gsimcli: a geostatistical procedure for the homogenisation of climatic time series

Sara Ribeiro[a;] , Júlio Caineta[a] , Ana Cristina Costa[a] · Roberto Henriques[a]

[a]NOVA IMS,Universidade Nova de Lisboa

Address: Campus de Campolide, 1070-312, Lisboa, Portugal

gsimcli: a geostatistical procedure for the homogenisation of climatic time series

Running head:

gsimcli: a geostatistical procedure for homogenisation

Authors:

Sara Ribeiro[a]

sribeiro@novaims.unl.pt

Júlio Caineta[a]

jcaineta@novaims.unl.pt

Ana Cristina Costa[a]

ccosta@novaims.unl.pt

Roberto Henriques[a]

roberto@novaims.unl.pt

[a]NOVA IMS,Universidade Nova de Lisboa

Address: Campus de Campolide, 1070-312, Lisboa, Portugal

Corresponding author:

Sara Ribeiro

NOVA IMS,Universidade Nova de Lisboa

Address: Campus de Campolide, 1070-312, Lisboa, Portugal

Phone: +351 213 828 610

E-mail: sribeiro@novaims.unl.pt

gsimcli: a geostatistical procedure for the homogenisation of climatic time series

Abstract

Climate data homogenisation is of major importance in monitoring climate change and in validating weather forecasts, general circulation and regional atmospheric models, modelling of erosion and drought monitoring, among other impact studies. Discontinuities in the time series, also named inhomogeneities, may lead to biased conclusions in such studies, so they should be detected and corrected. Previous studies have suggested a geostatistical stochastic approach, which uses Direct Sequential Simulation (DSS), as a promising methodology for the homogenisation of precipitation data series. Based on the spatial and temporal correlation between the neighbouring stations, DSS calculates local probability density functions at a candidate station to detect inhomogeneities. Here, we present a new method named gsimcli (Geostatistical SIMulation for the homogenisation of CLImate data), which is an improved and extended version of that approach. This technique is novel in its incorporation of spatial correlation metrics for the homogenisation of climate time series. The method's performance is assessed with annual and monthly precipitation, and monthly temperature data from two regions of the COST-HOME benchmark data set, and the results are compared using performance metrics. We also evaluate a semi-automatic version of the gsimcli method, which performs additional adjustments for sudden shifts. Both gsimcli versions provided similar results in the homogenisation of annual series. The gsimcli method was more efficient in the homogenisation of the benchmark's precipitation series than the original geostatistical approach. The gsimcli approach performed more closely to state-of-the-art procedures in the homogenisation of monthly data than in the homogenisation of annual data. We expect that the proposed procedure will open new perspectives for the development of techniques that detect and correct inhomogeneities in climate data with monthly and sub-monthly resolution.

Keywords

Climate data; data quality; benchmark; geostatistics; homogenization; precipitation; temperature.

1 Introduction

Climatic time series may be affected by non-natural irregularities caused by sudden or gradual changes on the surrounding environment of the weather station, or changes in the process of measurement and recording of the climate variable (e.g., Aguilar et al., 2003; Brunet and Jones, 2011; Trewin, 2010). Station relocations, repositioning at different heights and changes in the instrumentation are examples of the former. Gradual changes may be exemplified by slowly urban development around a weather station, contributing to the phenomenon known as urban heat island effect (Sahin and Cigizoglu, 2010). The presence of inhomogeneities can distort or even hide the true climatic signal, and thus bias the results of studies (e.g., Domonkos, 2013; Yozgatligil and Yazici, 2016). Several homogenisation methods have been developed in the last decades to detect inhomogeneities and to adjust the climatic time series in order to improve their temporal consistency (Domonkos et al., 2012; Ribeiro et al., 2015a). Homogenisation methods depend on the climate variable (temperature, precipitation, pressure, evaporation), on the temporal resolution of the observations (annual, seasonal, monthly or daily), on the availability of information on the history of the weather station, and on the spatial density of monitoring stations within the study area (Costa and Soares, 2009). Ribeiro et al. (2015a) classified the homogenisation methods according to their characteristics: non-parametric tests, classical tests, regression methods, Bayesian approaches, and procedures specifically proposed for the homogenisation of climate data series. Those authors also describe comparison studies that evaluated the efficiency of homogenisation methods, and summarise many methods applications. Domonkos et al. (2012) present a chronological review of the theoretical properties of the most relevant statistical tools that have been developed for the homogenisation of temperature series. Aguilar et al. (2003) and the World Meteorological Organization (2010) emphasise the importance of metadata in the homogenisation of climate time series. By using all the available metadata and stations' history, it is possible to anticipate and preview the type of problems that climate data may have and when they should appear. Since this is often unattainable, it is advisable to compare the stations' history with the data analysis, in a double check process.

Homogenisation approaches can be classified as absolute and relative. Absolute methods only consider the climatic time series of the station to be homogenised (candidate station), while relative homogenisation uses time series from neighbouring stations. Absolute homogenisation may be problematic, because it is difficult to determine if changes, or lack of changes, result from non-climatic or climatic influences without the support of the station's history information (Peterson et al., 1998). Absolute approaches are not recommended as they can even introduce more errors into the climate series (Begert et al., 2005; Guijarro, 2011; Venema et al., 2012). Relative homogenisation is preferred when the spatial density and coherence of the observed data allows it (Costa and Soares, 2009; Domonkos, 2013; Ribeiro et al., 2015a). Relative homogenisation relies on comparing the candidate time series to multiple reference series from surrounding stations in a pairwise fashion, or to a single composite reference series computed from multiple neighbouring stations (Venema et al., 2012). More specifically, time series comparisons can rely either on building one composite reference series for each candidate series, on using multiple reference comparisons for each candidate series, or on using multiple comparisons without defining which are the candidate and the reference series (Domonkos, 2013). Composite reference series are usually built as a weighted average of data from surrounding stations by using some measure of statistical similarity between them (Aguilar et al., 2003). The comparison series are computed as the difference (in case of temperature, pressure, etc.) or ratio (precipitation, wind, etc.) between the candidate and the reference. The comparison series are statistically tested, or a penalised likelihood criteria is used, to assess the significance of changes. Homogenisation corrections may be estimated directly from the comparison series as follows (Aguilar et al., 2003). If a series must be adjusted for a sudden shift, a common approach is to calculate separate averages on the comparison series for the two sections defined by the breakpoint. Then, the obtained means are compared by calculating their ratio or their difference, depending on the variable, and the resulting factor is then applied to the inhomogeneous part. When gradual inhomogeneities are detected, the usual approach is to de-trend the inhomogeneous section using the slope calculated on the ratio time series. When multiple references or pairwise estimates are available, a combination of those estimates is used (e.g., a mean or median). A different approach based on multiple reference series is used by MASH – Multiple Analysis of Series for Homogenisation (Szentimrey, 1999), which considers the adjustment-factors as the lower limits of confidence intervals to keep a low false alarm detection rate (Domonkos, 2013). Once a first correction has been performed, most methods perform a review (Venema et al., 2012).

Aguilar et al. (2003) recommend the adoption of a reverse chronological approach to adjust annual (monthly) series experiencing more than one discontinuity, in which the most recent homogeneous period is used as a standard and earlier periods are adjusted to reflect these current conditions. By doing so, incoming data in the future will still be homogeneous unless further changes occur in the monitoring station. Moreover, even if additional changes take place, another advantage of this strategy is that it allows for easier updating (Auer et al., 2005). Allen and DeGaetano (2000) argue that it is also reasonable to base adjustments on the longest stationary homogeneous period within a station's record, and then proceed chronologically, but with the decision to adjust earlier or more recent periods again based on the series length. One advantage of this approach is that the quantity of data that is subject to adjustment is minimised.

The selection of the homogenisation procedure is an effortful task. Domonkos (2015) refers three reasons for the complexity of the selection of the homogenisation procedure: first, the applicability of the method highly depends on the properties and the spatial and temporal structure of the climatic records to be homogenised; second, the efficiency of the homogenisation can be measured empirically only with synthetic test data sets, even though the observed efficiency might differ from the true efficiency due to the deviations in the test data set from the real data; and, third, metadata sometimes provide more reliable information than statistical tests. In 2008, the HOME project (COST Action ES0601) gathered a group of climate experts in order to compare, evaluate and develop homogenisation methods using a benchmark dataset of temperature and precipitation series (Venema et al., 2012). To create the COST-HOME benchmark datasets, known inhomogeneities and other data disturbances were inserted. Under this project, 25 contributions based on 13 statistical homogenisation algorithms were submitted before the release of the list of known/inserted inhomogeneities (the “truth”) in the data sets (blind contributions), and their results were evaluated with performance metrics. Later, some of the blind contributions were improved to address problems revealed by the results. One of the main conclusions of the HOME project is that the most efficient methods are those that deal with inhomogeneous neighbouring series, as well as with the interactions of multiple breakpoints and their effects on the calculation of correction terms, namely ACMANT (Domonkos et al., 2011a), MASH (Szentimrey, 1999, 2007, 2008), PRODIGE (Caussinus and Mestre, 1996, 2004) and USHCN (Menne and Williams Jr., 2009; Menne et al., 2009). According to Domonkos et al. (2012), these procedures provide the reconstruction and preservation of true climatic variability in observational time series with the highest reliability. Although more limited regarding some tasks, the Craddock method (Brunetti et al., 2006; Craddock, 1979) also had an excellent performance and it is recommended for practical use (Domonkos, 2013; Venema et al., 2012).

Several methods proposed in the literature have been developed as software packages, which intend to reduce the time consumed during the homogenisation process and to minimise the users’ interaction. Ribeiro et al. (2015a) describe their main characteristics, namely of ACMANT and its units ACMANT2 (Domonkos, 2015), Climatol (Guijarro, 2006), RHTest (Wang, 2008), AnClim and ProClimDB (Štepánek, 2008a, 2008b), and HOMER (Mestre et al., 2013). More recently, the ACMANT3 unit has been released (Domonkos and Coll, 2016). Some of the methods recommended by the HOME project are available in HOMER for monthly data, and HOM/SPLIDHOM for daily data (Mestre et al., 2011).

This article presents the gsimcli method, which is an extension of the geostatistical approach proposed by Costa and Soares (2009) and Costa et al. (2008a). Costa et al. (2008a) proposed to use the DSS – Direct Sequential Simulation algorithm (Soares, 2001) to calculate the local probability density function (pdf) at a candidate station's location. The DSS algorithm generates realisations of the climate variable through the resampling of the global pdf using the local mean and variance of the candidate station, which are estimated through a spatiotemporal model using Ordinary Kriging. The local pdf from each instant in time is then used to verify the existence of irregularities in the candidate station's series. Costa and Soares (2009) proposed to adjust the candidate series by replacing the inhomogeneous records with the mean (or median) of the pdfs calculated at the candidate station's location for the inhomogeneous periods. The capability of the geostatistical approach to detect inhomogeneities in real precipitation data was tested with very auspicious results by Costa and Soares (2009) and Ribeiro et al. (2016). However, the original geostatistical approach was considered slow, laborious and very computationally intensive.

The gsimcli method aims to provide more local information to the calculation of the local pdf of the candidate station, in order to better estimate the climatic signal of its surrounding area. Furthermore, we propose a different approach to adjust for sudden shifts in the inhomogeneous series, which is based on composite reference series derived from the estimated local pdf. Along with the implementation of the new methodology, a software package was developed, also named gsimcli, with the purpose of making its application easier and more direct. The gsimcli software and its source code are freely available on the internet (http://iled.github.io/gsimcli).

The gsimcli method's efficiency was assessed through the homogenisation of annual and monthly precipitation data from surrogate networks of the COST-HOME benchmark. This was also the main type of artificial data considered by researchers under the HOME project, because the surrogate data provide an estimate of the accuracy of the homogenisation algorithms. Unlike most of those researchers, we evaluated the gsimcli method's performance using precipitation data, which is more difficult to homogenise than temperature.

This article is organised as follows. Section 2 describes the methodology, including the gsimcli method formulation and the considered performance metrics. The study area and the surrogate precipitation data are addressed in Section 3. Several homogenisation exercises have been performed using the (original) geostatistical approach and different implementation strategies of the gsimcli method, as detailed in Section 3. The results of the different homogenisation exercises are presented and discussed in Section 4. Finally, the conclusion and future work are presented in Section 5.

2 Methodology

2.1 gsimcli method

Climate observations correspond to realisations (outcome values) of a spatiotemporal random variable $Z(u,t)$ that can take a series of values at any location in space u and instant in time t according to a probability distribution. The set of climate data measured at n locations u_α and in t_i time instants is

$$\{z(u_\alpha, t_i): \alpha = 0, 1, \dots, n-1; i = 1, \dots, T\},$$

where $\{z(u_0, t_i): i = 1, \dots, T\}$ denotes the set of values of the candidate station, and $\{z(u_\alpha, t_i): \alpha = 1, \dots, n-1;\ i = 1, \dots, T\}$ denotes the set of values of the reference stations. For each instant in time t_i, the DSS algorithm is applied in order to obtain a set of m equally probable realisations of $Z(u, t_i)$ using the whole set of climate data except the $z(u_0, t_i)$ value. In practice, m equally probable surfaces are simulated on a grid without taking into account the candidate's data for the period being tested.

The DSS algorithm simulates directly in the original data space and does not rely on multi-Gaussian assumptions. The simulated surfaces have the same statistical characteristics (auto-covariance, global sample mean and variance, and histogram) of the original variable (Soares, 2001). Because kriging interpolation requires a positive definite model of spatial variability, a variogram model must be specified. For long-term time series, it is advisable to split the series in smaller sections, in order to guarantee that the statistical properties are consistent within these sections, as recommended by Costa et al. (2008b) and Durão et al. (2010). Accordingly, the DSS algorithm should be applied independently on those smaller sections (e.g., by decade).

In the gsimcli method, the local pdf of the candidate station, for each instant in time (t_i), is defined by the set of spatiotemporal random variables that belong to a circular local neighbourhood centred at the candidate station's location:

$$\{Z^k(u_\alpha, t_i): r = 0, \dots, R;\ \alpha = 0, \dots, W_r;\ i = 1, \dots, T;\ k = 1, \dots, m\},$$

where W_r denotes the number of locations within a circle of radius r (*local radius parameter*) centred at the candidate station location (u_0). Accordingly, the estimated local pdf of the candidate station for a given instant in time t_0 is the set of simulated values:

$$\{z^k(u_\alpha, t_0): r = 0, \dots, R;\ \alpha = 0, \dots, W_r;\ k = 1, \dots, m\},$$

When r = 0, it is implied that the local pdf of the candidate station will only depend on the simulated values at its exact location. This parameter allows estimating the local pdf of the candidate station with data that contribute to better describe the climatic signal of the area on which the candidate is located. The corresponding empirical cumulative distribution function gives the estimated probability that the variable Z at location u_0 in space and instant t_0 in time is no greater than any given threshold z: $F^*(u_0, t_0; z) = Prob^*\{Z(u_0, t_0) \leq z\}$.

For the detection of irregularities (breakpoints, trend-type inhomogeneities and outliers), the method proceeds as proposed by Costa and Soares (2009). An irregular record $z(u_0, t_0)$ is identified if the interval of a specified probability p (*detection parameter*, e.g., 0.95), centred in the estimated local pdf of the candidate station for the instant t_0, does not contain the observed $z(u_0, t_0)$ value:

$$Prob^*\{Z(u_0, t_0) \leq z(u_0, t_0)\} < \frac{1-p}{2} \text{or} Prob^*\{Z(u_0, t_0) \leq z(u_0, t_0)\} < 1 - \frac{1-p}{2}$$

The detection and correction of irregularities, as well as missing values filling, are automatic procedures in the original geostatistical approach (Costa and Soares, 2009). Missing values are replaced by the mean of the local histogram of the candidate station for the corresponding time instant. Irregular values can be replaced by the mean, median, or other statistic (*correction parameter*) of the estimated local pdf for the inhomogeneous period(s). If the correction parameter is set to a percentile value equal to p (e.g., 0.95), irregularities are replaced with the percentile (1–

p)/2 or 1–(1–p)/2, depending on the irregularities being located in the lower or upper tail of the pdf, respectively. In such case, the values of the percentiles (*p*) used for detection and correction do not have to be the same. The geostatistical approach deals with trend-type inhomogeneities by correcting multiple irregularities within inhomogeneous periods. No further corrections and adjustments have been proposed by Costa and Soares (2009) and Costa et al. (2008a).

Once a candidate station is tested, the corrected time series is included in the detection process of the next candidate station as a reference time series for the calculation of the local pdf. Therefore, the detection of inhomogeneities in the second candidate station benefits from the corrections applied to the first candidate station, the third one will benefit from the previous two, and so on and so forth. Accordingly, it would be desirable to homogenise the most inhomogeneous series first, but those are unknown when homogenising real data. To overcome this limitation, the homogenisation sequence may be determined by an indicator of the level of the series inhomogeneity, such as the descending order of variance or the decreasing value of the difference between the station average and the network average (network deviation). The gsimcli software includes several alternative options to determine the order in which stations are tested: ID order, network deviation, random, variance (greater or lower), and the sequence specified by the user (e.g., to start with the series with more missing values in order to fill them in).

The automatic gsimcli method, previously described, can be extended to adjust for sudden shifts using a semi-automatic approach. Adjustments should be done cautiously and station history information should be used to support decisions, since corrections may introduce higher errors than the irregularities they try to remove. Moreover, Domonkos et al. (2011b) state that “not correcting some detected breaks may well sometimes lead to more accurate data”.

The homogenisation adjustments are estimated from a comparison series, which is computed as the ratio (in case of precipitation) between the automatically corrected candidate series and the corresponding composite reference series. This reference series is defined by the time series of the means $\bar{z}(u_\alpha, t_i)$ calculated from the local pdfs of the candidate station for each instant in time t_i:

$$\bar{z}(u_\alpha, t_i) = \frac{1}{m + W_r} \sum_{k=1}^{m} \sum_{\alpha=1}^{W_r} z^k(u_\alpha, t_i), \quad i = 1, \dots, T$$

The reason for considering the time series of the means, instead of another statistic, was that the mean and the median have very similar time series if the number of simulations is high enough (Ribeiro et al., 2016). Besides, the detection percentile should not be used because its time series has high variability, since in some instances in time it takes values corresponding to the lower tail of the local pdf, and for other instances it corresponds to the upper tail value.

The dates of the detected irregularities, together with the inspection of both the comparison series and the candidate series, serve to judge each detected inhomogeneity as a potential sudden shift, an outlier or a trend-type inhomogeneity. When decisions cannot be supported by stations' metadata, the comparison series can be statistically tested to assess the significance of such changes. In this study, we used the Buishand range test (Buishand, 1982) with a 5% significance level for this purpose.

Outlier and trend-type inhomogeneities are adjusted using the correction parameter, as suggested by Costa and Soares (2009), before applying any adjustments for sudden-shifts. The dates of sudden shifts are used to divide the comparison series into segments, and separate averages are calculated on each segment. Then, the obtained means are compared by calculating their ratio (in case of precipitation) with the mean of the most recent period. The resulting factors are then applied to the corresponding segments of the automatically corrected candidate series.

2.2 Performance metrics

From the climatologists' point of view, efficiency metrics are more appropriate to evaluate the homogenisation methods capability to improve the temporal consistency of the climatic time series than detection scores (Domonkos et al., 2011b). Domonkos (2013) discusses the problems that arise from the application of the hit rate and detection skill, which are the most traditional efficiency measures used by developers of homogenisation methods. In this study, we used the efficiency metrics proposed by Venema et al. (2012) to assess the homogenisation methods' performance.

A well-known statistical metric for measuring model performance is the root mean square error (RMSE):

$$\mathrm{RMSE(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - t_i)^2},$$

where the x_i are the homogenised values, the t_i are the true (fully homogeneous) values, and n is the sample size. The RMSE can be calculated for various time units of the observed series (e.g., month, year, and decade time units).

Venema et al. (2012) introduced a modified version of RMSE, the Centred RMSE (CRMSE), which is used as a basic accuracy metric of the data at the highest available resolution. The motivation of using CRMSE instead of RMSE in the HOME project was to eliminate the effect of unknown mean station effects (Domonkos, 2013). The Station CRMSE is defined as the RMSE of the anomalies relative to the mean bias, and it is computed on single station data directly:

$$\mathrm{CRMSE(X)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - t_i - \overline{\mathrm{X} - \mathrm{T}})^2},$$

where the upper stroke means arithmetical average, and X and T stand for homogenised and true (fully homogeneous) time series, respectively. This metric is similar to the standard deviation of the time series of the difference between the homogenised data and the truth.

The Station CRMSE quantifies the homogenisation efficiency for each station individually. The Network CRMSE measures the efficiency of the homogenisation of the network, as a whole. It is calculated using the mean CRMSE, by network. The Station [Network] Improvement evaluates the enhancement over the inhomogeneous data, and it is computed as the ratio of the Station [Network] CRMSE of the homogenised networks with the Station [Network] CRMSE of the same inhomogeneous networks. As in Venema et al. (2012), data corresponding to missing data or outliers were not taken into account in the above computations.

The performance metrics were also computed for the blind submissions to the HOME project using the homogenised series available at the HOME project's website (http://www.homogenisation.org; accessed May 2016).

3 Climate data and homogenisation framework

The HOME project (COST Action ES0601) included the creation of a benchmark data set containing real inhomogeneous data, as well as simulated data with inserted inhomogeneities. Venema et al. (2011) discuss the generation of this benchmark data set, the climate variables considered, which types of data are considered, how they have been produced, the ways to introduce artificial inhomogeneities, and additional specifications such as length, missing data and trends. The benchmark has different types of monthly datasets (temperature and precipitation) organised in three sections: real, surrogate, and synthetic data. Real inhomogeneous data is composed of temperature and precipitation monthly data series from a set of weather stations located in Europe, because of their importance for climate studies, and because they represent two important types of statistics (additive and multiplicative, respectively). These real data sets allow the comparison between different homogenisation methods with the most realistic type of data and inhomogeneities (Venema et al., 2011). The objective of the surrogate data set is to reproduce the structure of measured data accurately enough that it can be used as substitute for measurements. Surrogate climate networks reproduce the temporal cross-correlation structure of existing homogenised networks, as well as the temporal auto-correlation functions of the stations (Venema et al., 2012). Inhomogeneities were random and independently inserted in the surrogate data sets, with a normal distribution of the breakpoint sizes, and they were simultaneously introduced in multiple station series within a simulated network (Venema et al., 2012). These inhomogeneous surrogate data sets also include outliers, missing data periods and local station trends. Additionally, a stochastic nonlinear network-wide trend was added. The synthetic data sets are based on the surrogate networks, though the differences between the stations have been modelled as uncorrelated Gaussian white noise. The statistical properties of the synthetic data are those assumed by most statistical tests used for homogenisation. This data is easier to homogenise than the more realistic surrogate data (Venema et al., 2012).

3.1 Study area and data

Only surrogate series from the COST-HOME benchmark were subject to homogenisation. The following describes the precipitation data from networks 5 and 9 that have been homogenised. These networks have nine and five weather stations, respectively, and are both located in France (Figure 1). Network 9 includes five of the nine weather stations from network 5, but the time series are different in the two networks. The benchmark data set comprises precipitation monthly data for a period of 100 years (1900 – 1999). It also contains temporal intervals with missing data, which occur in the first decades (1900 – 1930) and in the beginning of the fifth decade (1940 – 1945). The lack of data intends to mimic the absence of weather stations in the beginning of the century, and the absence of measurements during the Second World War, respectively. Networks 5 and 9 cover a rectangular area of approximately 4000 km^2 (50 km x 80 km). These two networks were selected because they correspond to the precipitation networks homogenised by the MASH Marinova submission to the HOME project (MASH method operated by a first-time user named Marinova) described by Venema et al. (2012).

In this study, the monthly and annual precipitation data from those networks were subject to exploratory data analysis and homogenisation. The annual precipitation series were derived from the monthly series. As expected, the annual and monthly series from all stations have high variability and several potential outliers. Regarding network 5, station 21142001 has the highest precipitation values in the first decades. Considering the data from all nine stations, there are 102 years with missing precipitation data. The correlation coefficients of the stations' annual series vary between 0.496 and 0.847. The lowest correlation corresponds to two stations located at the centre of the network (21142001 and 21425001). The highest correlation corresponds to the stations 21142001 and 21386001. Considering the annual series from network 9, all stations have similar distributions, except station 21584001 that has higher values. The correlation coefficients of the stations' annual series vary between 0.498 (stations 21454001 and 21425001) and 0.883 (stations 21454001 and 21109001).

The main spatial patterns of the precipitation data were also investigated, particularly the presence of anisotropy. An attribute has an anisotropic behaviour when it exhibits a different spatial auto-correlation structure for different directions. The possible existence of anisotropy was analysed by producing an interpolated surface of the precipitation data for a sample of years using the Inverse Distance Weighting (IDW) interpolator. Although time consuming, these analyses were important, because if an attribute shows different auto-correlation structures in different directions, then an anisotropic variogram model should be developed to reflect these differences. The most commonly employed model for anisotropy is the geometric anisotropy, with the variogram reaching the same sill in all directions, but at different ranges. The interpolated surfaces obtained using IDW neither revealed an overall trend, nor an overall anisotropic pattern in any of the networks.

Considering that the variogram modelling is a very important stage of geostatistical methods, a thorough variography analysis was undertaken. Due to the variability of precipitation data, the lack of data in several decades, and, mainly, the reduced size of the monitoring networks, that analysis revealed to be a challenging task. Many experimental variograms exhibit a variability pattern such that the correlation between stations' data seems to be lost at short distances. Accordingly, the spatial features of precipitation occur at scales smaller than the distance between monitoring stations. For this reason, modelling the experimental variograms, and the nugget effect in particular, was not a straightforward task. Moreover, with very widely spaced data, a realistic estimate of the range parameter was also sometimes difficult to obtain. A way to overcome these drawbacks is the use of additional data provided by other weather stations located in the surrounding study area. However, such task could not be performed, since only the provided data sets by the HOME project could be used in the process.

In previous exploratory homogenisation activities with the original geostatistical approach, different variogram models for the spatial continuity structure of the data were assessed (Costa et al., 2015). The variogram models that lead to the best performance metrics are considered in this study. As recommended by Costa et al. (2008b) and Durão et al. (2010), variograms were first estimated by decade in order to account for possible long-term trends, or fluctuations, in the precipitation auto-correlation structure. This approach was followed in the case of annual data from network 5 (Table 1). Due to the small number of stations in network 9, a single variogram model for the whole 1900–1999 period was estimated for the annual data (Table 1). Previous exploratory homogenisation activities indicated that using all yearly data to infer a single variogram model for network 9 provided similar results to using the same decadal variogram models inferred for network 5 (Costa et al., 2015). Hence, estimating a single variogram model for the whole period is the recommended solution in case of small networks.

Due to the lack of data in the monthly series, a unique variogram model was estimated for the first, second and third decades (1900–1929) from network 5, for each month (Table A1 of the Appendix). For the same reason, the fourth and fifth decades' data were also combined in order to obtain another single variogram model. Seven variogram models were prepared for each monthly series, in a total of 84. The estimated variogram models for network 5 were also used in network 9 (Table A1 of the Appendix), since the reduced number of stations in this network did not allow to obtain a reliable estimate of the variogram model.

3.2 Specifications of the homogenisation exercises

Several homogenisation exercises were undertaken for the precipitation networks 5 and 9 from the COST-HOME benchmark using different sets of parameters (Table 2). We investigated the impact of two strategies on the definition of the simulation grids. The homogenisation exercises used a regular grid comprising 9882 cells (81 x 122 cells) for a grid cell size of 1 km, except one exercise with annual data that used a regular grid with 425 cells (17 x 25 cells) having a cell size of 5 km. Different values of the local radius parameter (*r*) were also considered, ranging from 1 to 5 cells (Table 2). All homogenisation exercises with the gsimcli method used the following common set of parameters:

- Candidates order = descending order of the stations' data variance;
- Number of simulations (m) = 500;
- Detection parameter (p) = 0.95;
- Correction parameter =percentile value of 0.975.

The annual series were homogenised using both the automatic and semi-automatic versions of gsimcli. Considering that the later did not significantly improve the method's efficiency, the monthly series were only homogenised using the automatic gsimcli. In the adjustments stage of the annual series, whenever the candidate series had missing values in the beginning of the time series, it was considered the existence of a breakpoint at the first date with data in the automatically corrected candidate series. Therefore, the missing values that were automatically estimated were also adjusted, despite the fact that these data are not used in the computation of the performance metrics.

The original geostatistical approach (Costa and Soares, 2009) was also used to homogenise the annual time series from networks 5 and 9. This was accomplished by setting the local radius parameter (*r*) to zero, and the correction parameter to the mean of the local pdf of the candidate station. No further adjustments were applied in these homogenisation exercises. The simulation grid was defined with cells of 1 km^2.

4 Results and discussion

The automatic gsimcli method homogenises candidate time series using the correction parameter derived from the estimated local pdf for the inhomogeneous periods. In the semi-automatic version, the automatically corrected candidate series are further adjusted using correction factors derived from comparison series. These are based on composite reference series corresponding to the series of means computed from the estimated local pdfs. The different parameters used in the homogenisation of the precipitation series are described in Section 3.2. The following sections detail the results of the precipitation data homogenisation.

4.1 Annual precipitation series

For illustration purposes, Figure 2 shows the candidate time series of station 21142001 from network 5, and the homogenised series that were obtained using the gsimcli method with the parameters specified for Test #6 (Table 2), as well as the corresponding composite reference series and comparison series. The Buishand range test identified a significant sudden shift in 1952 in this candidate time series. No other significant breakpoints were identified in the segments before and after this year.

The irregular years identified in the automatic stage of gsimcli, as well as the years corresponding to significant sudden-shifts identified by the Buishand's test are listed in Table A2 of the Appendix. This table also presents the years defined by HOME project as breakpoints and outliers. It is important to note that these irregularities were introduced in the monthly time series of the benchmark data set. Certain inhomogeneities might only be evident at certain timescales of variability (Yan and Jones, 2008). In this study, those monthly irregularities were considered as annual breakpoints for comparison purposes, thus the detection results should be analysed with caution. Those inhomogeneities might not be detected as breakpoints in the homogenisation exercises, since the annual amounts of precipitation may smooth those monthly irregularities.

The number of years with detected irregularities by the automatic gsimcli does not seem to be dependent of the local radius parameter (*r*), since it is similar in the different homogenisation exercises. It is higher than the number of breakpoints defined by the HOME project. In some cases, a sequence of more than two consecutive years with irregularities is detected, which can be assumed as the detection of a trend-type inhomogeneity in the candidate series by the automatic gsimcli (e.g., in the station 21454001 from network 5 there are breakpoints detected consecutively from 1911 to 1914, from 1940 to 1946, and from 1987 to 1993).

The breakpoint years detected by the Buishand's test are similar for all homogenisation exercises, varying from zero to three, thus the performance metrics obtained with the automatic and semi-automatic gsimcli are also similar (Table 3).The automatically corrected candidate series from stations 21454001, 21501003 and 21584001 from network 5, and station 21109001 from network 9, were considered as homogeneous by the Buishand's test in all homogenisation exercises. One additional breakpoint year (1984) was identified in the automatically corrected candidate series from station 21310001 from network 5, and two additional years (1906 and 1918) in station 21584001 from network 9, using the homogenisation Test #6. Only one breakpoint year (1926) was identified in station 21425001 from network 9 using the homogenisation Test #1, whereas all other homogenisation exercises identified two breakpoint years (1917 and 1926) in this station. The breakpoint year of 1937 was not identified in station 21711001 from network 9 using the homogenisation Test #2.

The performance metrics were computed for the homogenisation exercises considering the application of the automatic gsimcli method (without adjustments for sudden shifts), and the semi-automatic version (with the additional adjustments stage) (Tables 3 and 4). The performance metrics were also computed for the homogenisation activities undertaken with the original geostatistical approach, and for the blind submissions to the HOME project that homogenised networks 5 and 9. All the homogenisation exercises undertaken with the annual precipitation data from network 9 (Table 4) made the data more inhomogeneous, i.e. had a Station improvement quotient over the inhomogeneous data above one. However, the original geostatistical approach was the only homogenisation activity undertaken that made the data from network 5 (Table 3) more inhomogeneous. The higher number of stations in network 5 might explain the better results obtained for this network than for network 9. All the values of the Station CRMSE of the gsimcli method are at least 24% smaller than those of the original geostatistical approach. Considering the

Network CRMSE, the efficiency increase of the gsimcli method is greater for the automatic version (at least 44%) than for the semi-automatic one (at least 24%). Accordingly, the gsimcli method is more efficient than the original geostatistical approach. Nonetheless, the gsimcli method underperformed all the blind submissions to the HOME project, except the absolute method (h008 - PMFred abs) for network 5.

Considering the performance of the automatic and semi-automatic versions of gsimcli, both provide similar results. For the Station's CRMSE and Improvement, the semi-automatic gsimcli was more efficient (in average, 2%) for network 5, and less harmful (in average, 11%) for network 9. Regarding the Network's CRMSE and Improvement, the automatic gsimcli provided better results than the semi-automatic gsimcli (in average, 35% in network 5 and 16% in network 9). These results seem to indicate that the automatic gsimcli increases the temporal consistency of the regional climate signal more than the semi-automatic version.

In network 5 (Table 3), the smallest Network metrics were obtained for the homogenisation Test #6 with both the automatic (Network CRMSE = 2.86; Network Improvement = 1.07), and the semi-automatic (Network CRMSE = 3.78; Network Improvement = 1.41) versions of gsimcli. The efficiency of the semi-automatic gsimcli Test #6 was lower than the Climatol (h010) and AnClim main (h018) procedures by 17% and 21%, respectively, in terms of Station CRMSE. However, all automatic versions of gsimcli were more efficient (at least 2%) than the Climatol (h010) in terms of Network CRMSE. The efficiency of the automatic gsimcli Test #6 was lower than the C3SNHT (h006), AnClim main (h018) and PRODIGE main (h002) procedures by 11%, 12% and 13%, respectively, in terms of Network CRMSE. Considering the results of network 9 (Table 4), the automatic gsimcli homogenisation Test #5 was the less harmful (Network CRMSE = 2.87; Network Improvement = 1.53), whereas using the semi-automatic version the "best" homogenisation Test was #2 (Network CRMSE = 3.08; Network Improvement = 1.64). These results indicate that using the local radius parameter (*r*) with values greater than 1 does not conclusively increase the gsimcli's efficiency. However, using larger grid cells generally improves the gsimcli method efficiency and decreases the processing time, since the size of the simulation grid cells has a significant impact in the computational effort. These results are consistent with a preliminary sensitivity analysis of the gsimcli's parameters that was undertaken using monthly precipitation data from the benchmark's network 16 (Ribeiro et al., 2015b), which is located in Austria and comprises 15 stations.

4.2 Monthly precipitation series

The monthly series were homogenised with the automatic gsimcli method as described in Section 3.2. Even though the performance metrics of the homogenisation exercises provided similar values, the best results were obtained in Tests #8 and #9 (Table 5), which used a local radius parameter greater than zero. In the previous homogenisation exercises, using a local radius parameter (*r*) equal to zero provided similar results to the Tests #8 and #9. These results might be explained by the fact that the correction parameter was the percentile of 0.975, whereas the original geostatistical approach used the mean as the correction parameter in the homogenisation of the annual series. This suggests the high importance of the correction parameter in the overall homogenisation efficiency.

The efficiency of the automatic gsimcli was higher than the C3SNHT (h006), AnClim main (h018) and Climatol (h010) procedures by 24%, 19% and 8%, respectively, in terms of Station CRMSE. However, it underperformed the PRODIGE monthly (h021) and the MASH Marinova (h009) procedures by 9% and 22%, respectively. It is noticeable that, in comparison with other procedures, the efficiency of gsimcli in the homogenisation of monthly series is higher when compared to the annual series.

5 Conclusion

In the original geostatistical approach (Costa and Soares, 2009; Costa et al., 2008a), the detection and correction stages of the homogenisation process were automatic procedures based on individual pieces of data. The proposed gsimcli algorithm includes a new parameter (local radius) that aims to provide more local information to the calculation of the local pdf in order to reproduce the climatic signal of that location more realistically. Moreover, the gsimcli method may include another stage that aims at further adjusting the candidate time series by examining the characteristics of segments of data (semi-automatic version). Both automatic and semi-automatic versions of the gsimcli method proved to be more efficient in the homogenisation of the benchmark's precipitation series than the geostatistical approach proposed by Costa and Soares (2009) and Costa et al. (2008a).

The semi-automatic version of gsimcli uses comparison series that can be statistically tested in order to proceed with further inhomogeneities detection and adjustments. In this study, both gsimcli versions provided similar results in the homogenisation of annual precipitation series. We used the Buishand's test in the semi-automatic gsimcli, but the application of other techniques should be investigated.

Even though the geostatistical homogenisation made the data slightly more inhomogeneous in many experiments, the gsimcli approach outperformed a few procedures in the homogenisation of monthly precipitation data (Table 5), and Climatol in the homogenisation of monthly temperature series (as shown in Table S3). It is also important to point out that the benchmark's networks are relatively small, and that the gsimcli method is more appropriate for larger networks. Ribeiro et al. (2016) tested the gsimcli method with a real data set of 66 monitoring stations from Portugal (0.0015 stations/km^2 in a simulation grid with 1 km^2 cells), whereas networks 5 and 9 have nine and five stations, respectively (0.0009 and 0.0005 stations/km^2 in the simulation grids with 1 km^2 cells, respectively).

Geostatistical techniques are suitable for variables that exhibit spatial correlation, which is modelled by the variogram. A higher number of observations that are spatially well distributed allows for a more accurate estimation of the variogram, thus improving the quality of the kriging predictions. A major limitation of this study was the reduced number of points available to estimate the variogram models. The modelling was particularly difficult for the shorter lag distances, which tend to have very few pairs of points. This is an important weakness of gsimcli, since the variogram's behaviour near the origin is the most important to characterise. Accordingly, further research with larger networks should be pursued.

Another direction for future research is the application of Direct Sequential Cosimulation (coDSS; Soares, 2001), which is an extension of the DSS algorithm that allows incorporating covariates such as elevation. Such extension of the gsimcli procedure could be suited for homogenising climatological networks from mountainous regions. However, the variography analysis would be even more challenging, because the coDSS algorithm requires a linear model of coregionalisation (i.e., modelling the spatial correlation structure through the simple and cross variograms). Another potential drawback is that the computational effort would highly increase.

The proposed approach is a valuable contribution to this research field, particularly the new methods' capability for filling missing values, and irregularities filtering. However, data corresponding to missing data or outliers were not taken into account in the computation of the performance metrics.

Acknowledgements

The authors gratefully acknowledge the financial support of "Fundação para a Ciência e Tecnologia" (FCT), Portugal, through the research project PTDC/GEO-MET/4026/2012

("GSIMCLI - Geostatistical simulation with local distributions for the homogenization and interpolation of climate data").

Supporting Information

The automatic version of the gsimcli method (without adjustments for sudden shifts) was also used to homogenise monthly temperature data from network 4 (Figure S1) of the HOME benchmark, considering different sets of parameters (Table S2). The Supporting Information describes these analyses and their results, which were evaluated with the performance metrics (Table S3).

Titles of Figures in the Supporting Information:

- Figure S1 – Location of stations from network 4 in the North of the Bay of Biscay (Digital Elevation Model source: Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled seamless SRTM data V4, http://srtm.csi.cgiar.org; accessed November 2015).

Titles of Tables in the Supporting Information**:**

- Table S1 – Variogram models of the monthly temperature series from network 4.
- Table S2 – Parameters of the homogenisation exercises of monthly temperature data from network 4.
- Table S3 – Performance metrics of the monthly temperature series from network 4 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidelines on Climate Metadata and Homogenization. In: WMO/TD No. 1186, WCDMP No. 53, Llansó P (ed.). World Meteorological Organization, Geneva, Switzerland, 55 pp.

Allen RJ, De Gaetano AT. 2000. A method to adjust long-term temperature extreme series for non climatic inhomogeneities. *Journal of Climate* **13**(2): 3680–3695. doi:10.1175/1520-0442(2000)013<3680:AMTALT>2.0.CO;2.

Auer I, Bohm R, Jurkovic A, Orlik A, Potzmann R, Schoner W, Ungersbock M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, Moisselin JM, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajic-Capka M, Zaninovic K, Majstorovic Z, Szalai S, Szentimrey T, Mercalli L. 2005. A new instrumental precipitation dataset for the greater alpine region for the period 1800-2002. *International Journal of Climatology* **25**(2): 139–166. doi:10.1002/joc.1135.

Begert M, Schlegel T, Kirchhofer W. 2005. Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology* **25**(1): 65–80. doi:10.1002/joc.1118.

Brunet M, Jones P. 2011. Data rescue initiatives: bringing historical climate data into the 21st century. *Climate Research* **47**(1): 29–40. doi: 10.3354/cr00960.

Brunetti M, Maugeri M, Monti F, Nanni T. 2006. Temperature and precipitation variability in Italy in the last two centuries from homogenized instrumental time series. *International Journal of Climatology* **26:** 345–381. doi: 10.1002/joc.1251.

Buishand T. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**(1): 11–27. doi:10.1016/0022-1694(82)90066-X.

Caussinus H, Mestre O. 1996. New mathematical tools and methodologies for relative homogeneity testing. In: *Proceedings of the First Seminar for Homogenization of Surface Climatological Data*, Hungarian Meteorological Service (Ed.). Budapest, Hungary, 63–82.

Caussinus H, Mestre O. 2004. Detection and correction of artificial shifts in climate series. *Applied Statistics* **53:** 405–425. doi:10.1111/j.1467-9876.2004.05155.x.

Costa AC, Soares A. 2009. Homogenization of climate data: Review and new perspectives using geostatistics. *Mathematical Geosciences* **41**(3): 291–305. doi:10.1007/s11004-008-9203-3.

Costa AC, Negreiros J, Soares A. 2008a. Identification of inhomogeneities in precipitation time series using stochastic simulation. In: *geoENV VI – Geostatistics for Environmental Applications*, Soares A, Pereira MJ, Dimitrakopoulos R (eds.). Springer, 275–282. doi:10.1007/978-1-4020-6448-7_23.

Costa AC, Durão R, Pereira MJ, Soares A. 2008b. Using stochastic space-time models to map extreme precipitation in southern Portugal. *Natural Hazards and Earth System Sciences* **8**(4): 763–773. doi:10.5194/nhess-8-763-2008.

Costa AC, Soares A, Henriques R, Ribeiro S, Caineta J, 2015. Final Scientific Report of the Project. GSIMCLI – Geostatistical simulation with local distributions for the homogenization and interpolation of climate data, PTDC/GEO-MET/4026/2012. NOVA Information Management School, 30 November 2015, 72 pp.

Craddock JM. 1979. Methods of comparing annual rainfall records for climatic purposes. *Weather* **34:** 332–346. doi:10.1002/j.1477-8696.1979.tb03465.x.

Domonkos P. 2013. Measuring performances of homogenization methods. *Időjárás* **117**(1): 91–112.

Domonkos P. 2015. Homogenization of precipitation time series with ACMANT. *Theoretical and Applied Climatology* **122**(1): 303–314. doi:10.1007/s00704-014-1298-5.

Domonkos P, Coll J. 2016. Homogenisation of temperature and precipitation time series with ACMANT3: method description and efficiency tests. *International Journal of Climatology*. doi:10.1002/joc.4822.

Domonkos P, Poza R, Efthymiadis D. 2011a. Newest developments of ACMANT. *Advances in Science Research* **6:** 7–11. doi:10.5194/asr-6-7-2011.

Domonkos P, Venema V, Mestre O. 2011b. Efficiencies of homogenisation methods: our present knowledge and its limitation. In: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.). World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, 11–24.

Domonkos P, Venema V, Auer I, Mestre O, Brunetti M. 2012. The historical pathway towards more accurate homogenisation. *Advances in Science Research* **8**(1): 45–52. doi:10.5194/asr-8-45-2012.

Durão RM, Pereira MJ, Costa AC, Delgado J, del Barrio G, Soares A. 2010. Spatial-temporal dynamics of precipitation extremes in southern Portugal: a geostatistical assessment study. *International Journal of Climatology* **30**(10): 1526–1537. doi:10.1002/joc.1999.

Guijarro J. 2006. Homogenization of a dense thermo-pluviometric monthly database in the Balearic Islands using the free contributed R package “Climatol”. In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.). World Meteorological Organization, Budapest, Hungary, WMO-TD No. 1493, 25–36.

Guijarro J. 2011. Influence of network density on homogenization performance. In: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Vincze E (eds.). World Meteorological Organization, Budapest, Hungary, WCDMP-No. 78, 11–18.

Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT), http://srtm.csi.cgiar.org (accessed November 2015).

Menne MJ, Williams Jr. CN. 2009. Homogenization of temperature series via pairwise comparisons. *Journal of Climate* **22**(7)**:** 1700–1717. doi:10.1175/2008JCLI2263.1.

Menne MJ, Williams Jr. CN, Vose RS. 2009. The U. S. historical climatology network monthly temperature data, version 2. *Bulletin of American Meteorology Society* **90**: 993–1007. doi:10.1175/2008BAMS2613.1.

Mestre O, Gruber C, Prieur C, Caussinus H, Jourdain S. 2011. SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology* **50**(11): 2343–2358. doi:10.1175/2011JAMC2641.1.

Mestre O, Domonkos P, Picard F, Auer I, Robin S, Lebarbier E, Bohm R, Aguilar E, Guijarro J, Vertachnik G, Klancar M, Dubuisson B, Štepánek P. 2013. HOMER: a homogenization software – methods and applications. *Időjárás* **117**(1): 47–67.

Peterson TC, Easterling DR, Karl TR, 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, Forland EJ, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D. 1998. Homogeneity adjustments of in situ atmospheric climate data: A review. *International Journal of Climatology* **18**(13): 1493–1517. doi:10.1002/(SICI)1097-0088(19981115)18:13<1493::AID-JOC329>3.0.CO;2-T.

Ribeiro S, Caineta J, Costa AC. 2015a. Review and discussion of homogenisation methods for climate data. *Physics and Chemistry of the Earth*. doi:10.1016/j.pce.2015.08.007.

Ribeiro S, Caineta J, Costa AC, Soares A. 2015b. Establishment of detection and correction parameters for a geostatistical homogenisation approach. *Procedia Environmental Sciences* **27**: 83–88. doi:10.1016/j.proenv.2015.07.115.

Ribeiro S, Caineta J, Costa AC, Henriques R, Soares A, 2016. Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation. *Atmospheric Research* **171**: 147–158. doi:10.1016/j.atmosres.2015.11.014.

Sahin S, Cigizoglu HK. 2010. Homogeneity analysis of Turkish meteorological data set. *Hydrological Processes* **24**(8): 981–992. doi:10.1002/hyp.7534.

Soares A. 2001. Direct sequential simulation and cosimulation. *Mathematical Geology* **33**(8): 911–926. doi:10.1023/A:1012246006212.

Štepánek P. 2008a. AnClim – Software for Time Series Analysis. Dept. of Geography, Fac. of Natural Sciences, MU, Brno. <http://www.climahom.eu/AnClim.html>.

Štepánek P. 2008b. ProClimDB – Software for Processing Climatological Datasets. CHMI, Regional Office Brno. <http://www.climahom.eu/ProcData.html>.

Szentimrey T. 1999. Multiple analysis of series for homogenization (MASH). In: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, WMO, WCDMP-No. 41, WMO-TD No. 1962, 27–46.

Szentimrey T. 2007. Manual of homogenization software MASH v3.02. Hungarian Meteorological Service, 65 pp.

Szentimrey T. 2008. Development of MASH homogenization procedure for daily data. In: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Lakatos M, Szentimrey T, Bihari Z, Szalai S (eds.). Budapest, Hungary 2006, WCDMP-No.71, pp. 123–130.

Trewin B. 2010. Exposure, instrumentation, and observing practice effects on land temperature measurements. WIREs *Climate Change* **1:** 490–506. doi:10.1002/wcc.46.

Venema VKC, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Müller-Westermeier G, Lakatos M, Williams CN, Menne MJ, Lindau R, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquaotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P, Brandsma T. 2011. Description of the COST-HOME monthly benchmark dataset and the submitted homogenized contributions. HOME – Advances in Homogenisation Methods of Climate Series: an Integrated Approach (COST Action ES0601), Meteorological Institute, University of Bonn, Germany, 18 Juli 2011, 32 pp.

Venema V, Mestre O, Aguilar E, Auer I, Guijarro J, Domonkos P, Vertacnik G, Szentimrey T, Štěpánek P, Zahradníček P, Viarre J, Muller-Westermeier G, Lakatos M, Williams C, 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, Prohom Duran M, Likso T, Esteban P, Brandsma T. 2012. Benchmarking homogenization algorithms for monthly data. *Climate of the Past* **8**(1): 89–115. doi:10.5194/cp-8-89-2012.

Wang XL. 2008. Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal t or F test. *Journal of Applied Meteorology and Climatology* **47**(9): 2423–2444. doi:10.1175/2008JAMC1741.1.

World Meteorological Organization, 2010. *Guide to Climatological Practices*. WMO No. 100, third edition, World Meteorological Organization, Geneva, Switzerland.

Yan ZW, Jones PD. 2008. Detecting inhomogeneity in daily climate series using wavelet analysis. *Advances in Atmospheric Sciences* **25**(2): 157–163. doi:10.1007/s00376-008-0157-7.

Yozgatligil C, Yazici C. 2016. Comparison of homogeneity tests for temperature using a simulation study. *International Journal of Climatology* **36**(1): 62–81. doi:10.1002/joc.4329.

Appendix

Table A1

Table A2

Tables' captions

Table 1 – Variogram models of the annual precipitation series from networks 5 and 9.

Table 2 – Parameters of the homogenisation exercises with the gsimcli method.

Table 3 – Performance metrics of the annual precipitation series from network 5 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Table 4 – Performance metrics of the annual precipitation series from network 9 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Table 5 – Performance metrics of the monthly precipitation series from networks 5 and 9 for the homogenisation exercises undertaken and for the blind contributions to the HOME project.

Table A1 – Variogram models of the monthly precipitation series from networks 5 and 9.

Table A2 – List of the years with breakpoints and outliers defined by the HOME project (the "truth"), and of the irregular years that were detected in the homogenisation exercises. Years marked in bold are correctly detected breakpoints (with a tolerance of 2 years), and years marked in bold and underlined are correctly detected outliers.

Figures' captions

Figure 1 – Location of stations from networks 5 and 9 in France (Digital Elevation Model source: Jarvis et al., 2008).

Figure 2– Graphic (a) shows the candidate time series of station 21142001 from network 5, and the corresponding homogenised series using the automatic and semi-automatic versions of gsimcli with the parameters specified for Test # 6. Other graphics show the composite reference series (b) and

the comparison series (c) used in the homogenisation of this candidate series with the semi-automatic gsimcli.