Math Geosci (2009) 41: 291–305
DOI 10.1007/s11004-008-9203-3

SPECIAL ISSUE

Homogenization of Climate Data: Review and New Perspectives Using Geostatistics

Ana Cristina Costa · Amílcar Soares

Received: 28 July 2008 / Accepted: 20 October 2008 / Published online: 28 November 2008

Abstract The homogenization of climate data is of major importance because non-climatic factors make data unrepresentative of the actual climate variation, and thus the conclusions of climatic and hydrological studies are potentially biased. A great deal of effort has been made in the last two decades to develop procedures to identify and remove non-climatic inhomogeneities. This paper reviews the characteristics of several widely used procedures and discusses the potential advantages of geostatistical techniques. In a case study, the geostatistical simulation approach is applied to precipitation data from 66 monitoring stations located in the southern region of Portugal (1980–2001). The results from this procedure are then compared with those from three well established statistical tests: the Standard normal homogeneity test (SNHT) for a single break, the Buishand range test, and the Pettit test. Promising results from the case study open new research perspectives on the homogenization of climate time series.

Keywords Data quality · Composite reference series · Homogeneity tests · Nonparametric tests

1 Introduction

A homogeneous climate time series is defined as one where the variations are caused only by variations in climate (Aguilar et al. 2003). Non-climatic factors may hide the

A.C. Costa (✉)
ISEGI, Universidade Nova de Lisboa, Lisboa, Portugal
e-mail: ccosta@isegi.unl.pt

A. Soares
CERENA, Instituto Superior Técnico, Lisboa, Portugal
e-mail: asoares@ist.utl.pt

true climatic signals and patterns, and thus potentially bias the conclusions of climate and hydrological studies. Frequent factors are monitoring stations relocations, changes in instrumentation, changes of the surroundings, instrumental inaccuracies, and changes of observational and calculation procedures. Unfortunately, few long-term climate time series are free of irregularities (Auer et al. 2005). Consequently, it is an important task to assess the homogeneity of long climate records before they can be reliably used. Several techniques have been developed for non-climatic inhomogeneities detection and adjustment, i.e., homogenization. If the identified irregularities are due to non-climatic factors then adjustments are performed to compensate for the biases produced by the inhomogeneities. Since there is no one single best technique to be recommended, the following four steps are commonly followed (Aguilar et al. 2003): (i) metadata analysis and basic quality control, (ii) creation of reference time series, (iii) breakpoint detection, and (iv) data adjustment.

Most of the procedures that have been proposed to identify and remove non-climatic inhomogeneities are not proper for immediate application on data with low temporal resolution (i.e., daily or hourly data). In fact, well-established statistical methods for homogeneity testing sub-monthly precipitation data are lacking (Aguilar et al. 2003; Auer et al. 2005; Wijngaard et al. 2003). Furthermore, adjusting daily and hourly data is not straightforward. Thus, the World Meteorological Organization (WMO) makes no recommendations regarding adjusting sub-monthly data. As an alternative, the WMO advises that data should be carefully evaluated for the impacts of inhomogeneities and that portions of time series with homogeneity problems be excluded from the analysis before using sub-monthly data in long-term climate change analysis (Aguilar et al. 2003).

This paper reviews the characteristics of the procedures that are most commonly used for the homogenization of climate time series and discusses the potential advantages of using geostatistical techniques for inhomogeneities detection and adjustment. First, several basic concepts and general issues concerning the homogenization of climate data are introduced (Sect. 2). The most commonly used inhomogeneities detection procedures are reviewed in Sect. 3. The methodological framework of the geostatistical simulation approach is described in Sect. 4, and the case study results are summarized in Sect. 5. Finally, different homogenization approaches are discussed by highlighting their advantages and limitations and new research perspectives are suggested in Sect. 6.

2 Absolute and Relative Approaches

Two groups of homogeneity testing techniques can be distinguished and are usually referred to as 'absolute' and 'relative' methods. In the first set of procedures, the statistical tests are applied to each station data separately. In the second set, the testing procedures use records from neighboring stations (named reference stations) which presumably are homogeneous. While both approaches are worthwhile and valid, they both have drawbacks. Using only data from an individual station is problematic because it is difficult to determine if changes or lack of changes result from non-climatic or climatic influences (Peterson et al. 1998). To overcome this problem, metadata support from station history information is essential for evaluating the breaks detected.

Relative methods intend to isolate the non-climatic influences. They assume that within a geographical region, climatic patterns will be identical and that observations from all sites within the region will reflect this identical pattern. The data collected at all sites within the same climatic region should be highly correlated, have similar variability, and differ only by scaling factors and random sampling variability. Problems arise when the inhomogeneities in the climate data series are caused by simultaneous changes in the observational network, such as simultaneous changes in the measuring technique, as relative tests become insensitive since all series are affected at the same time (Tuomenvirta 2001; Wijngaard et al. 2003). Furthermore, ambiguous conclusions are possible when several neighboring stations have inhomogeneities themselves (Boissonnade et al. 2002; Reeves et al. 2007; Tayanç et al. 1998). The most common approach for selecting reference stations is to form Pearson correlation matrices between the candidate site and neighboring stations' data, and to take as reference the highest correlated ones (Boissonnade et al. 2002; Tayanç et al. 1998). Other approaches extract principal components from the whole data network, or use an independent data source thought to be homogeneous (Aguilar et al. 2003). Some procedures search for breakpoints or artificial trends in a composite reference series (or, alternatively, in the data when a suitable composite series cannot be built), while some statistical tests compare the candidate data series with data from reference stations (i.e., tests for the difference between medians).

Using composite reference series—ratio series for precipitation and difference series for temperature—is a standard procedure in the detection of non-climatic homogeneities. The assumption from this approach is that the composite reference series includes the regional climate trends and fluctuations present in the data of the candidate, but does not contain discontinuities itself during the period of analysis of the discontinuity in the candidate station. Composite reference series are computed as a weighted average of data from neighboring stations by using some measure of statistical similarity (usually the correlation coefficient or an inverse function of the distance) between them. Romero et al. (1998) proposed a combined use of those measures in order to increase the contribution of the records from closer stations, both in spatial and correlation terms. Alexandersson and Moberg (1997) proposed the construction of ratio (difference) reference series, which are generally used in precipitation (temperature) studies. The most usual approach to obtain adjustment factors is to calculate separate averages on the difference or ratio series for the two sections defined by a breakpoint (Aguilar et al. 2003). When abrupt changes are identified in the time series, the obtained means are compared by calculating their ratio or difference and the obtained factor is applied to the inhomogeneous part. When dealing with gradual trends or breakpoints superimposed on trends, the inhomogeneous section is de-trended, using the slope calculated on the difference or ratio time series.

3 Review on Homogenization Procedures

The approaches underlying the homogenization techniques are quite different and typically depend on the type of elements (temperature, precipitation, pressure, evaporation, etc.), the temporal resolution of the observations (annual, seasonal, monthly, or sub-monthly), the availability of metadata (station history information) and the monitoring station network density or spatial resolution. Several widely used techniques for inhomogeneities detection and homogenization will be summarized. A few descriptions and references were obtained from Aguilar et al. (2003) and Peterson et al. (1998). Comparisons between the procedures are provided by Ducré-Robitaille et al. (2003) and Reeves et al. (2007). The listing of methods is presented in alphabetical order, without distinguishing between absolute and relative procedures since absolute tests can also be used in relative approaches by applying the test to the composite reference series.

Subjective methods, such as the double-mass analysis (Kohler 1949), will not be presented because they should only be used for exploratory analysis. Other common homogenization procedures referred to in the literature, but not described here, are the Craddock test (Craddock 1979), the Potter's method (Potter 1981), and the well known Wilcoxon–Mann–Whitney test (Mann and Whitney 1947; Wilcoxon 1945). The procedures summarized are intended to illustrate the variety of approaches that are commonly used. The literature is replete with techniques, but most of them are similar or variations of the methodologies described here.

3.1 Buishand Range Test

The Buishand range test (Buishand 1982) is a parametric test and supposes, under the null hypothesis, that the values of the testing variable are independent and identically normally distributed. Under the alternative hypothesis, it assumes that a step-wise shift in the mean (a break) is present. This test is capable of locating the period (month or year) where a break is likely, but it is more sensitive to breaks in the middle of a time series (Wijngaard et al. 2003).

3.2 Kruskal–Wallis Test

The Kruskal–Wallis test (Kruskal 1952; Kruskal and Wallis 1952) is a well known nonparametric (or distribution free) test used to compare two or more independent groups of sampled data. This test is an alternative to the ANOVA F independent group test (which compares the means of several groups), when the assumption of normality is not met. One of the assumptions of the Kruskal–Wallis test is that the observations are drawn randomly and independently from their respective populations. For a two-tailed test, the alternative hypothesis is that not all of the samples come from identical populations. If all of the population distributions have the same shape (normal or not), these hypotheses are also sometimes written as testing the equality of the central tendency of the populations (i.e., testing whether all the independent samples have been drawn from populations possessing equal medians). The Kruskal–Wallis test gives little information about the probable date for a shift in the median and no information about the magnitude of the break.

3.3 Mann–Kendall Test

An assumption of trend tests is that trends are consistently increasing or decreasing, otherwise known as monotonic changes. The nonparametric Mann–Kendall test has traditionally been used to test randomness against the monotonic trend. Since the first proposals of the test by Mann (1945) and Kendall (1975), the test was extended in order to include seasonality (Hirsch et al. 1982; Hirsch and Slack 1984), multiple time series (Lettenmaier 1988), and covariates (Libiseller and Grimvall 2002). Yue and Wang (2004) discuss several approaches that use the effective sample size to modify the test statistic in order to eliminate the effect of serial correlation.

3.4 Multiple Analysis of Series for Homogenization (MASH)

The MASH method was developed in the Hungarian Meteorological Service (Szentimrey 1994, 1999); it is a relative homogeneity test procedure that does not assume that the reference series are homogeneous. Possible break points and shifts can be detected and adjusted through mutual comparisons of series within the same climatic area. The role of series (candidate or reference series) changes step by step in the course of the procedure. MASH is a multiple break points detection technique that takes into account the significance and the efficiency of the test. Moreover, it provides not only estimated break points and shift values, but the corresponding confidence intervals as well. Hence, the series can be adjusted by using the point and interval estimates. Another feature of MASH is that an additive or a cumulative model can be used depending on the climate elements (e.g., temperature, precipitation). An interesting feature of the software developed for this method (MASH system) is that the probable dates of break points provided by metadata information can be used automatically. In the case of having monthly series for all the 12 months, the MASH system also allows the monthly, seasonal, and annual series to be homogenized together. More recently, Szentimrey (2003) introduced a new procedure to evaluate the homogenization results in the MASH system. The verification procedure evaluates the quality of the homogenized series by the joint comparative mathematical examination of the original and the homogenized series systems.

3.5 Pettit Test

Pettit (1979) developed a nonparametric test that is capable of locating the period (month or year) where a break is likely. The null hypothesis is that the data is independent, identically distributed random quantities, and the alternative is that a stepwise shift in the mean is present. The test statistic is related to the Mann–Whitney statistic. Like other tests, the Pettit test is more sensitive to breaks in the middle of a time series (Wijngaard et al. 2003).

3.6 Regression-Based Methods

This section reviews several regression-based techniques that have been proposed for the homogenization of climate time series. A two-phase regression technique for detecting a change point in the trend of a time series is described by Solow (1987). In this method, the regression lines before and after the year that is being tested are constrained to meet at that point. Easterling and Peterson (1995) developed a variation on this technique in which the regression lines are not constrained to meet, and where a linear regression is fitted to the part of the reference series before the year being tested and another one after the year being tested. This test is repeated for all years of the time series (with a minimum of 5 years in each section), and the year with the lowest residual sum of the squares is considered the year of a potential discontinuity. The time series is then divided into two at that year, and both sub-series are similarly tested. This subdividing process continues until no significant breaks are found or the time series are too short to test. Reeves et al. (2007) discuss a number of variants of this two-phase regression technique.

Allen et al. (1998) describe the ellipse test or accumulated residual method which uses the cumulative residuals from the linear regression between the candidate series (dependent variable) and data from a neighboring station (independent variable), or the average observations of several surrounding stations inside the same climatic region. The candidate series can be considered homogeneous if the cumulative residuals are not biased. The bias hypothesis can be tested using an ellipse defining the confidence limits. Plotting the cumulative residuals against time, using the time scale (interval) of the variable under analysis, the accumulated residual curve is obtained. If all the cumulative residuals lie inside the ellipse then the hypothesis of homogeneity is not rejected for the significance level considered. This test is capable of locating the period (year) where a break is likely to occur. Costa and Soares (2006) proposed an extension of this method that takes into account the contemporaneous relationship between several candidate series from the same climatic area. Instead of using the residuals from a linear regression model, the proposed technique uses the residuals from a seemingly unrelated regression equations (SUR) model, thus named SUR+Ellipse test.

Vincent (1998) proposed a multiple linear regression approach based on the application of four regression models to determine whether the tested series is homogeneous, has a trend, a single step, or trends before and/or after a step. The dependent variable is the series of the candidate station and the independent variables are the series of a number of surrounding stations. The first model determines whether the candidate series is homogeneous (in this case, the remaining models are not used). The data series is considered homogeneous if the residuals from the regression are independent normal variables with zero mean and constant variance. If there is a significant autocorrelation in the residuals (assessed by generalized Durbin–Watson tests and by the correlogram), then a second regression is calculated in which a linear trend is included. If the autocorrelation in the residuals of this second model exist, then the model is discarded and a third model is examined. The third regression is calculated for sequential increases in the time at which a step can occur. The minimum residual sum of squares from these regressions identifies the time of a break. If autocorrelations exist in the residuals from the regression with the step, a fourth model is considered. The last regression accounts for trends before and after the identified step. The existence of a trend provides an indication of multiple inhomogeneities in the candidate series. In this case, the series is subdivided at the position in time of the identified step and each segment is tested separately starting with the first model. The estimated parameters corresponding to steps and trends provide the magnitude of each inhomogeneity. Adjustments are then applied to bring each segment into agreement with the most recent homogeneous part of the series. The method proposed by Vincent (1998) has been recently improved by Reeves et al. (2007).

3.7 Standard Normal Homogeneity Test (SNHT)

The standard normal homogeneity test (Alexandersson 1986) is one of the most widely-used homogeneity tests. The null hypothesis is that the data are independent, identically normally distributed random quantities, and the alternative is that a step-wise shift in the mean (a break) is present. The SNHT is a likelihood ratio test that is usually performed on a composite reference series. There are now variations of this test to account for more than one discontinuity, testing for inhomogeneous trends rather than just breaks, and inclusion of change invariance (Alexandersson and Moberg 1997). The SNHT for a single break is capable of locating the period (month or year) where a break is likely, and it detects breaks near the beginning and the end of a series relatively easily (Ducré-Robitaille et al. 2003; Wijngaard et al. 2003).

3.8 Von Neumann Ratio Test

Von Neumann (1941) proposed a nonparametric test where the statistic is defined as the ratio of the mean square successive (year-to-year) difference to the variance. The null hypothesis is that the data are independent, identically distributed random quantities and the alternative is that the time series is not randomly distributed. Under the null hypothesis of a constant mean, the expected value of the test statistic is equal to two (Buishand 1982). The von Neumann ratio test is not location specific, which means that it gives no information about the date of the break.

3.9 Wald–Wolfowitz Runs Test

The Wald–Wolfowitz runs test (Wald and Wolfowitz 1943) is a well-known nonparametric test for randomness. The null hypothesis is that the process that generates the set of numerical data is random (with respect to the median) over time. For the two-tailed test, the alternative hypothesis is that the data set is not randomly distributed. For one of the one-tailed tests, the alternative hypothesis is that a trend effect is present in the data, whereas for the other one the alternative hypothesis is that a systematic or periodic effect is present in the data. This test is sensitive to shifts and trend, but gives little information about the probable dates for breaks (Tayanç et al. 1998).

4 Geostatistical Simulation Approach

A geostatistical approach, using direct sequential simulation (DSS), was proposed by Costa et al. (2008) for inhomogeneities detection. The DSS algorithm (Soares 2001) is used to calculate the local probability density function (pdf) at a candidate station's location, using spatial and temporal observations only from nearby reference stations, without taking into account the candidate's data. Afterwards, the local pdf from each instant in time (e.g., year) is used to verify the existence of irregularities. A breakpoint is identified whenever the interval of a specified probability p

(e.g., 0.95), centred in the local pdf, does not contain the observed (real) value of the candidate station. In practice, the local pdfs are provided by the histograms of simulated maps. Thus, this rule implies that if the observed (real) value lies below or above the pre-defined percentiles of the histogram of a given instant in time then it is not considered homogeneous. If irregularities are detected in a candidate series, the time series can be adjusted by replacing the inhomogeneous records with the mean, or median, of the pdf(s) calculated at the candidate station's location for the inhomogeneous period(s). The methodology of the geostatistical simulation approach can be summarized as follows.

4.1 Methodological Framework

Let $\{z(u_\alpha, t_i)\colon \alpha = 0, 1, \ldots, n-1; i = 1, \ldots, T\}$ be the set of climate data measured at n locations u_α and in t_i time instants (e.g., years). The n monitoring stations do not have to be all informed at the same T time instants (i.e., a number of z-values can be missing). Let $\{z(u_0, t_i)\colon i = 1, \ldots, T\}$ denote the candidate time series. The set of climate observations correspond to outcome values or realizations 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. Using the set of time series corresponding to the reference stations, $\{z(u_\alpha, t_i)\colon \alpha = 1, \ldots, n-1; i = 1, \ldots, T\}$, the DSS algorithm is applied in order to obtain a set of m equally probable realizations of $Z(u, t)$ at the candidate station's location and all instants in time: $\{z^k(u_0, t_i)\colon k = 1, \ldots, m; i = 1, \ldots, T\}$. For a given instant in time t_0, the set of simulated values $\{z^k(u_0, t_0)\colon k = 1, \ldots, m\}$ defines the local histogram of the candidate station at location u_0 for that instant t_0. 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: $F^*(u_0, t_0; z) = \text{Prob}^*\{Z(u_0, t_0) \le z\}$.

An inhomogeneous record $z(u_0, t_0)$ is identified if the interval of a specified probability p (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.

$$\begin{aligned} &\text{Prob}^*\{Z(u_0, t_0) \le z(u_0, t_0)\} < \frac{1-p}{2} \quad \text{or} \\ &\text{Prob}^*\{Z(u_0, t_0) \le z(u_0, t_0)\} > 1 - \frac{1-p}{2}. \end{aligned} \tag{1}$$

The proposed algorithm assumes that the global pdf used to generate realizations of $z^k(u_0, t_i)$ by DSS is representative of all monitoring stations—the candidate and the neighbors. In those situations where a candidate monitoring station has a pdf clearly different from the global pdf—local trends induced by topography or proximity to coast —then direct sequential simulation can be implemented with local distributions.

4.2 Direct Co-Simulation with Local Distributions

Among the sequential algorithms of stochastic simulation, one advantage of direct sequential simulation and co-simulation is precisely the use of original variables instead of the transformed Gaussian (sequential Gaussian simulation) or indicator (sequential indicator simulation). Direct sequential simulation and co-simulation have been applied in several soil and air quality characterization studies (Carvalho et al. 2006; Horta and Soares 2008). The use of original (non-transformed) variables and the generation of a simulated value by re-sampling the global pdf did open the door to new ways of this re-sampling approach. Carvalho et al. (2006) proposed to re-sample local distributions, taken from a secondary image instead of the global pdf, in an application of data fusion of satellite images. Horta and Soares (2008) proposed the re-sampling of joint distributions for co-simulation of a set of variables of a contaminated site. In this case, the first covariate $Z_1(u_\alpha)$ is simulated by direct sequential simulation at the location u_α. Based on previously simulated values $z_1^k(u_0)$, the conditional pdfs $F_z[Z_2(u_\alpha) \mid Z_1(u_\alpha) = z_1^k(u_0)]$ are calculated from the bi-distribution $F_{z_1,z_2}[Z_1(u_\alpha), Z_2(u_\alpha)]$. After estimating the local mean and variance at each location u_0, identified with the estimated simple collocated co-kriging and corresponding estimation variance $[z_2(u_0)]_{sck}$ and $\sigma_{sck}^2(u_0)$, the simulated value $z_2^k(u_0)$ is re-sampled from the conditional pdf $F_z[Z_2(u_\alpha) \mid Z_1(u_\alpha) = z_1^k(u_0)]$ as in the usual direct sequential procedure (Soares 2001).

In this study, to obtain a set of m equally probable realizations of $Z(u_\alpha, t)$ at the candidate station's location u_0 at a time period t_0, we suggest re-sampling the local pdf $F(u_0; z) = \text{Prob}\{Z(u_0) \le z\}$, rather than the global pdf. Hence, this new version of DSS with local distributions can be summarized as follows.

i) For each time period t_i, estimate the local cumulative distribution $F(u_0; z) = \text{Prob}\{Z(u_0) \le z\}$ at each candidate station location u_0, using past experimental data $z(u_0, t_j; j < i)$ and/or experimental data $z(u_\alpha, t_0)$ from a local neighborhood.
ii) At the location u_0 and time period t_0, estimate the local mean and variance, identified by simple kriging conditioned to other experimental data $z(u_\alpha, t_0) \ne z(u_0, t_0)$ and previous simulated values $z^k(u_0, t_0)$.
iii) The simulated value $z^k(u_\alpha, t_i)$ is re-sampled from the local pdf $F(u_\alpha, z)$ as in the usual direct sequential procedure (Soares 2001).

5 Case Study Results

The inhomogeneities detection procedures used in the case study followed the hybrid approach proposed by Wijngaard et al. (2003) for the European Climate Assessment (ECA) dataset, and used as testing variable the number of wet days per year, where a wet day is defined as a day with at least 1 mm of precipitation. For illustration purposes, the geostatistical simulation approach was applied to the testing variable data from 4 candidate stations using data from 62 surrounding stations (reference stations) located in the southern region of Portugal (Fig. 1). The daily precipitation series were compiled from the ECA dataset and the National System of Water Resources Information (SNIRH—Sistema Nacional de Informação de Recursos Hídricos) database, and are available through free downloads from the ECA website (http://eca.knmi.nl) and the SNIRH website (http://snirh.inag.pt), respectively. Some authors run the relative tests once, relying that the reference series are homogeneous, while others engage

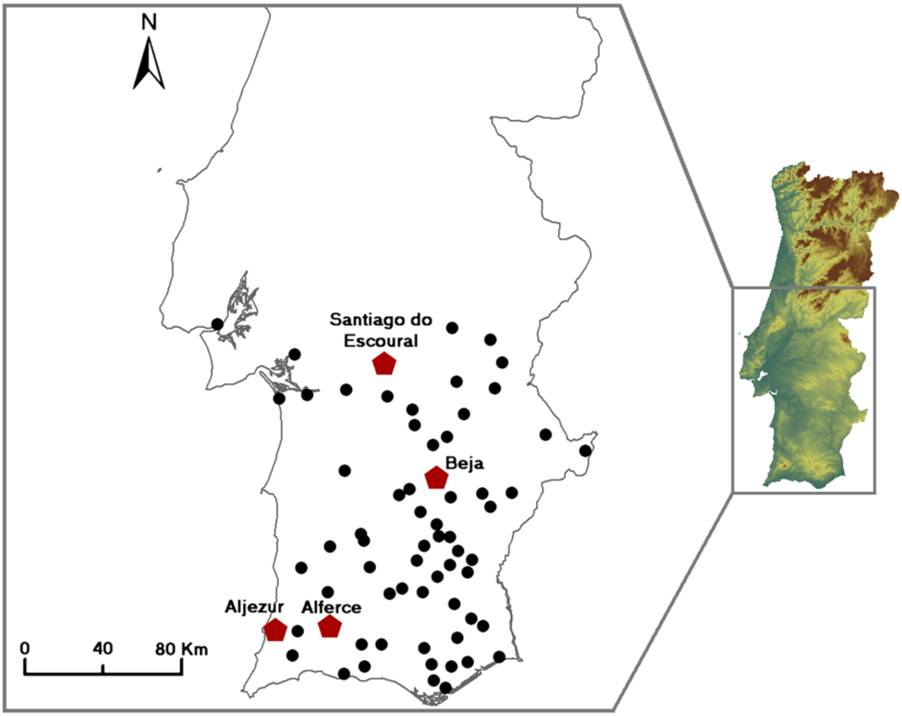

Fig. 1 Locations of the 66 monitoring stations in the southern region of Portugal; candidate stations are marked with *pentagons*

in an iterative procedure in which all the stations in the data set are seen consecutively as candidates and references (Aguilar et al. 2003). Following the latter methodology, the local pdfs of each year of the candidate series, derived from 50 simulated maps in 1 km × 1 km grid cells, were computed using data not only from the 62 references but also from the other 3 candidate stations. The analyzed period was 1980–2001. The geostatistical simulations used a spherical semivariogram model fitted to the testing variable data from the complete set of 66 monitoring stations. The spatial dimension was modeled using an isotropic semivariogram with a range of 72 km, and the temporal one with the range equal to 1.8 years.

The results from the geostatistical simulation procedure were compared to the results from three well-established homogeneity tests that used two reference series for each candidate and the full length of the series. The SNHT, Pettit and Buishand range tests were applied to composite (ratio) reference series (Alexandersson and Moberg 1997), which were derived from the testing variable data. Further methodological details and additional results from this approach are described by Costa and Soares (2006). The geostatistical approach allowed identifying several inhomogeneities by comparing the observed (real) values of the candidate series, for each year, with the 2.5% and the 97.5% percentiles of the corresponding histograms of 50 simulated maps ($p = 0.95$ in (1)). This methodology identified not only the same break years (or within one-year range) as the other three testing procedures, but also revealed

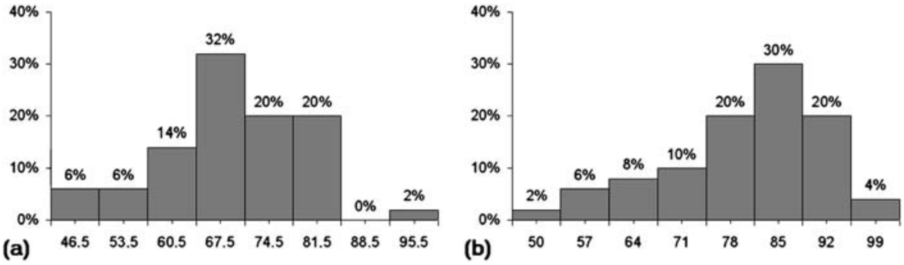

Fig. 2 Histograms of the 50 simulated realizations of the annual wet day count at Santiago do Escoural location for 1988 (**a**) and 1996 (**b**), computed without data from Santiago do Escoural (the real values are 86 days in 1988, and 96 days in 1996)

inhomogeneities in other years that were not detected by any of the three statistical tests at a 5% significance level.

For Aljezur (SNIRH 30E.01) station, the four approaches considered the series as homogeneous. The series from Beja (ECA 666) was considered as homogeneous by the three statistical tests, whereas the geostatistical approach identified a break in 1991. For Alferce (SNIRH 30G.01) station, the SNHT concluded the series as homogeneous, but the Buishand and Pettit tests detected a break in 1984. Similarly, the geostatistical approach identified a breakpoint in 1983. In fact, at Alferce, the minimum simulated realization of the annual wet day count in 1983 was equal to approximately 33 days, whereas the observed value for this series was 30 days, thus a breakpoint was detected in this year. The candidate series from Santiago do Escoural (SNIRH 22H.02) was considered as inhomogeneous by all techniques. The SNHT detected a break in 1989, the Buishand and Pettit tests identified a breakpoint in 1988, and the geostatistical technique detected breaks in 1987, 1988 (Fig. 2(a)) and 1996 (Fig. 2(b)). At Santiago do Escoural, the maximum simulated realization of the annual wet day count in 1987 was equal to 89 days, whereas the observed value was 91 days, so a break was detected in 1987 by the geostatistical simulation approach.

All break years identified by the three well established statistical tests considered were also detected by the new technique. Moreover, the geostatistical technique allowed for the identification of breaks near the end of the series that were not detected by the other methods. These promising results indicate the geostatistical simulation approach as a valuable tool for inhomogeneities detection in climate time series.

6 Discussion and Perspectives

More than one undocumented inhomogeneity may be present in a climate time series. In the ideal case, all possible breakpoints should be identified jointly before their mean shift magnitudes are estimated (Reeves et al. 2007). However, the number of multiple breakpoints detection procedures is limited, thus this is an active current area of statistical research (Reeves et al. 2007). Consequently, the tests for single break detection are sometimes used iteratively by systematically dividing the tested series into smaller segments when a break is detected, and then performing the test on those segments. Techniques that use series from surrounding stations, sometimes run the test once, relying the reference to be homogeneous, or engage in an iterative procedure in which all stations in the data set are seen consecutively as candidates and references (Aguilar et al. 2003; Auer et al. 2005). Procedures based on test iteration such as those are powerful but computationally intensive, and thus can be time consuming and exacting work.

The multiple analysis of series for homogenization (MASH) method seems to be one of the most comprehensive procedures for homogenization, although the MASH system may not be readily available for many researchers. Wijngaard et al. (2003) state that, generally, a combination of statistical methods and methods relying on metadata information is considered to be the most effective when tracking down inhomogeneities. Menne and Williams (2005) evaluated three hypothesis test statistics to ascertain whether multiple tests can be combined to improve overall confidence in undocumented inhomogeneities detection. These authors also evaluated different composite reference series formulations. Using Monte Carlo case studies, Menne and Williams (2005) concluded that for reasonably well correlated time series and homogeneous references, the choice of reference series formulation has relatively little impact on candidate series inhomogeneities detection skill and, consequently, the choice of the test statistic has a greater impact. However, Menne and Williams (2005) argue that those circumstances are probably rare in practice; thus, the choice of reference series formulation has implications that are more important in breakpoints detection than the choice of the test statistic. In fact, the common period of observations between the candidate series and time series from neighboring stations might be too short to properly select and weight the individual series and construct a reliable composite reference series. Moreover, if too many distant (or less correlated) neighboring stations are used, the resulting reference may not reflect the true climatic signal of the candidate station properly (Boissonnade et al. 2002). These difficulties may increase dramatically with the increase in spatial variability of the data caused by the inherent variability of the element (e.g., precipitation), the time series resolution (e.g., monthly data) or the network location (e.g., Mediterranean region).

Nonparametric tests are not as efficient or powerful as the equivalent parametric procedures, provided that the underlying assumptions are satisfied because they have less information to use to determine significance. For this reason, the Normal assumption is usually relaxed for annual climatic data, and parametric tests such as the SNHT are sometimes preferred. Most of the statistical procedures, including nonparametric tests, require serially independent data. When sample data is serially correlated, the presence of serial correlation in time series will affect the ability of the tests to correctly assess the significance of inhomogeneities detection. However, it is a standard procedure to relax this hypothesis for annual data. The autocorrelation assumption, while acceptable for some annual climate series, is not realistic for daily or monthly series where there is much empirical evidence of autocorrelation (Reeves et al. 2007). Both parametric and nonparametric tests described above look at one or a few of the characteristics of a frequency distribution. These characteristics do not include nonlinear effects, nor do they consider non-climatic influences that affect data in a nonuniform manner, such as certain weather events, seasons, etc. The daily data may reflect more of a mixture of populations and are more likely to be affected by nonlinear and nonuniform weather events than data that is averaged over a longer time interval. In order to overcome these difficulties, the homogenization of high temporal resolution climate databases is usually performed by using traditional procedures with monthly or annual totals, or other variables derived from the daily series (e.g., Wijngaard et al. 2003; Feng et al. 2004).

As stated before, further investigation is required to develop procedures for the homogenization of sub-monthly climate data. The geostatistical simulation approach seems to be a very promising procedure for this research field, as kriging techniques have proven to succeed in the estimation of missing daily precipitation records (e.g., Kyriakidis et al. 2004; Teegavarapu and Chandramouli 2005). Moreover, multivariate simulation algorithms might be used for the homogenization of highly variable elements, such as precipitation, making use of information from explanatory physiographic variables (e.g., elevation). Further research on this subject is clearly required. The geostatistical simulation approach has also a number of potential advantages over the traditional approaches used for the homogenization of data with lower temporal resolution. Geostatistical techniques allow dealing with the problem of missing values and varying availability of stations through time, by using different sets of neighboring stations at different time periods (years, months, etc.), and by including shorter and incomplete records. Additionally, the geostatistical approach avoids the iterative construction of composite reference series because it increases the contribution of records from closer stations, both in spatial and correlation terms, by accounting for the joint spatial and temporal dependence between observations. Multiple breaks can be detected simultaneously; thus, this method might be less time consuming than other testing techniques that are used iteratively. Another advantage is that the geostatistical approach seems to be able to identify breakpoints near the start and end of the time series, while traditional approaches have less power in detecting them (Aguilar et al. 2003).

The inherently high (temporal and spatial) variability of precipitation makes homogenization of precipitation records more difficult to accomplish than other elements (e.g., temperature). Therefore, it is reasonably intuitive that the number of simulated realizations, used by the geostatistical approach to infer the local probability density functions (pdfs), should be higher for precipitation data than for temperature records. But how many simulated realizations should be used to accurately infer the local pdfs? This and other issues require further investigation and open new perspectives on the homogenization of climate data.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J (2003) Guidelines on climate metadata and homogenization. World Meteorological Organization, WMO-TD No. 1186, WCDMP No. 53, Geneva, Switzerland, 55

Alexandersson H (1986) A homogeneity test applied to precipitation data. J Climatol 6:661–675

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

Allen RG, Pereira LS, Raes D, Smith M (1998) Statistical analysis of weather data sets. In: Food and Agriculture Organization of the United Nations (ed) Crop evapotranspiration – Guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56, Rome, Annex IV

Auer I, Böhm R, Jurković A, Orlik A, Potzmann R, Schöner W, Ungersböck M, Brunetti M, Nanni T, Maugeri M, Briffa K, Jones P, Efthymiadis D, Mestre O, Moisselin J-M, Begert M, Brazdil R, Bochnicek O, Cegnar T, Gajić-Čapka M, Zaninović K, Majstorović, Ž, Szalai S, Szentimrey T, Mercalli L (2005) A new instrumental precipitation dataset for the greater alpine region for the period 1800–2002. Int J Climatol 25(2):139–166

Boissonnade AC, Heitkemper LJ, Whitehead D (2002) Weather data: cleaning and enhancement. In: Dischel RS (ed) Climate risk and the weather market: financial risk management with weather hedges. Risk Waters, pp 73–93

Buishand TA (1982) Some methods for testing the homogeneity of rainfall records. J Hydrol 58:11–27

Carvalho J, Delgado-García J, Soares A (2006) Merging Landsat and SPOT digital data using stochastic simulation with reference images. In: Caetano M, Painho M (eds) Proceedings of accuracy 2006 – 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences. Instituto Geográfico Português, pp 567–577

Costa AC, Soares A (2006) Identification of inhomogeneities in precipitation time series using SUR models and the Ellipse test. In: Caetano M, Painho M (eds) Proceedings of accuracy 2006 – 7th international symposium on spatial accuracy assessment in natural resources and environmental sciences. Instituto Geográfico Português, pp 419–428

Costa AC, Negreiros J, Soares A (2008) Identification of inhomogeneities in precipitation time series using stochastic simulation. In: Soares A, Pereira MJ, Dimitrakopoulos R (eds) geoENV VI – Geostatistics for Environmental Applications. Springer, Berlin, pp 275–282

Craddock JM (1979) Methods of comparing annual rainfall records for climatic purposes. Weather 34:332–346

Ducré-Robitaille J-F, Vincent LA, Boulet G (2003) Comparison of techniques for detection of discontinuities in temperature series. Int J Climatol 23(9):1087–1101

Easterling DR, Peterson TC (1995) A new method for detecting and adjusting for undocumented discontinuities in climatological time series. Int J Climatol 15:369–377

Feng S, Hu Q, Qian W (2004) Quality control of daily meteorological data in China, 1951–2000: a new dataset. Int J Climatol 24(7):853–870

Hirsch RM, Slack JR (1984) A nonparametric trend test for seasonal data with serial dependence. Water Resour Res 20(6):727–732

Hirsch RM, Slack JR, Smith RA (1982) Techniques of trend analysis for monthly water quality data. Water Resour Res 18(1):107–121

Horta A, Soares A (2008) Data integration model for soil degradation risk assessment. In: Proceedings of the 8th international geostatistics congress, 1–5 December 2008, Santiago, Chile

Kendall MG (1975) Rank correlation methods. Charles Griffin, London

Kohler MA (1949) Double-mass analysis for testing the consistency of records and for making adjustments. Bull Am Meteorol Soc 30:188–189

Kruskal WH (1952) A nonparametric test for the several sample problem. Ann Math Stat 23:525–540

Kruskal WH, Wallis WA (1952) Use of ranks in one-criterion variance analysis. J Am Stat Assoc 47:583–621

Kyriakidis PC, Miller NL, Kim J (2004) A spatial time series framework for simulating daily precipitation at regional scales. J Hydrol 297:236–255

Lettenmaier DP (1988) Multivariate nonparametric tests for trend in water quality. Water Resour Bull 24(3):505–512

Libiseller C, Grimvall A (2002) Performance of partial Mann-Kendall test for trend detection in the presence of covariates. Environmetrics 13:71–84

Mann HB (1945) Non-parametric test against trend. Econometrika 13:245–259

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60

Menne MJ, Williams CN (2005) Detection of undocumented changepoints using multiple test statistics and composite reference series. J Climate 18(20):4271–4286

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. Int J Climatol 18(13):1493–1517

Pettit AN (1979) A non-parametric approach to the change-point detection. Appl Statist 28(2):126–135

Potter KW (1981) Illustration of a new test for detecting a shift in mean in precipitation series. Mon Weather Rev 109:2040–2045

Reeves J, Chen J, Wang XL, Lund R, Lu Q (2007) A review and comparison of changepoint detection techniques for climate data. J Appl Meteorol Clim 46:900–915
Romero R, Guijarro JA, Ramis C, Alonso S (1998) A 30-year (1964–1993) daily rainfall data base for the Spanish Mediterranean regions: First exploratory study. Int J Climatol 18(5):541–560
Soares A (2001) Direct sequential simulation and cosimulation. Math Geol 33(8):911–926
Solow A (1987) Testing for climatic change: an application of the two phase regression model. J Clim Appl Meteorol 26:1401–1405
Szentimrey T (1994) Statistical problems connected with the homogenization of climatic time series. In: Heino R (ed) Climate variations in Europe. Proceedings of the European workshop held in Kirkkonummi (Majvik), Finland, May 1994, Publications of the Academy of Finland, pp 330–339
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-TD No. 962, WCDMP No. 41, pp 27–46
Szentimrey T (2003) Multiple analysis of series for homogenization (MASH); Verification procedure for homogenized time series. In: Fourth seminar for homogenization and quality control in climatological databases. Budapest, Hungary, WMO-TD No. 1236, WCDMP No. 56, pp 193–201
Tayanç M, Dalfes HN, Karaca M, Yenigün O (1998) A comparative assessment of different methods for detecting inhomogeneities in Turkish temperature data set. Int J Climatol 18(5):561–578
Teegavarapu RSV, Chandramouli V (2005) Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. J Hydrol 312:191–206
Tuomenvirta H (2001) Homogeneity adjustments of temperature and precipitation series – Finnish and Nordic data. Int J Climatol 21(4):495–506
Vincent L (1998) A technique for the identification of inhomogeneities in Canadian temperature series. J Climate 11:1094–1104
Von Neumann J (1941) Distribution of the ratio of the mean square successive difference to the variance. Ann Math Stat 13:367–395
Wald A, Wolfowitz J (1943) An exact test for randomness in the non-parametric case based on serial correlation. Ann Math Stat 14:378–388
Wijngaard J, Klein Tank AMG, Können GP (2003) Homogeneity of 20th century European daily temperature and precipitation series. Int J Climatol 23(6):679–692
Wilcoxon F (1945) Individual comparison by ranking methods. Biometrics 1:80–83
Yue S, Wang CY (2004) The Mann-Kendall test modified by effective sample size to detect trend in serially correlated hydrological series. Water Resour Manag 18(3):201–218