15
2	Montgai	-1.06	0.21	0.32	0.20	0.20	-0.01
2	Oliana	0.59	0.15	0.28	1.15	0.12	-0.11
2	Ponts	1.32	0.27	0.47	-0.73	0.38	0.21
2	Psegur	0.31	-0.01	0.21	-0.83	-0.21	-0.22
2	Senterada	1.54	0.25	0.33	-0.13	0.00	-0.06
2	Sotonera	-0.12	0.11	0.37	0.26	0.16	-0.05
2	Talarn	-0.20	0.12	0.37	-0.01	0.00	-0.01
2	Terradets	0.09	0.09	0.35	0.07	0.02	-0.09

tected 5.5 break points per series on average (10 break points per century), than with HOMER, which detected 3.6 (7 break points per century).

To evaluate the similarity between the results obtained by these homogenization methods, break points per year for each station were compared, considering that the timing of a break point can differ by up to 1 year between both methods due to the difference in the rest of break points. For T_{\max} in cluster 1 (2), ACMANT detected 143 (126) break points of which 59 (51) were also detected by HOMER. For T_{\min} in cluster 1 (2), ACMANT detected 113 (117) break points of which 49 (42) were also detected by HOMER.

As HOMER is an interactive method that allows the user to introduce known break points, all dates identified in the metadata were included. However, some were removed during the homogenization procedure because the magnitude of the break was less than 0.05 °C and their presence didn't show an improvement in the correction of the series compared with the QC data. From the 8 (15) metadata-supported breaks stored for cluster 1 (2), ACMANT detected 2 (4) break points for T_{\min} and 4 (6) T_{\max} (all of them detected also by HOMER), while HOMER detected 7 (12) for T_{\min} and 6 (13) for T_{\max} .

4.2 SCC comparison

SCC values are useful indicators to visualize the temporal linear relationship between time series before and after homogenization (Freitas et al., 2013), and in indicating the presence of large inhomogeneities when they exist. In general, the variance of the SCC for the series in cluster 1 was greater than that for cluster 2. The minimum correlation value for the first cluster was 0.65 while for the second cluster the minimum was 0.90, as shown in Figs. 5 and 6. One reason for this may be a large error in the Vielha observatory data (cluster 1) that was detected after the homogenization process: from 2004 to 2007 the seasonal cycle of temperature seems to be inverted or lagged by a few months for both T_{\max} and T_{\min} (Fig. 7), although the origin of this error is unknown. This error was detected and corrected adequately only for T_{\max} with ACMANT. With HOMER, we failed to be recognized due to the smoothness of the annual values. This type of oversight could be avoided using the CLIMATOL QC check that is also included in HOMER. In ACMANT homogenization of T_{\min} the seasonal cycle error remained untouched, while using HOMER, it was even propagated to earlier sections of the series.

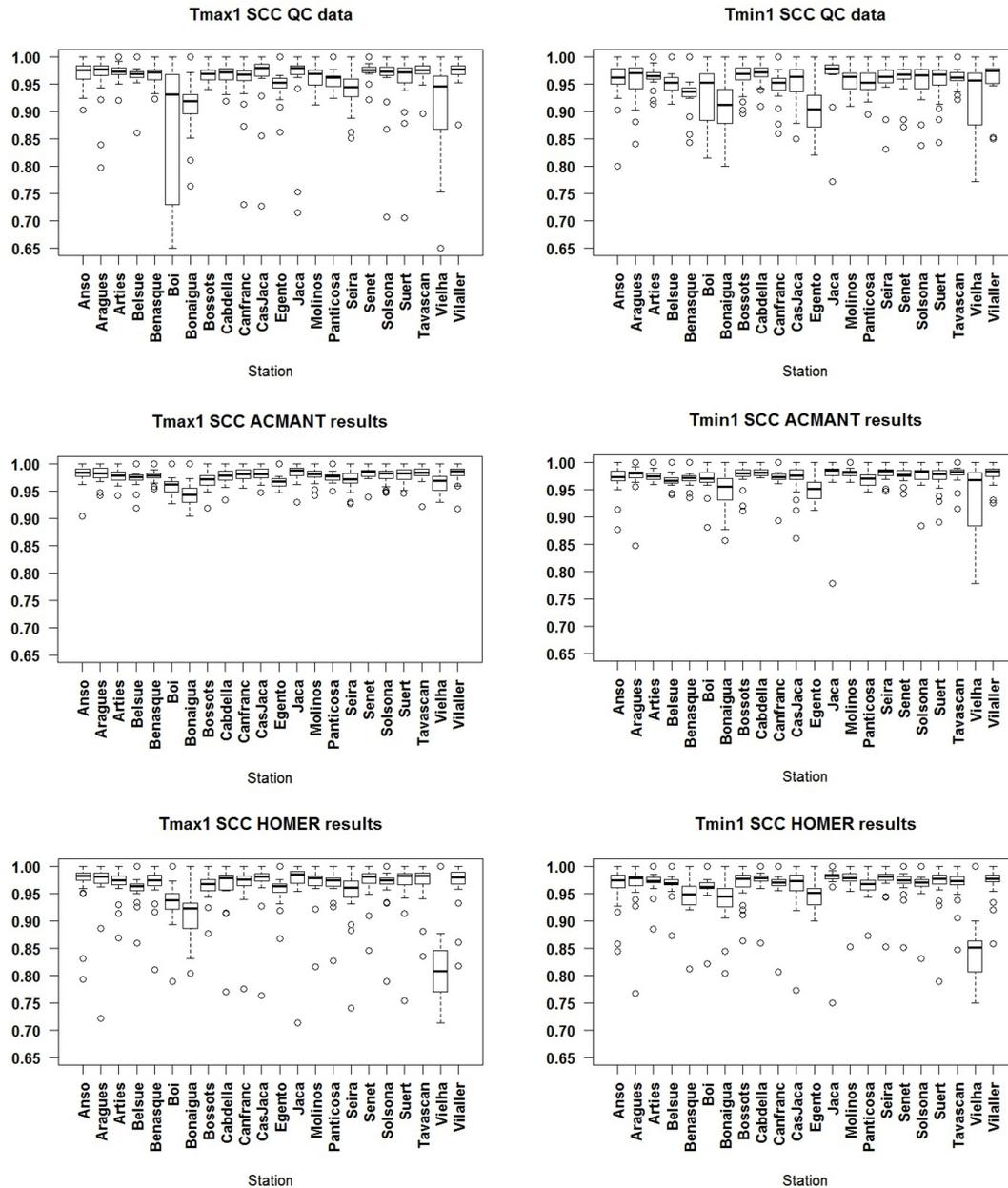


Figure 5. Boxplots of Spearman Correlation Coefficients of QC data (top panels), and ACMANT (middle panels) and HOMER (bottom panels) homogenized data for maximum (left panels) and minimum (right panels) temperature for stations in cluster 1.

4.3 Trend analysis

After homogenization, spatial gradients of trend slopes became smaller, and the number of significant positive trends was reduced as shown in Table 1.

For T_{max} , all 5 significant trends of the 12 series in cluster 1 were positive in the QC data. After homogenization with HOMER, the number of significant and positive trends decreased to 1, while with ACMANT it increased to 9. For cluster 2, 6 of the 12 series had positive and significant trends in the QC data. HOMER didn't return any significant trends

for this cluster, while with ACMANT 8 significant positive trends were obtained. None of the homogenization methods returns negative and significant trends for T_{max} . For T_{min} of cluster 1, only 4 positive significant trends occurred in the QC data. HOMER returned 1 positive and 1 negative significant trends, while the ACMANT homogenized series produced 2 positive and zero negative trends. In cluster 2, the QC data presented 3 positive and 2 negative significant trends. After homogenizing, HOMER kept 1 positive but zero negative significant trends, while with ACMANT all trends were not significant.

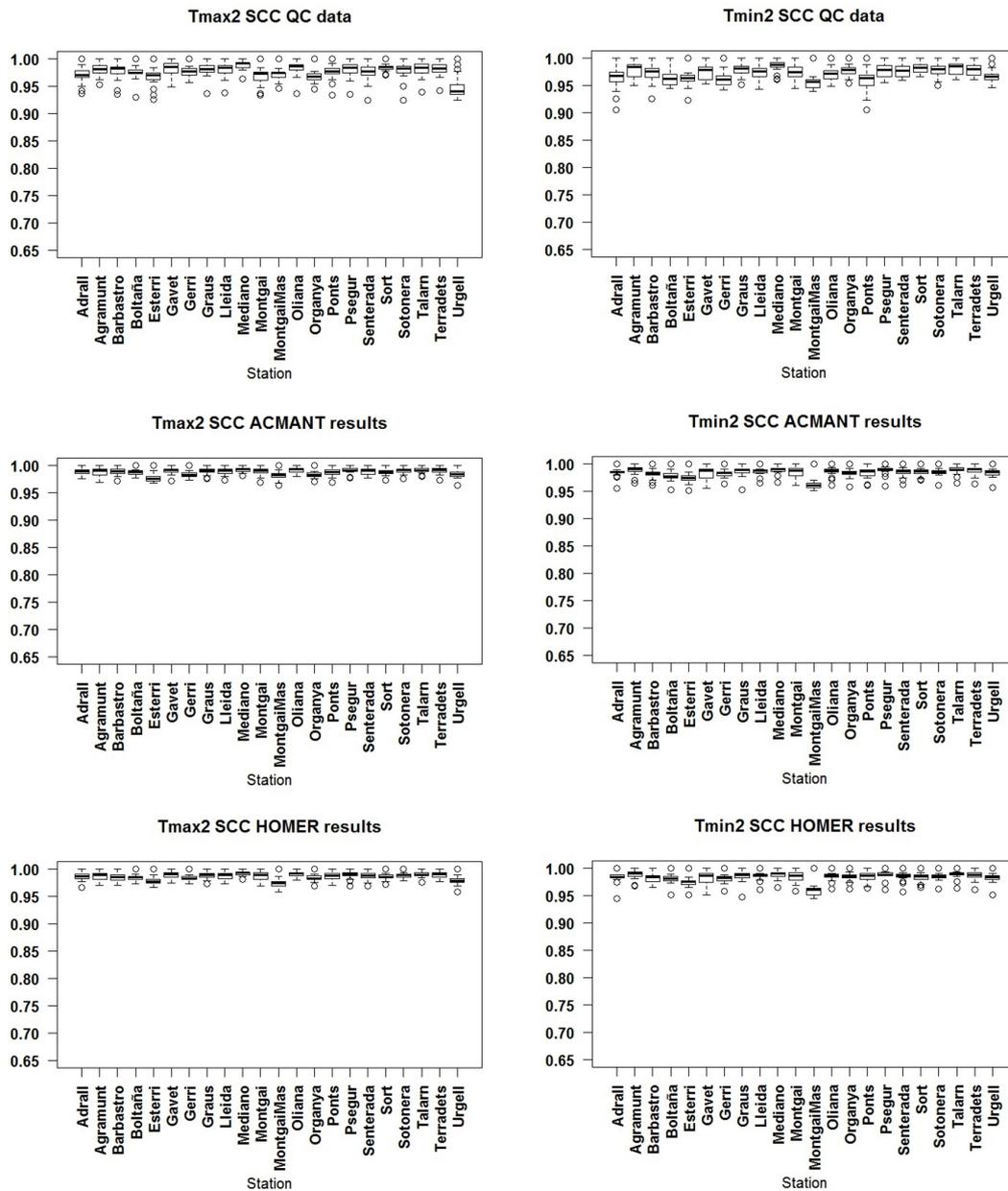


Figure 6. Boxplots of Spearman Correlation Coefficients of QC data (top panels), ACMANT (middle panels) and HOMER (bottom panels) homogenized data for maximum (left panels) and minimum (right panels) temperature for stations in cluster 2.

The relatively large differences between the HOMER homogenization results and ACMANT homogenization results in the mean T_{max} trends and the number of significant positive trends in T_{max} series are unexpected results and their origin requires further analysis.

5 Discussion and conclusions

ACMANT and HOMER are two modern, partly similar, multiple break point homogenization methods, but they have distinct strengths and weaknesses. While automatic methods

such as ACMANT are easy to use for large datasets, human intervention and the consideration of metadata is possible only with interactive methods like HOMER.

In this case study of Pyrenees temperatures, ACMANT detected and corrected more breaks than HOMER, which is in agreement with the developed sensitivity of ACMANT to detect short-term biases. Concerning breaks justified by metadata, HOMER detected a larger number than ACMANT, showing the advantage of using interactive homogenization methods. Note however, that one cannot conclude on the accuracy of methods from the number of detected breaks, since

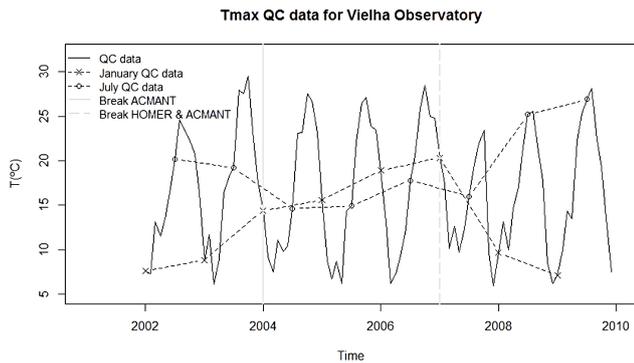


Figure 7. Vielha QC series for T_{\max} from 2002 to 2009 continuous line), January data (dashed line and asterisks) and July data (dashed line and circles). Detected breaks of ACMANT (HOMER and ACMANT) are indicated by gray continuous (dashed) vertical lines.

we do not know the number and exact position and size of breaks in the observed dataset. Detailed evaluation of efficiencies requires the use of artificially developed benchmark datasets.

We have identified a serious error in Vielha T_{\min} and T_{\max} series, which spectacularly affected the SCC values in cluster 1. This error was corrected well with ACMANT for T_{\max} , but not for T_{\min} , since the control of seasonal changes is included only in the homogenization of T_{\max} with ACMANT. Concerning the homogenization with HOMER, the program outputs indicated the error, but the indications were left out of consideration, due to the smoothness of the annual means. This rare error in the data handling of Vielha time series points to the necessity of applying a thorough, multi-functional data quality control, since ideally, homogenization procedures should be applied on datasets that are free from such large errors.

In this case study, average trends for all stations in the period 1961–1990 for T_{\max} using HOMER (ACMANT) are $0.12\text{ °C decade}^{-1}$ ($0.43\text{ °C decade}^{-1}$) and, for T_{\min} , using HOMER (ACMANT) the average trend slope is $0.03\text{ °C decade}^{-1}$ ($0.05\text{ °C decade}^{-1}$). Comparing these trends with other studies of homogenized Pyrenees temperature data reveals mixed results. A single-station study of annual T_{\max} and T_{\min} for 1882–1970 (Bücher and Dessens, 1991) described the opposite to what is found here, with a negative trend identified for T_{\max} and a positive for T_{\min} . Previous homogenization of three stations using HOMER (Esteban et al., 2012) showed significant and positive trends on annual T_{\max} for all stations, although no significant trend was identified for annual T_{\min} in the period 1935–2008; during the shorter 1950–2008 period however, T_{\max} and T_{\min} were found to be positive and significant in all of the three stations (Esteban et al., 2012). Finally, a study of annual T_{mean} for the Pyrenees over 1950–2010 showed an increase of $0.2\text{ °C decade}^{-1}$ (Cuadrat et al., 2013). Three factors can explain these differences in the detected trends. First, the

different time periods in focus; second, the homogenization methods applied, and third, the differences in the number and geographical distribution of stations.

In conclusion, the high SCC results achieved indicate that the homogenization was generally successful with both HOMER and ACMANT, although the difference in T_{\max} trend slopes and particularly the handling of Vielha error points to the need of further methodological analysis.

Acknowledgements. This work was supported by the FP7-SPACE-2013-1 project (Uncertainties in Ensembles of Regional Reanalyses, UERRA) and the MONTCORTES project (CGL2012-33665). We grateful to three anonymous reviewers for comments that helped improve this manuscript.

Edited by: I. Auer

Reviewed by: three anonymous referees

References

- Aguilar, E., Auer, I., Brunet, M., Peterson, T. C., and Wieringa, J.: Guidelines on climate metadata and homogenization, WCDMP Report No. 53, WMO-TD 1186, World Meteorological Organization, Geneva, 2003.
- Barros, V. R., Field, C. B., Dokken, D. J., Mastrandrea, M. D., Mach, K. J., Bilir, T. E., Chatterjee, M., Ebi, K. L., Estrada, Y. O., Genova, R. C., Girma, B., Kissel, E. S., Levy, A. N., MacCracken, S., Mastrandrea, P. R., and White, L. L. (Eds.): IPCC 2014: Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects, in: Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, UK and New York, NY, USA, 2014.
- Bradley, R. S.: Paleoclimatology: reconstructing climates of the Quaternary, Academic Press, University of Massachusetts, Amherst, USA, 1999.
- Brunet, M., Saladié, O., Jones, P., Sigró, J., Aguilar, E., Moberg, A., Lister, D., Whalther, A., and Almarza, C.: A case-study/guidance on the development of long-term daily adjusted temperature datasets, World Meteorological Organization WCDMP Report No. 66, WMO-TD-1425, World Meteorological Organization, Geneva, 2008.
- Bücher, A. and Dessens, J.: Secular Trend of Surface Temperature at an Elevated Observatory in the Pyrenees, *J. Climate*, 4, 859–868, 1991.
- Caussinus, H. and Lyazrhi, F.: Choosing a linear model with a random number of change-points and outliers, *Ann. Inst. Statist. Math.*, 49, 761–775, 1997.
- Caussinus, H. and Mestre, O.: Detection and correction of artificial shifts in climate series, *J. Roy. Stat. Soc. C*, 53, 405–425, 2004.
- Cuadrat, J. M., Serrano, R., Saz, M.Á., Tejedor, E., Prohom, M., Cunillera, J., Esteban, P., Soubeyroux, J. M., and Deaux, N.: Creación de una base de datos homogeneizada de temperaturas para los Pirineos (1950–2010), *Geographicalia*, 63–64, 63–74, 2013.

- Domonkos, P.: Efficiency evaluation for detecting inhomogeneities by objective homogenisation methods, *Theor. Appl. Climatol.*, 105, 455–467, 2011a.
- Domonkos, P.: Adapted Caussinus-Mestre algorithm for networks of temperature series (ACMANT), *Int. J. Geosci.*, 2, 293–309, 2011b.
- Domonkos, P.: The ACMANT2 software package, in: Eighth Seminar for Homogenization and Quality Control in Climatological Databases, World Climate Monitoring Program (WCDMP), www.c3.urv.cat/publications/publications.html (last access: 1 March 2015), 2014.
- Esteban Veà, P., Prohom Duran, M., and Aguilar E.: Tendencias recientes e índices de cambio climático de la temperatura y la precipitación en Andorra, Pirineos (1935–2008), *Pirineos*, 167, 87–106, 2012.
- Freitas, L., Gonzalez Pereira, M., Caramelo, L., Mendes, M., and Filipe Nunes, L.: Homogeneity of monthly air temperature in Portugal with HOMER and MASH, *Idojaras*, 117, 69–90, 2013.
- Hawkins, D.M.: On the choice of segments in piecewise approximation, *J. Inst. Math. Appl.*, 9, 250–256, 1972.
- Kilibarda, M. and Bajat, B.: plotGoogleMaps: the R-based web-mapping tool for thematic spatial data, *Geomatica*, 66, 37–49, 2012.
- López-Moreno, J. I., Goyette, S., and Beniston, M.: Climate change prediction over complex areas: spatial variability of uncertainties and predictions over the Pyrenees from a set of regional climate models, *Int. J. Climatol.*, 28, 1535–1550, 2008.
- López-Moreno, J. I., Goyette, S., and Beniston, M.: Impact of climate change on snowpack in the Pyrenees: Horizontal spatial variability and vertical gradients, *J. Hydrol.*, 374, 384–396, 2009.
- Mestre, O., Domonkos, P., Picard, F., Auer, I., Robin, S., Lebarbier, E., Böhm, R., Aguilar, E., Guijarro, J., Vertachnik, G., Klancar, M., Dubuisson, B., and Stepanek, P.: HOMER: a homogenization software – methods and applications, *Idojaras*, 117, 47–67, 2013.
- Picard, F., Lebarbier, E., Hoebeke, M., Rigail, G., Thiam, B., and Robin, S.: Joint segmentation, calling, and normalization of multiple CGH profiles, *Biostatistics*, 12, 413–428, 2011.
- Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom, M., Lisko, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, *Clim. Past*, 8, 89–115, 2012, <http://www.clim-past.net/8/89/2012/>.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, Academic Press, Cornell University, Ithaca, New York, USA, 2011.
- Willett, K., Williams, C., Jolliffe, I. T., Lund, R., Alexander, L. V., Brönnimann, S., Vincent, L. A., Easterbrook, S., Venema, V. K. C., Berry, D., Warren, R. E., Lopardo, G., Auchmann, R., Aguilar, E., Menne, M. J., Gallagher, C., Hausfather, Z., Thorarinsdottir, T., and Thorne, P. W.: A framework for benchmarking of homogenisation algorithm performance on the global scale, *Geosci. Instrum. Method. Data Syst.*, 3, 187–200, doi:10.5194/gi-3-187-2014, 2014.