IDŐJÁRÁS
Quarterly Journal of the Hungarian Meteorological Service
Vol. 117, No. 1, January–March 2013, pp. 69–90

Homogeneity of monthly air temperature in Portugal with HOMER and MASH

Luís Freitas[1]*, Mário Gonzalez Pereira[1,2], Liliana Caramelo[1], Manuel Mendes[3], and **Luís Filipe Nunes[4]**

[1] *Centre for Research and Technology of Agro-Environment and Biological Sciences (CITAB), University of Trás-os-Montes and Alto Douro, Apartado 1013, 5001-801 Vila Real, Portugal*

[2] *Instituto Dom Luiz – Universidade de Lisboa, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, Edifício C8, Piso 3, 1749-016 Lisboa, Portugal*

[3] *Instituto de Meteorologia, IM, I.P., Rua C, Aeroporto de Lisboa, 1749-077 Lisboa, Portugal*

[4] *Laboratory for Systems, Instrumentation and Modeling in Science and Technology for Space and the Environment, Faculty of Sciences of University of Lisbon, Campo Grande, Edifício C1, Gabinetes 1.4.21 e 1.4.39, P-1749-016 Lisboa, Portugal*

**Corresponding author E-mail: pedro-fafe@hotmail.com*

(Manuscript received in final form November 25, 2012)

Abstract–In this paper we focus on the homogeneity of Portuguese monthly mean air temperature with two purposes: i) to detect and correct eventual inhomogeneities in the dataset; and, ii) to compare the homogenized time series with different methods. The dataset used in this study comprises time series of minimum (TN) and maximum (TX) monthly mean air temperature recorded in weather stations located in the northern region of the continental part of Portugal, from 1941 to 2010. MASH and HOMER were the methods used in this study to homogenize the Portuguese air temperature database. The former was selected for being one of the most widely used by the homogenization community, while the latter was selected because it is one of the most recent homogenization methods, and the combination of detection methods resulted in that, along with MASH, HOMER exhibited the best results in the comparative analysis performed within the COST Action ES0601 (HOME). A high number of break points were identified in both minimum and maximum air temperature time series, but differences in the number, size and temporal location of the breaks detected by both methods must be underlined. The homogenization process was assessed by comparing results obtained with correlation, trend, and principal component analysis using non-homogenized (NH) and homogenized datasets with both methods. Correlation analysis reveals a higher increase in the similarity in homogenized TX than in TN in relation with NH time series. Decrease in the amplitude of the tendencies and in the number of statistically significant trends is higher in homogenized TX than in TN, independently of the homogenization method. On the other hand, the number of statistically significant principal components tend to decrease with the application of homogenization procedures, while the explained variance by the first principal components of homogenized datasets is tendentiously higher than for non-homogenized datasets.

Key-words: Homogenization, temperature, MASH, HOMER, Portugal.

1. Introduction

The existence of long and reliable instrumental climate records registered in a sufficiently dense network is fundamental to assess climate variability and climate change and to validate climate models. Climate research results are also dependent on the quality of the datasets, in particular on its homogeneity (*Venema et al.*, 2012). A homogeneous climate time series can be defined as the one whose variability is only caused by changes in weather and climate (*Aguilar et al.*, 2003). However, long instrumental records are rarely homogeneous because they include non-climatic signals which must be removed. Results from the homogenization of Western Europe climate records points to the existence of inhomogeneities in mean temperature series every 15 to 20 years (*Venema et al.*, 2012). In fact, any weather observation network, that operates for a long period of time, undergoes changes in its functioning due, for example, to instrumentation failure or damage, changes on its surrounding (e.g., urbanization), relocation and substitution of weather stations. For these reasons, it is expected that the Portuguese maximum and minimum air temperature datasets present heterogeneities that need to be detected and corrected.

In the last decades, inhomogeneity detection techniques have been developed based on classical statistical tests (*Alexandersson*, 1986; *Gullett et al.*, 1990), regression models (*Vincent*, 1998), or Bayesian approaches (*Perreault et al.*, 2000). More recently, new procedures were particularly developed to detect and correct multiple change-points using reference series (*Szentimrey*, 1999; *Mestre*, 1999; *Caussinus* and *Mestre*, 2004; *Menne* and *Williams*, 2005) and changes in the mean and variance (*Toreti et al.*, 2012). Review papers and comparison studies of homogenization methods have been published regularly (*Peterson et al.*, 1998; *Ducré-Robitaille*, 2003, *Reeves et al.*, 2007, *Venema et al.*, 2012). Some authors have been focusing their interest in specific aspects of the homogenization procedure such as the cause of inhomegeneities (*Trewin*, 2010), use of reference series (*Menne* and *Willians*, 2005, *Domonkos*, *et al.*, 2012), ability of homogenization methods (*Menne* and *Willians*, 2005), or to test automatic homogenization methods by the introduction of perturbed parameter experiments (*Williams et al.*, 2012).

The inventory and evaluation of existing detection and correction methods and the need of an objective comparative analysis to assess their performance was included in the scientific programme of the COST Action HOME ES0601: Advances in Homogenization Methods of Climate Series: an integrated approach (HOME). HOME results include the publication of a comparison study, based on 25 blind contributions and 22 contributions made after knowing the location and size of the heterogeneities, performed with a large number of different versions of 9 main methods (*Venema et al.*, 2012). This study was based on a benchmark dataset of monthly air temperature and precipitation and on different error metrics to assess the performance of the methods. Results of this comparison suggests that: (i) the assessment of the methods is dependent on the error metric considered; (ii) in general, all relative methods contribute to homogenized temperature data; but, (iii) only the methods with best performance are able to improve the quality of precipitation datasets; and, (iv) the list of methods with better performance includes Craddock (*Craddock*, 1979), PRODIGE (*Caussinus* and *Mestre*, 2004), MASH (*Szentimerey*, 2007), ACMANT (*Domonkos*, 2011), and USHCN methods (*Menne* and *Williams*, 2009).

HOME main objective was to develop a general homogenization method for homogenizing climate and environmental datasets which was accomplished in 2011 with the release of a free software package (HOMER), implemented in R language (*HOME*, 2011). It should be noted that ACMANT is a modified and automated version of PRODIGE, and that HOMER integrates PRODIGE, ACMANT, and USHCN.

Consequently, the purpose of this study is twofold: (i) to analyze the homogeneity of minimum and maximum air temperatures in northern Portugal; and, (ii) to compare the homogenized maximum and minimum air temperatures Portuguese datasets with HOMER and MASH. A review of the main characteristics of the procedures used to control the quality of the data and methods of homogenization will be undertaken in order to justify the options taken in this study and to highlight the methodological differences between MASH and HOMER.

2. Dataset description

The dataset that we analyze here is representative of the monthly mean maximum and minimum air temperature fields (hereafter TX and TN, respectively) in the northern region of the continental part of Portugal for the 1941–2010 period. Monthly time series were calculated from daily values, following the WMO directives in what concerns to the existence of missing values in daily time series. Specifically, a monthly value should only be computed if no more than five consecutive daily values or less than ten daily values throughout the month are missing (*WMO*, 2011).

Daily values of TX and TN were recorded at weather stations managed by the Portuguese Meteorological Institute (IM). Location and characteristics of these weather stations are presented in *Fig. 1* and *Table 1*, respectively. This network comprises both classical weather stations (CWS), collecting data since the mid-1800s, and automatic weather stations (AWS), installed in the end of the 20th century. In cases where AWS were installed in approximately the same location of the CWS, the time series from both weather stations were merged, the type of station in *Table 1* was set to CWS/AWS, and the date of the fusion was stored as metadata. Maximum distance between an AWS and CWS used to produce the merged time series was 4.7 km (in Vila Real), which is a much lower distance than those used in previous studies (*Stepanek* and *Mikulova*, 2008; *Vicente-Serrano et al.*, 2010).

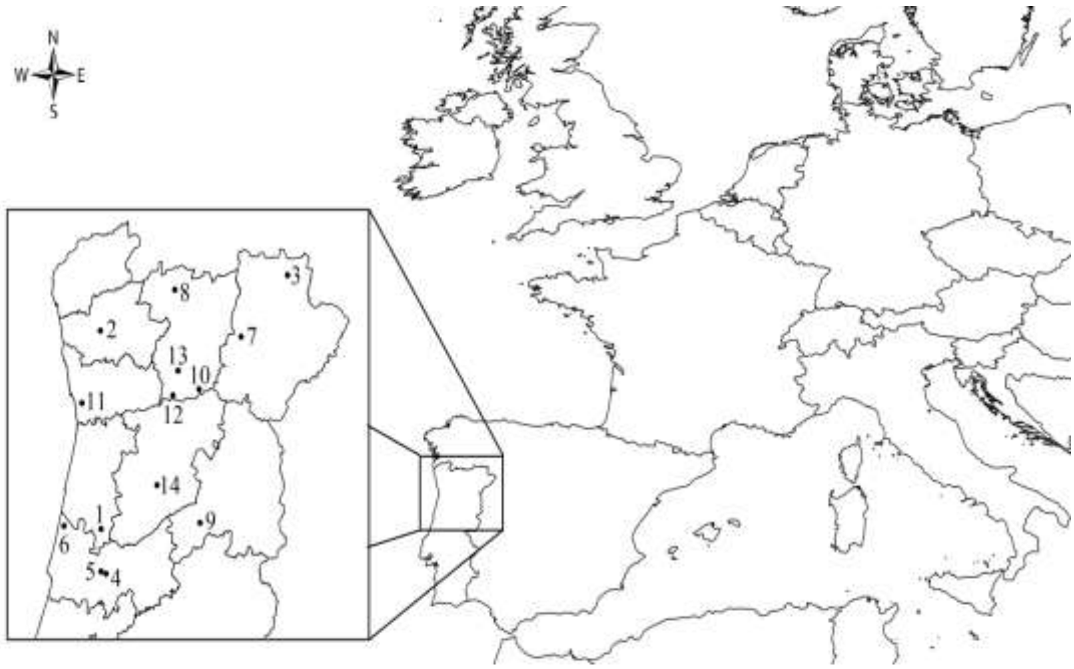

Fig. 1. Location of the weather stations of the Portuguese Institute of Meteorology (IM) network, in northern Portugal. Addition characteristics of these stations are provided in *Table 1*.

In this network, weather stations are well distributed and located both in low and high altitude (ranging from 14 m to 1380 m), in densely populous coastal areas and sparsely populated inner regions within the country territory (*Fig. 1*). The northern Portugal is characterized for being the region with the highest density of mountains and river basins in the country as well as by a diverse land use/occupation (*Freitas et al.*, 2012). Independently of the proximity to the Atlantic Ocean or the altitude, all weather stations considered in this study are located in a region of C_s type of climate, which is a temperate climate with dry period in summer (*AEMET-IM*, 2011). In more detail, the climate of this northern region is essentially of C_{sb} type, which corresponds to a temperate climate with dry or temperate summer, except a small part, in the northeast, which is of type C_{sa}, also temperate but with dry or hot summer. The recently published Iberian Climate Atlas (*AEMET-IM*, 2011) provides a brief history of the complete IM network and additional description and characteristics of the temperature dataset. Results of the exploratory preliminary statistical analysis of minimum and maximum air temperature datasets for the 1941–2010 period are presented and discussed in *Freitas et al.* (2012).

Table 1. Characteristics of the weather stations of the Portuguese Institute of Meteorology (IM) network, located in northern Portugal including: identification code (ID); stations name; station type; altitude (m); start and ending dates; and, amount of missing values (in %), accounted for the 1941–2010 period. When the entire time series results from measurements from a CWS (or AWS), the type is simply CWS (or AWS); in the cases where a CWS was replaced by a AWS, the type is CWS/AWS

ID	Station Name	Type	Altitude (m)	Start year	End year
1	Anadia (AN)	AWS	45	1941	2010
2	Braga (BR)	CWS/AWS	65	1931	2010
3	Bragança (BG)	CWS	690	1932	2010
4	Coimbra B. (CB)	CWS	35	1941	2010
5	Coimbra G. (CG)	CWS	141	1864	1996
6	Dunas Mira (DM)	CWS	14	1935	2005
7	Mirandela (MI)	CWS/AWS	250	1926	2010
8	Montalegre (MO)	CWS/AWS	1050	1880	2010
9	Penhas D. (PD)	CWS/AWS	1380	1932	2010
10	Pinhão (PI)	CWS/AWS	130	1941	2010
11	Porto S.P. (PS)	CWS	93	1863	2005
12	Régua (RE)	CWS	56	1933	2010
13	Vila Real (VR)	CWS/AWS	561	1928	2010
14	Viseu (VI)	CWS/AWS	443	1925	2010

3. Methodological procedures

This section is devoted to the description of the methods used to perform the quality control of the data, the homogeneity analysis and to compare the homogenized datasets with those methods. The quality control of the datasets comprises a preliminary exploratory statistical analysis to characterize the potential and limitations of the datasets as well as to identify and correct missing values and outliers. Main technical features of the procedures used in this study will be briefly discussed to validate the followed methodology and to underline the major differences between the approaches of the two selected methods to homogenize the Portuguese air temperature dataset.

3.1. Quality control with homogenization methods packages

In this study two homogenization methods were used: (i) the most recent version of MASH, (Version MASHv3.03), initially developed in the Hungarian Meteorological Service by *Szentimrey* (1994, 1999); and, (ii) HOMER, developed in the framework of COST Action ES0601 (*HOME*, 2011). We start with presenting the homogenization methods because, in addition to being able to detect and correct inhomogenieties, these softwares comprise additional functions to perform fast quality control. On this subject, with MASH it is obligatory to use available functionalities to fill the missing values and perform automatic correction of outliers. On the other hand, HOMER provides a fast quality control of the data, which includes functions of the CLIMATOL R package (*Guijarro*, 2011), which allow the user to perform/estimate station density, correlogram, histograms, boxplots, and cluster analysis. With respect to the detection of heterogeneities, MASH relies on multiple references series while HOMER combines three detection algorithms: pairwise – univariate detection (*Caussinus* and *Lyazrhi*, 1997), joint detection (*Picard et al.*, 2011), and ACMANT – bivariate detection (*Domonkos et al.*, 2012). To correct the datasets, MASH uses multiple comparison techniques whereas HOMER uses ANOVA. MASH is provided with a user guide, while a brief description of HOMER can be found in *Mestre* and *Aguilar* (2011) or in *Freitas et al.* (2012).

3.2. Outlier detection

It is recommended to use different methods for outlier detection because, in general, one single method/criteria is not sufficient to identify real outliers nor to exclude false detections (*Stepanek et al.*, 2009). Consequently, in this study, abnormal high and low values were only classified as outliers if two criteria were simultaneously verified: (i) values above/below the upper/lower thresholds defined as the upper/lower quartiles plus/minus the interquartile range times a coefficient (usually equal to 1.5 to detect outliers and equal to 3.0 to detect extreme values); and, (ii) pairwise comparison which is based on the difference time series between candidate and best neighbor time series, which can be defined as the closer stations and/or those presenting higher correlation (*Stepanek et al.*, 2009; *Syrakova* and *Stefanova*, 2009). This latter procedure can be performed in HOMER by visual inspection of the plots of the difference between candidate and best neighbor time series. As mentioned in the previous section, in addition to this analysis, MASH has an independent and automatic procedure to detect and correct outliers that is executed before detection procedures.

3.3. Missing values correction

The existence of missing data in climate time series can be solved with temporal interpolation, using data of the same time series before and after the data gap, or with spatial interpolation, using data from nearby weather stations (*WMO*, 2011). Complex estimation methods, such as weighted averages, spline functions, linear regression, and kriging, which take into account the correlations with other elements, can also be used to complete the time series. *Brunetti et al.*, (2006) adopted a procedure to fill the gaps on monthly precipitation and temperature Italian time series, with estimates based on the highest correlated reference series. For temperature, this method is based on the differences between incomplete and reference temperature series. *Staudt et al.* (2007), replace the missing values on monthly time series of Spanish minimum and maximum temperatures by weighted means of the best-correlated synchronous data. The method used by *Syrakova* and *Stefanova* (2009) to fill the gaps in Bulgarian monthly temperature is based on the stability of the differences between the time series at neighboring highly correlated stations. More recently, *Vicente-Serrano et al.* (2010) tested three different procedures to fill missing data in daily precipitation time series: (i) the nearest neighbor, (ii) inverse distance weighted interpolation; and, (iii) linear regression methods, concluding that the nearest-neighbor method provided the best results. Both homogenization methods used in this study (MASH and HOMER) have corrected databases as final result with respect to inhomogeneities and missing values using multiple comparison and ANOVA, respectively.

3.4. Reference time series

Reference series or reference sections are used in detection procedures in many homogenization methods, such as ACMANT, AnClim/ProClimDB, Climatol, RHTestV3, and MASH (*WMO*, 2011). Reference series are also used to assess the quality of the homogenization (*Kuglitsch et al.*, 2009). These reference series do not need to be homogeneous (*Szentimrey*, 1999; *Zhang et al.*, 2001; *Causinus* and *Mestre*, 2004), but must encompass the same climatic signal as the candidate series (*Della-Marta* and *Wanner*, 2006) and, in this sense, are usually produced as weighted averages of the time series from surrounding stations

(*Peterson* and *Easterling*, 1994; *Sahin* and *Cigizoglu*, 2010). *Stepanek* and *Mikulova* (2008) discuss the advantages and disadvantages producing weighted reference series based on the distance between stations or on the correlation between candidate and potential time series, while *Della-Marta* and *Wanner* (2006) argue about the benefits of using weighted reference series in comparison with a single reference station. The selection procedure of the surrounding stations to produce the reference series can be based on the distance between stations or on the correlation between candidate and potential time series. Both criteria present advantages and disadvantages that must be underlined. Distance-based methods preserve the geographical vicinity, but time series from near stations with different climatic signals (e.g., due to altitude) can be selected. Using high correlated neighbor time series, both the candidate and reference series present similar variability (which reduces differences/ratios time series variability), but stations affected with similar/coincident inhomogeneities with the candidate can be selected (*Stepanek* and *Mikulova*, 2008). Weighted reference series are considered more representative of the climatic region and, for being less prone to potential inhomogeneities in the neighbor series than single reference station, are more characteristic of the climate variability at smaller scale (*Della-Marta* and *Wanner*, 2006).

In this study, reference time series are used in the detection procedure, because this is the methodology adopted in MASH and ACMANT, and to assess the quality of the homogenized time series. For the reasons presented before, weighted reference series were produced with AnClim software (*Stepanek*, 2008) using difference series to evaluate the correlation coefficients as suggested in *Alexandersson* and *Molberg* (1997), *Peterson et al.* (1998), *Stepanek* and *Mikulova* (2008), and *Domonkos et al*. (2012). Since our database is affected by only a few number of missing values and the objective is to assess the quality of the homogenization process not of the data completion process, reference series were produced to present the same data gaps than the uncorrected time series. This is achieved by using uncorrected time series (with the data gaps) and neighbor time series without missing values (in order to exclude neighbor time series missing value in the reference series).

3.5. Homogenization methods performance assessment

In contrast to comparative studies performed with synthetic databases, when type, size, and location of inhomogeneities are known a priori (as in *Venema et al.*, 2012), the homogenization methods performance assessment must be executed with real data, by comparing the results obtained with different techniques using non-homogenized (hereafter NH) and homogenized data with MASH (hereafter HM) and HOMER (hereafter HH). This section is devoted to present the methodology used to assess the quality of the corrected dataset and, consequently, methods used in the homogenization process.

i. Correlation analysis

The main objective of correlation analysis is to evaluate the strength of the temporal linear relationship through the computation of the Spearman correlation coefficient, SCC (*Pereira et al.*, 2011). In this sense, to assess potential improvement in the similarity between time series before and after the homogenization process, correlation analysis was applied to annual time series to compute: (i) the correlation matrix between time series of non-homogenized and homogenized time series with MASH and with HOMER datasets; and, (ii) the SCC between each candidate and corresponding reference series. Since our objective is to assess the quality of the homogenization process, and not of the interpolation procedures used in MASH and HOMER to fill the data gaps, SCC was computed between time series with the same missing values than in NH datasets.

ii. Trend analysis

The existence of trends is in the basis of climate change studies (*Raj* and *Azeez*, 2012). In this study, the Mann-Kendal non-parametric test is used to estimate the existence, magnitude and statistical significance of potential trends in the NH, HM, and HH time series, in order to assess the impacts of homogenization methods. This test is suggested for trend analysis by the WMO (*Sneyers*, 1990) and has been used in many published works on climate change and climate variability (e.g., *Moberg* and *Jones*, 2004; *Brunetti et al.*, 2006; *Rodrigo* and *Trigo*, 2007).

iii. Principal component analysis (PCA)

When PCA is applied on a dataset, a new set of time series is produced as linear combination of the original ones. The new time series are the so-called principal components (PC), while the coefficients used to compute them are the elements of the empirical orthogonal functions (EOF). From the mathematical point of view, EOFs are the eigenvectors of the variance-covariance or the correlation matrix of the original dataset, the PCs are obtained by projecting the original time series into the EOF, and the eigenvalues are a measure of the explained variance, i.e., the proportion of the total variance explained by each PC. Obtained PCs are uncorrelated and sorted by decreasing order of variance, while EOFs are orthogonal to each other and constitute a vector base. There are different versions of this multivariate statistical technique, but it is easy to find their description/characteristics (*Jolliffe*, 2005; *Wilks*, 2011). PCA has multidisciplinary applications and is used in data analysis as an exploratory tool (for outlier detection, cluster identification, data visual examination, and interpretation), data preprocessing (dimensionality and noise reduction), modeling, and to identify spatial and temporal patterns and modes of variability such as NAO and ENSO (*Wold et al.*, 1987; *Jolliffe*, 2005; *Pozo-Vazquez et al.*, 2005). PCA results are dependent on the scaling of the original matrix (*Wold et al.*, 1987; *Jolliffe*, 2005), but statistical significance can be assessed, e.g., with cross-validation, bootstrap, or jackknifing techniques (*Romanazzi*, 1993; *Jolliffe*, 2005). PCA outputs, in particular the amount of explained variance by each PC, are dependent on the similarity of the time series (*Jolliffe*, 2005). This characteristic of PCA will be used in this study to assess homogenization results.

4. Obtained results

Preliminary exploratory statistical analysis reveals the existence of a very small number of missing values. Time series most affected by this problem present multiple consecutive missing values or their last record (end date) is before 2010. Results for maximum temperature are very similar to that for minimum temperature. The great majority of the low number of outliers detected above and below the defined thresholds based on the quartiles of their own time series was not confirmed with pairwise comparison with neighboring time series. The final number of outliers considered in HOMER for minimum and maximum temperatures were 10 and 11, respectively, which corresponds to 0.1% of total number of monthly values in each dataset or to less than 1 missing values per time series in each dataset. As mentioned in Section 3.2, MASH has an automatic procedure to detect and correct outliers which is not controlled by the user.

Temporal location and size of the breaks detected in minimum and maximum air temperature time series with MASH and HOMER are shown in *Table 2*. It should be pointed out that breaks marked with a star (*), noticeable only in the detection list of MASH, correspond to shifts of equal value but opposite sign in two consecutive years, that will most likely be an annual outlier than a break point and, from this point forward, will not be considered as breaks. Consequently, the number of breaks detected with HOMER (39 in TN and 32 in TX) is higher than with MASH (32 in TN and 24 in TX). Since the original data only have one significant decimal digit, the physical meaning of a great number of these breaks can be questioned. The number of shifts smaller than 0.1ºC detected with MASH is much higher (12 breaks in TN and 19 in TX) than with HOMER (5 breaks in TN and 1 in TX). On the other hand, the number of coincident breaks detected in TN with both methods is 18 (which corresponds to 56% and 46% of total number of breaks detected with MASH and HOMER, respectively) and 9 in TX (37% of MASH and 28% of HOMER total breaks, respectively). If the analysis is restricted to breaks with shifts greater or equal to 0.1ºC, the number of coincident breaks in TN is 14 (which corresponds to 70% and 41% of total number of breaks detected with MASH and HOMER, respectively) and 5 in TX (100% of MASH and 16% of HOMER total number of detected breaks, respectively). These results suggest that MASH could be able to detect smaller shifts but an overall small number of break points.

Table 2. Location (and magnitude) of break points detected on minimum and maximum air temperature during 1941-2010 period, with MASH and HOMER. Coincident detections with both methods, defined with utmost 18 months apart are presented in bold. Detections in two consecutive years with symmetrical shifts are marked with a star (*)

ID	Minimum air temperature (TN) MASH	Minimum air temperature (TN) HOMER	Maximum air temperature (TX) MASH	Maximum air temperature (TX) HOMER
1	**1963**(0.11), 1996(0.12)	1944(-0.08), 1950(0.04), **1964**(-0.45), 1970(0.49), 1984(0.29)	1944*(0.12)	1977(-0.22)
2	-	1963(-0.37), 1987(0.24) 1992 (0.53)	**1949**(0.14)	**1950**(-0.76), 1959(-0.49) 1971(-0.27), 1981(0.23)
3	1947*(0.17), 1972(0.08), **1980**(0.14)	1962(0.27), **1980**(-0.49)	1962(-0.05), 1969(-0.14), 1977(-0.03)	1965 (0.27) 1972(0.39) 1993 (0.45)
4	**1979**(-0.31)	1966(-0,18), **1979**(0,78)	1943*(0.46), 1949(-0.03), 1961(0.06), 1963(0.01), 1988*(-0.03), 2000*(0.14)	1953(-0.26), 1992(0.42)
5	**1982**(-0.04)	1950(-0.41), 1967(0.19) **1982**(0.14)	1969(-0.06), **1971**(0.15)	1949(-0,35), **1971**(-0,51)
6	1965(0.15), **1971**(0.15), 1976 (0.09), **1979**(0.26), 1985*(0.39), **1993**(0.05), 1996 (-0.07)	**1969**(-0.75), **1980**(-1.21), 1987(1.26), **1994** (-0.20)	1949*(0.09), 1986*(-0.16)	-
7	1951(-0.28), **1966**(0.21)	**1967**(-0.60), 1989(0.31), 1998(-1.54)	-	-
8	-	1950(-0.67)	**1953**(-0.14), 1976(-0.02), 1979(-0.04), 1994*(-0.18), 1996*(-0.16), 1998*(0.08)	**1951**(0.41), 1974(0.25) 19992(0.35)
9	1963*(-0.128)	-	-	1972(0.12), 1988(0.40)
10	**2003**(-0.54), 2007(0.16)	1958(-0.33) **2004**(1.11)	**1953**(0.04), 1965*(0.06), 1977*(0.12), 1995(0.06), **1998**(-0.27), 2000(-0.12), 2003(0.09)	**1951**(-0.01), 1974(-0.59), 1991(-0.39), **1996**(1.14)
11	-	1986(0.12), 1990(0.32)	-	1951(0.40), 1955(0.19), 1973(-0.42), 1990(0.41)
12	1968(0.11), **1977**(-0.05), 1980(-0.17), **1984**(-0.12), **1986***(-0.12), 1996(-0.03)	**1978**(0.56), **1984**(-0.03), **1987**(0.80), 2000 (0.28)	-	1995 (0,52)
13	**1943***(-0.28), 1948(-0.06), 1966 (0.07), **1974**(0.12), **1986**(-0.09)	**1944**(0.81), 1946(0.45), **1973**(-0.49), **1986** (0.04), 1993 (0.04)	**1954**(0.09), 2000(0.15)	**1953**(-1.22), 1959(0.25), 1991(-0.41)
14	1955(-0.21), **1958**(-0.42), 1969 (0.11), **1982**(-0.58), **1994**(0.80), 1996(-0.08), 1999(-0.08)	**1957**(0.88), **1982**(0.27), **1994**(0.90)	1950(0.05), **1978**(-0.08), **1981**(-0.06), 1992*(0.26), **1994** (0.45)	**1977**(0.29), **1982**(0.35), **1994**(-1.70)

Correlation matrices between non-homogenized time series of maximum (minimum) air temperature, TXNH (TNNH), as well as between homogenized time series with MASH, TXHM (TNHM) and with HOMER, TXHH (TNHH) were computed. Boxplots of the Spearman correlation coefficient (SCC) values obtained for homogenized time series with HOMER are higher having lower dispersion in relation to non-homogenized and homogenized with MASH (*Fig. 2*). In general, SCC values between homogenized TX and TN time series is higher than those obtained between non-homogenized times series. Median value of the difference between TXHH and TXNH correlation matrix is higher (0.09, which corresponds to an increase of 9%) than the difference between TXHM and TXNH correlation matrix (0.02, which corresponds to a general increase of 2%). For minimum temperature, the median of the difference between TNHH and TNNH is equal to 0.13, while between TNHM and TNNH it is equal to 0.05.

Spearman correlation coefficient values obtained between reference series and non-homogenized and homogenized with MASH and HOMER corresponding time series (*Fig. 3*) reveals: (i) higher SCC values between reference and homogenized time series with MASH in every stations and for both TX and TN than between reference and non-homogenized time series; (ii) higher SCC values between reference and homogenized time series with HOMER for TX than between reference and non-homogenized time series but lower values for TN in 6 weather stations. Median of the SCC values obtained for maximum and minimum air temperature homogenized time series with HOMER are similar (94.3% and 89.3%) to those obtained with MASH (91.8% and 88.8%) but higher than for non-homogenized time series (88.2% and 86.4%), in particular for maximum air temperature. At this respect, the increase in the SCC can be underlined computed between the reference and one of the corresponding series: (i) TXHM and TXHH time series in Vila Real and Viseu (of 14.7% and 13.0%, respectively); and, (ii) TNHH and TNHM time series in Vila Real (of 13.7% and 8.4%, respectively).

Trend analysis for TN performed with Mann-Kendal test assuming a statistical significance level of 99% (Table 3) reveals that: (i) only a small number of non-homogenized times series presents statistically significant trends (5 in TNNH and TNHM datasets and only 1 in TNHH dataset); (ii) almost all time series present positive trends except Mirandela and Dunas de Mira; (iii) a reduction in the number of statistical significant trends is only verified for TNHH dataset; and, (iv) with the homogenization procedures, the trend of two time series, after being homogenized, became statistically significant (time series of Bragança, with MASH and of Montalegre with HOMER). Results obtained for TX shows that: (i) there is a lower number of statistically significant trends (2 in TXNH and only 1 in TXHM); (ii) the number of non-homogenized and homogenized time series with negative and positive trends are similar; but, (iii) all statistically significant trends are positive; and, (iv) homogenization procedures lead the loss of statistical significance of the trends in one homogenized time series with MASH and in two homogenized time series with HOMER.

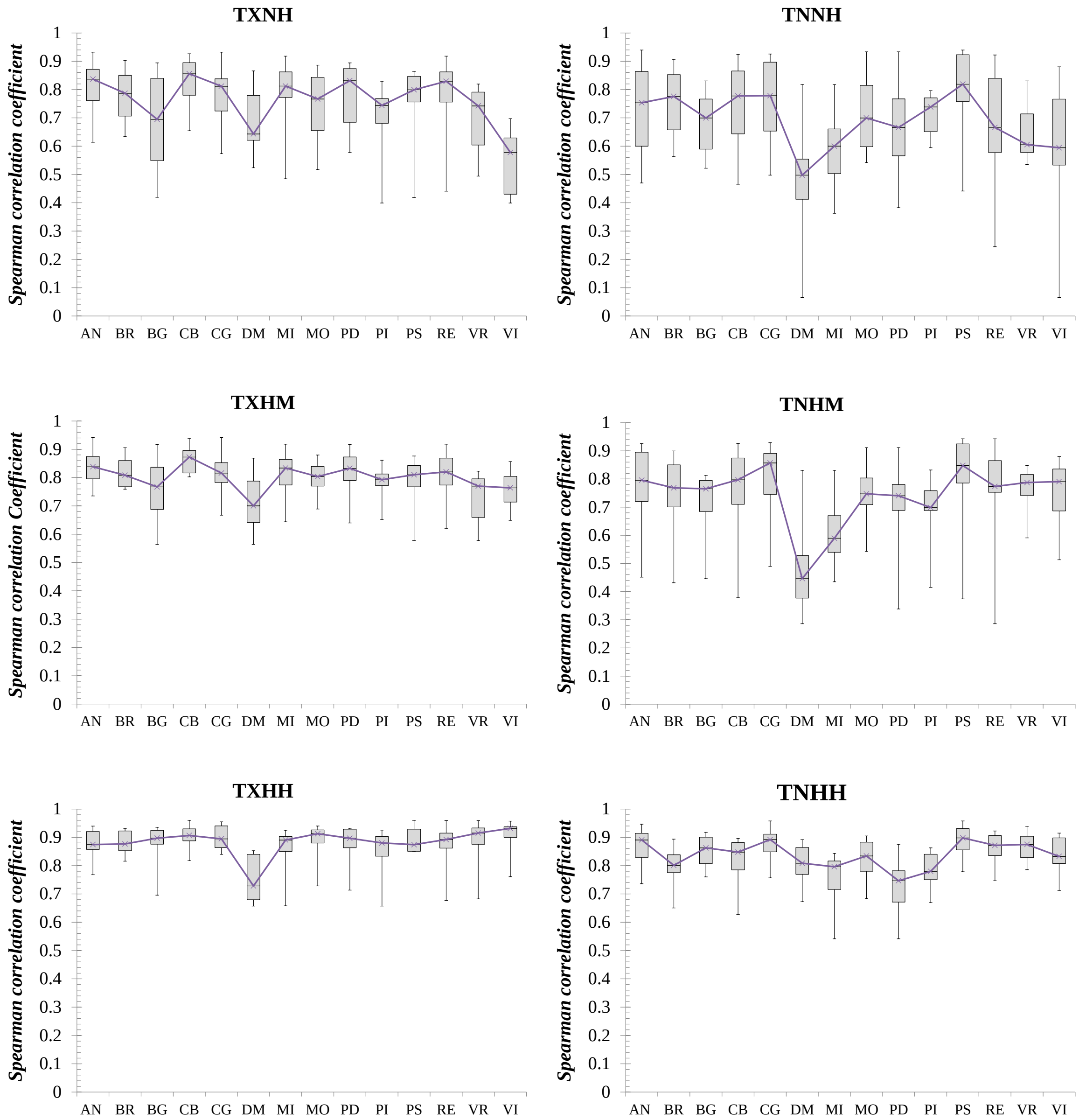

Fig. 2. Boxplot of Spearman correlation coefficient (SCC) between annual time series of non-homogenized (NH), homogenized with MASH (HM) and with HOMER (HH) maximum (top panel) and minimum air temperatures (bottom panel), from weather stations located in northern part of the continental Portugal (*Table 1* and *Fig. 1*), for 1941–2010 period. SCC was evaluated taking into account missing values of NH time series. The bottom/top indicates the lower/upper quartiles, and the band near the middle of the box is the median. The lower/upper end of the whiskers represents the minimum/maximum values.

Results obtained with PCA performed on non-homogenized and homogenized datasets (Table 3) can be summarized as follows: (i) only a small number of PCs are statistical significances (1 PC for TXHM, TXHH, and TNHH and 2 PCs for TNNH, TXNH, and TNHM); (ii) the explained variance by the first PC of homogenized datasets is greater than the explained variance by the first PC of non-homogenized ones; (iii) explained variance of first PC are higher for homogenized datasets with HOMER than with MASH.

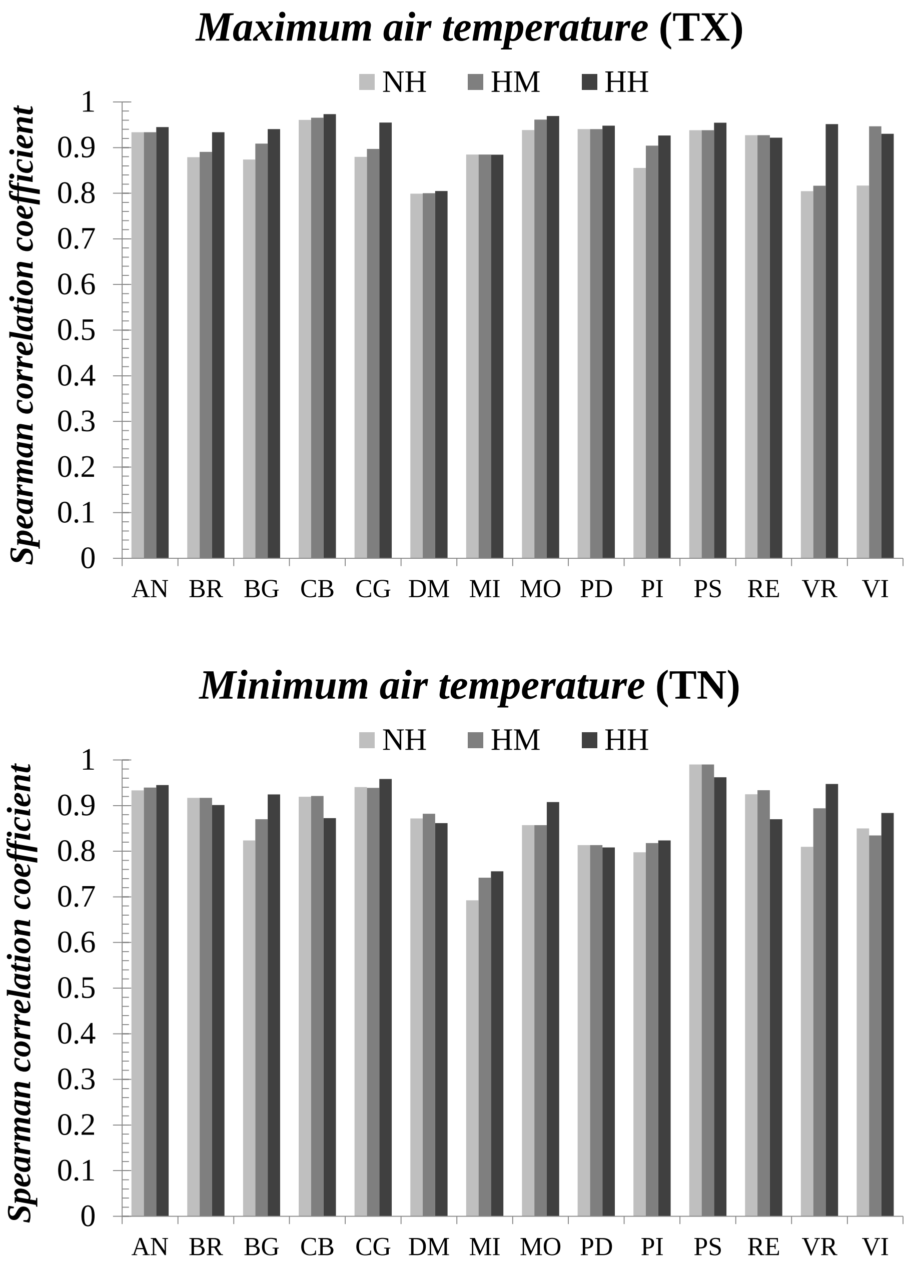

Fig. 3. Spearman correlation coefficient (SCC) between annual reference series and time series of non-homogenized (NH), homogenized with MASH (HM) and HOMER (HH) of maximum air temperature (top panel) and minimum air temperature (bottom panel), for the 1941–2010 period.

5. Discussion and conclusions

The IM network analyzed here includes stations located near the coast and a few meters above sea level and inland stations at higher altitude. Moreover, all weather stations are located in the same climatic region (temperate with dry and hot summer), which is a necessary condition to perform homogenization analysis. Results of the preliminary exploratory data analysis reveals time series with no extremes and only a small amount of outliers and missing values except in cases where times series does not cover the entire analysis period of 1941–2010. This is an important characteristic of the dataset, because missing values can have profound impact on reference series and, consequently, in the detection procedures (*Menne* and *Williams*, 2005; *Syrakova* and *Stefanova*, 2009). In addition, since missing values are treated differently in MASH and HOMER, a small number of data gaps cannot be associated with potential significant differences between homogenized datasets with both methods. On the other hand, heterogeneities are to be expected in TX and TN datasets, since this network is in operation for a long time, and during this period experienced adjustments were carried out on its structure (e.g., replacement of instruments), on its type (changes from classical to automatic sensors), and spatial distribution (e.g., relocation, cessation, and installation of new stations). For these reasons, we may conclude that maximum and minimum air temperature datasets in northern Portugal are examples of databases in good position to be analyzed for homogeneity.

MASH and HOMER were the methods used to perform the homogeneity analysis of TX and TN datasets. The selection criterion was, primarily, the high performance shown by these two methods during the comparison study performed in the framework of the COST Action HOME, using monthly temperature benchmark databases but also the large methodological differences between these two methods, discussed in previous sections. In fact, HOMER was not compared with other methods in *Venema et al.* (2012), because it became available later, but its results from the combination of the methods had the best performance. Craddock method was also included in the list of algorithms with best performance, but because it is a subjective method (uses visual detection of breaks), was not used in this study.

Time series were corrected with both methods from the most recent observations to the oldest. This procedure is consistent with the general believe that current sensors and data acquisition systems are more reliable than previous ones. Both methods uses interpolation to produce homogenized time series without missing values, but MASH also uses extrapolation to fill the data gaps in the extremes of the time series. MASH identifies the location of the break with the year of the shift, while HOMER is able to estimate the month of the change also (not shown in *Table 2*).

The total number of breaks detected in TN with both methods is higher than in TX, and the number of breaks detected with HOMER is higher than with MASH, in both climatic elements. The same conclusion is supported by considering the number of breaks with amplitudes above increasing thresholds. In addition, the amplitude of the breaks detected with HOMER is, in general, higher than the amplitude of the breaks detected with MASH (*Table 2*). The weather stations of Vila Real and Coimbra B were selected as examples of inland and coastal weather stations, located at higher and lower altitudes, respectively (*Fig. 4* and *Fig. 5*), to illustrate the differences between the non-homogenized and homogenized time series with MASH and HOMER. It should also be mentioned that maximum air temperature time series in Mirandela is the only one without inhomogeneities.

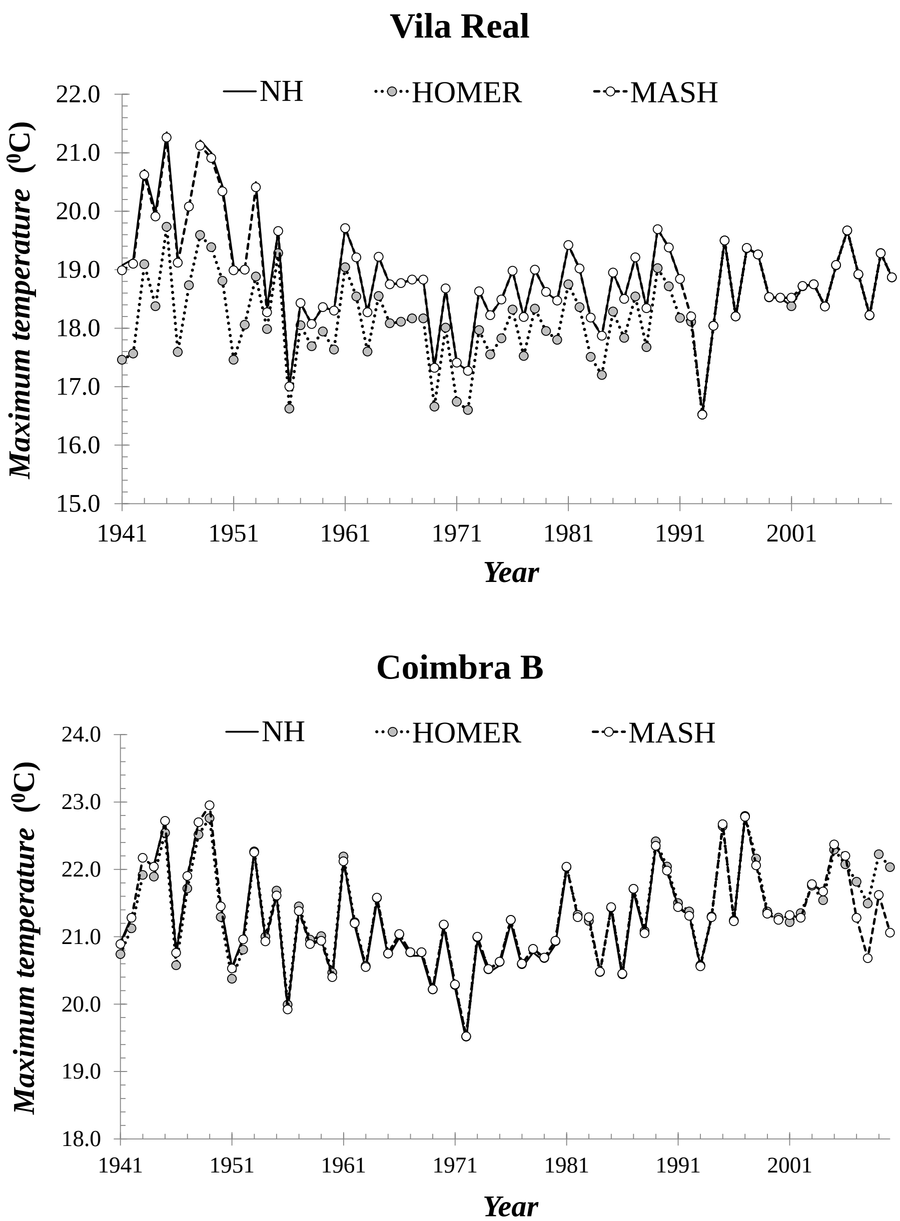

Fig. 4. Non-homogenized (NH) and homogenized time series (with HOMER and MASH) of maximum air temperature (TX) recorded in Vila Real and Coimbra B weather stations. Coimbra B is an example of weather station located in near the coast at low altitude, while Vila Real is an example of weather station located at mountainous region of the interior.

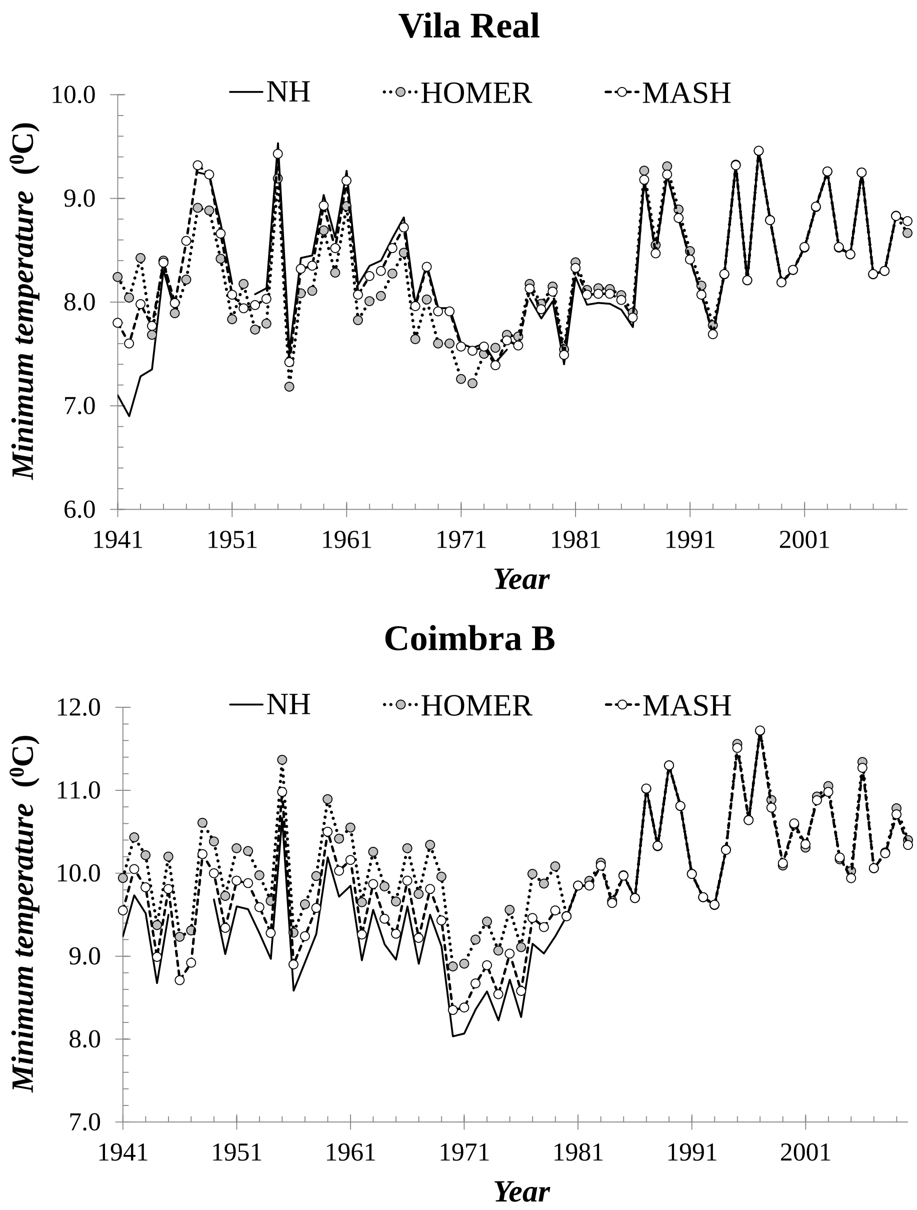

Fig. 5. As in *Fig. 4*, but for minimum air temperature (TN).

Correlation, trend, and principal component analysis were used to assess the homogenization process performance by comparing the results obtained with using non-homogenized and homogenized datasets. Boxplots of the Spearman correlation coefficient (SCC) statistical values obtained between homogenized time series with HOMER are higher and have much lower dispersion than those obtained between homogenized with MASH and non-homogenized time series (*Fig. 2*). For maximum air temperature, the homogenized time series of Dunas de Mira weather station (with both methods) presents the lowest values of the statistics, and it is responsible for the high dimension of the lower whisker. This result is more perceptible in homogenized time series with HOMER than with

MASH as boxplots for other time series are much more alike. For TN, the dispersion is much higher than for TX, and Dunas de Mira time series is also among those presenting lower statistics values. Results obtained with correlation analysis between reference, non-homogenized, and homogenized time series are also consistent with the increase of the similarity of the datasets with the application of both homogenization procedures. Values of SCC increases for all homogenized time series with MASH, but for a few time series (both in TX and TN), SCC values obtained for homogenized time series with HOMER are smaller than for non-homogenized. Notwithstanding this fact, an overall small increase in the median SCC values is conspicuous.

Trend analysis performed on TX and TN time series reveals a small reduction in the number of statistically significant tendencies after homogenization, but a general decrease in the slope, more significant for homogenized time series with HOMER than with MASH, must be underlined. Results obtained using different statistical significance levels (97.5% and 95%) are similar except for the expected higher number of statistically significant trends.

Results from PCA are consistent with those previously obtained with other methodologies and also suggests that homogenization leads to an increase of the resemblance in the spatial and temporal variability of both TN and TX. This behavior is more evident for TX than for TN. In general, the first EOF presents elements with equal sign, which reflects similar behavior in the entire region. Then, the following EOF represents small scale features of variability (e.g., contrast between north and south or between east and west). The magnitude of each feature can be measured by the explained variance of the corresponding mode of variability. In this study, the explained variance by the first PC is higher for homogenized than for non-homogenized datasets, independently of the climatic element (*Table 3*). This difference is higher for homogenized time series with HOMER than with MASH.

Table 3. Explained variance of the statistically significant principal components of non-homogenized (NH), homogenized with MASH (HM) and with HOMER (HH) minimum and maximum air temperature datasets, for the 1941–2010 period

	N	**NH**	**HM**	**HH**
Maximum temperature	1	78.6%	81.0%	89.4%
	2	8.6%	–	–
Minimum temperature	1	75.2%	76.3%	84.8%
	2	9.7%	8.3%	–

In resume, the most important conclusion from this study is that both methods contribute to correcting the inhomogeneities detected in both TN and TM datasets, and that there is no clear evidence of the better performance of one method relative to one another. Results obtained from the correlation analysis, trend analysis and principal component analysis point to a general increase on the spatial and temporal similarity of the time series as should be expected in datasets of the same climatic region. Apparently, these results are independent of the location and altitude of the weather stations. However, these conclusions should be taken with caution, because earlier studies reveal that the evaluation of methods performance is dependent on the metrics used for this purpose (*Venema et al.*, 2012) and on the quality and characteristics of databases (*Freitas et al.*, 2012).

Finally, it should be noted that, to the best of our knowledge, this study is the first effort to compare HOMER with other homogenization methods using observed datasets. The other known study assessing HOMER performance was recently presented in the 7th seminar for homogenization and quality control on climatological databases, but using the HOME benchmark datasets (*Domonkos*, 2012). Furthermore, besides the study of *Freitas et al.* (2012), to assess HOMER potential and limitations, this is the first consistent attempt to homogenize maximum and minimum air temperature Portuguese datasets, using more than one method, MASH and HOMER in particular. The other known homogenization study performed with Portuguese data, was performed by *Soares* and *Costa* (2009), which used precipitation data from stations located in the southern part of the country, as a case study, to compare the potential advantages of geostatistical techniques. As a final point, it should also be emphasized the number of different methods/measures used to compare the homogenized Portuguese air temperature datasets.

References

AEMET, IM, 2011: Iberian Climate Atlas. Closas-Orcoyen S.L, Madrid.

Aguilar E, Auer I, Brunet M, Peterson T.C, Wieringa J., 2003. *Guidelines on Climate Metadata and Homogenization*. World Meteorological Organization: *WMO–TD No. 1186, WCDMP No. 53*, Switzerland, Geneva.

Alexandersson, A., 1986: A homogeneity test applied to precipitation data. *J. Climatol., 6*, 661–675.

Alexandersson, H. and *Moberg, A.*, 1997: Homogenization of Swedish temperature data. 1. Homogeneity test for linear trends. *Int. J. Climatol. 17*, 25–34.

Brunetti M., Maugeri M., Monti F., Nanni T., 2006: Temperature and precipitation variability in Italy in the last two centuries from homogenized time series. *Int. J. Climatol. 26, 345–381* .

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

Caussinus, H. and *O. Mestre*, 2004: Detection and correction of artificial shifts in climate series. *Appl. Statist. 53*, 405–425.

Craddock, J.M., 1979: Methods of comparing annual rainfall records for climatic purposes. *Weather 34*, 332–346.

Della-Marta, P.M. and *Wanner, H.*, 2006: A Method of Homogenizing the Extremes and Mean of Daily Temperature Measurements. *J. Climate 19*, 4179–4197.

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

Domonkos, P., Venema, V., Auer, I., Mestre, O., and *Brunetti, M.*, 2012: The historical pathway towards more accurate homogenisation. *Adv. Sci. Res. 8, 45–52*

Ducré-Robitaille, J-F., Vincent, L.A., and *G. Boulet*, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol. 23*, 1087–1101.

Freitas,L., Pereira M.G., Amorim L., Caramelo L., Mendes M., and *Nunes F.L*, 2012: Portuguese temperature dataset homogeneity with HOMER *(submitted to WMO serial World Climate Data and Monitoring Program)*

Guijarro, J.A., 2011: Climatol version 2.0, an R contributed package for homogenization of climatological series. State Meteorological Agency, Balearic Islands Office, Spain, http://webs.ono.com/climatol/climatol.html

Gullet, D.W., Vincent, L., and *Sajecki, P.J.F.*, 1990: Testing for Homogeneity in Temperature Time Series at Canadian Climate Stations. CCC Report No. 90-4, *Atmospheric Environment Service, Downsview, Ontario*.

HOME, 2011: Homepage of the COST Action ES0601 - *Advances in Homogenisation Methods of Climate Series: an Integrated Approach* (HOME), http://www.homogenisation.org.

Jolliffe, I.T , 2005:Principal component analysis. In (B.S. Everitt and D.C. Howell) *Encyclopedia of Statistics in Behavioral Science, vol. 3,* Wiley, New York, 1580–1584.

Kuglitsch, F.G., Toreti, A., Xoplaki, E., Della-Marta, P.M., Luterbacher, J., and *Wanner, H.*, 2009: Homogenization of daily maximum temperature series in the Mediterranean, *JGR 114*, 1–6.

Menne, M.J. and *Williams Jr., C.N.*, 2009: Homogenization of temperature series via pairwise comparisons. *J. Climate 22*, 1700–1717.

Menne, M.J. and *Williams jr. C.N.*, 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate 18*, 4271–4286.

Mestre O., 1999. Step-by-step procedures for choosing a model with change-points. In *Proceedings of the second seminar for homogenisation of surface climatological data*, Budapest, Hungary, *WCDMP-No.41, WMO-TD No.962*, 15–26.

Mestre, O. and *Aguilar E.*, 2011: Homogenization (personal communication). COST ES0601 Training School, 5–7th of October, Tarragona.

Moberg, A. and *Jones, P.D.*, 2004: Regional Climate Models simulations of daily maximum and minimum near-surface temperatures across Europe compared with observed station data for 1961-1990. *Clim. Dynamics 23*, 695–715.

Pereira, M.G., Caramelo, L., Gouveia, C., Gomes-Laranjo, J., and *Magalhaes, M.* 2011: Assessment of weather-related risk on chestnut productivity. *NHESS*, 2729–2739.

Perreault L., Bernier J., Bobée B., and *Parent E.*, 2000. Bayesian change-point analysis in hydrometeorological time series. Part. 1. The Normal model revisited. *J. Hydrol. 235*, 221–241.

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

Peterson T.C., Easterling, D.R., Karl, T.R., Groisman, P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Førland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P. and *Parker, D.*,1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol. 18*, 1493–1517.

Peterson, T.C., Easterling, D.R., Karl, T.R., Groismanm P., Nicholls, N., Plummer, N., Torok, S., Auer, I., Boehm, R., Gullett, D., Vincent, L., Heino, R., Tuomenvirta, H., Mestre, O., Szentimrey, T., Salinger, J., Forland, E.J., Hanssen-Bauer, I., Alexandersson, H., Jones, P., and *Parker, D.*, 1998: Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol. 18*, 1493–1517.

Picard, F., Lebarbier, E., Hoebeke, M., Rigaill, G., Thiam B., and *Robin, S.*, 2011: Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics.* doi:10.1093/biostatistics/kxq076

Pozo-Vazquez, D., Gamiz-Fortis, S.R., Tovar-Pescador, J., Esteban- Parra, M.J., Castro-Dıez, Y., 2005: North Atlantic winter SLP anomalies based on the autumn ENSO state. *J. Climatol. 18*, 97–103.

Raj, P.P.N. and *Azeez, P.A.*, 2012: Trend analysis of rainfall in Bharathapuzha River basin, Kerala, India. *Int. J. Climatol. 32*, 533–539.

Reeves, J., Chen, J., Wang, X.L., Lund, R., and *Lu, Q.*, 2007: A review and comparison of changepoint detection techniques for climate data. *J. Appl. Meteorol. Climatol. 46*, 900–915.

Rodrigo, F.S. and *Trigo, R.M.*, 2007: Trends in daily rainfall in the Iberian Peninsula from 1951 to 2002. *Int. J. Climatol. 27*, 513–529.

Romanazzi, M., 1993: Jackknife estimation of the eigenvalues of the covariance matrix. *Comput. Statist. Data Anal. 15*, 179–198.

Sahin, S. and *Cigizoglu, H.K.*, 2010: Homogeneity analysis of Turkish meteorological data set. *Hydrol. Process. 24*, 981–992.

Sneyers, R., 1990: On the statistical analysis of series of observations., WMO *Tech. Note 143, 145*, Geneva, Switzerland.

Staudt, M., Esteban-Parra, M.J., and *Castrı-Diez, Y.*, 2007: Homogenization of long-term monthly Spanish temperature data. *Int. J. Climatol. 27*, 1809–1823.

Štěpánek, P., 2008: AnClim – software for time series analysis (for windows). *Dept. of Geography, Fac. of Natural Sciences*, MU, Brno. http://www.climahom.eu/AnClim.html.

Štěpánek, P. and *Mikulová, K.*, 2008: Homogenization of air temperature and relative humidity monthly means of individual observation hours in the area of the Czech and Slovak Republic. In: *5th Seminar for Homogenization and Quality Control in Climatological Databases*. Hungarian Met. Service, Budapest, Hungary, 147−163.

Štěpánek, P., Zahradniéek, P., and *Skalák, P.*, 2009: Data quality control and homogenization of the air temperature and precipitation series in the Czech Republic in the period 1961–2007, *Adv. Sci. Res. 3*, 23–26.

Syrakova, M. and *Stefanova, M.*, 2009: Homogenization of Bulgarian temperature series. *Int. J. Climatol. 29*, 1835–1849.

Szentimrey, T., 1994: Statistical problems connected with the homogenization of climatic time series. In *Proceedings of the European Workshop on Climate Variations*, Kirkkonummi, Finland, *Publications of the Academy of Finland, 3/94*, 330–339.

Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH). *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, WMO, *WCDMP-No. 41*, 27–46.

Szentimrey, T., 2007: Manual of homogenization software MASHv3.02”, Hungarian Meteorological Service.

Toreti, F.G., Kuglitsch, A., Xoplaki, E., and. Luterbacher, J., 2012: A Novel Approach for the Detection of Inhomogeneities Affecting Climate Time Series. *J. Appl. Meteorol. Climatol.,vol. 51*, 317–326.

Trewin, B., 2010: Exposure, instrumentation, and observing practice effects on land temperature measurements. *WIREs Clim. Change, 1*, 490–506.

Venema, V., 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., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Duran, M.P., Likso, T., Esteban, P., and Brandsma, T., 2012: Benchmarking monthly homogenization algorithms. *Climate of the Past 8*, 89–115.

Vicente-Serrano, S., Beguería, S., Lopez-Moreno, J.I., García-Verac, M.A., and *Stepanek, P.*, 2010: A complete daily precipitation database for northeast Spain: reconstruction, quality control, and homogeneity. *Int. J. Climatol. 30*, 1146–1163.

Vincent, L.A., 1998: A technique for the indentification of inhomogeneities in Canadian temperature series. *J. Clim. 11*, 1094–1104.

Wilks, D.S., 2011. Statistical Methods in the Atmospheric Sciences, vol 100, 3rd Edition,*International Geophysics 100*. Academic Press, Oxford, UK

Williams, C.N., Menne, M.J., and *Thorne, P.W.*, 2012:. Benchmarking the performance of pairwise homogenization of surface temperatures in the United States. *J. Geophys. Res.-Atmos. 117, D05116*.

Wold, S., Geladi, P., Esbesen, K. and *Ohman, J.*, 1987: Multi-way principal components- and PLS-analysis. *J. Chemometrics 1*, 41–56.

World Meteorological Organization, 2011: Guide to Climatological Practices, WMO/No. 100., Geneva.

Zhang, X.B., Vincent, L.A., Hogg, W.D., and *Niitsoo, A.*, 2001: Temperature and precipitation trends in Canada during the 20th century, *Atmos. Ocean 38,* 395–429.