Atmospheric Research 171 (2016) 147–158

Contents lists available at ScienceDirect

Atmospheric Research

journal homepage: www.elsevier.com/locate/atmosres

Detection of inhomogeneities in precipitation time series in Portugal using direct sequential simulation

Sara Ribeiro [a,*], Júlio Caineta [a], Ana Cristina Costa [a], Roberto Henriques [a], Amílcar Soares [b]

[a] NOVA IMS, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisboa, Portugal
[b] CERENA — Centro de Recursos Naturais e Ambiente, Instituto Superior Técnico, Avenida Rovisco Pais, 1049-001 Lisboa, Portugal

ARTICLE INFO

Article history:
Received 30 December 2014
Received in revised form 13 October 2015
Accepted 24 November 2015
Available online 7 December 2015

Keywords:
Discontinuities
Homogenization
Homogeneity tests
Geostatistical simulation

ABSTRACT

Climate data homogenisation is of major importance in climate change monitoring, validation of weather forecasting, general circulation and regional atmospheric models, modelling of erosion, drought monitoring, among other studies of hydrological and environmental impacts. The reason is that non-climate factors can cause time series discontinuities which may hide the true climatic signal and patterns, thus potentially bias the conclusions of those studies. In the last two decades, many methods have been developed to identify and remove these inhomogeneities. One of those is based on a geostatistical simulation technique (DSS — direct sequential simulation), where local probability density functions (pdfs) are calculated at candidate monitoring stations using spatial and temporal neighbouring observations, which then are used for the detection of inhomogeneities. Such approach has been previously applied to detect inhomogeneities in four precipitation series (wet day count) from a network with 66 monitoring stations located in the southern region of Portugal (1980–2001). That study revealed promising results and the potential advantages of geostatistical techniques for inhomogeneity detection in climate time series. This work extends the case study presented before and investigates the application of the geostatistical stochastic approach to ten precipitation series that were previously classified as inhomogeneous by one of six absolute homogeneity tests (Mann–Kendall, Wald–Wolfowitz runs, Von Neumann ratio, Pettitt, Buishand range test, and standard normal homogeneity test (SNHT) for a single break). Moreover, a sensitivity analysis is performed to investigate the number of simulated realisations which should be used to infer the local pdfs with more accuracy. Accordingly, the number of simulations per iteration was increased from 50 to 500, which resulted in a more representative local pdf. As in the previous study, the results are compared with those from the SNHT, Pettitt and Buishand range tests, which were applied to composite (ratio) reference series. The geostatistical procedure also allowed us to fill in missing values in the climate data series. Finally, based on several experiments aimed at providing a sensitivity analysis of the procedure, a set of default and recommended settings is provided, which will help other users to apply this method.

1. Introduction

Several environmental and atmospheric studies depend on climate data, in which precipitation data assume a vital role. However, its measurement and recording is prone to systematic and random errors (Sevruk et al., 2009; Teegavarapu and Chandramouli, 2005). Systematic errors may occur due to the growth of trees or urbanisation around the location of the weather station or to precipitation gauge malfunctions, such as water loss during measurement, adhesion loss on the surface of the gauge and raindrop splash from the collector. Random errors include sporadic faults which happen during the process of collecting, recording and transmitting precipitation data records (Brunet and Jones, 2011). These non-natural errors are critical as they affect the continuity of precipitation data and ultimately influence the results of models that use precipitation as input. Indices calculated from daily precipitation data, such as the number of wet days per year (wet day count), are also influenced by the errors in the measurement. Spurious shifts often have the same magnitude as the climate signal, such as long-term variations, trends or cycles, and might lead to wrong considerations about the results of the studies (Caussinus and Mestre, 2004).

In order to obtain trustful results, climate data should be free from non-climatic irregularities. Hence, the detection and the correction of these errors are absolutely necessary before any reliable climate study is based on instrumental series (Auer et al., 2005; Brunetti et al., 2012; Domonkos, 2013; Tuomenvirta, 2001). Moreover, the World Meteorological Organization (WMO) emphasises the importance of homogenisation in one of the ten climate monitoring principles: "The quality and homogeneity of data should be regularly assessed as a part of routine operations." (World Meteorological Organization, 2010).

---

[*] Corresponding author. Tel.: +351 213 828 610.
E-mail addresses: sribeiro@novaims.unl.pt (S. Ribeiro), jcaineta@novaims.unl.pt (J. Caineta), ccosta@novaims.unl.pt (A.C. Costa), roberto@novaims.unl.pt (R. Henriques), asoares@tecnico.ulisboa.pt (A. Soares).

http://dx.doi.org/10.1016/j.atmosres.2015.11.014
0169-8095/

Homogenisation includes the following steps (Štěpánek et al., 2006): detection, verification and possible correction of outliers, creation of reference series, homogeneity testing (through various homogeneity tests), determination of inhomogeneities in the light of test results and metadata, adjustment of inhomogeneities and filling in missing values. Various methods have been used in the homogenisation of climate data (Aguilar et al., 2003; Beaulieu et al., 2008; Domonkos et al., 2012; Peterson et al., 1998), and their efficiency is dependent on the climate variable, analysed time period, availability of data or other stations located in the same climatic region which may be used as reference series (Costa and Soares, 2009b). Homogenisation methods can be classified into different groups, depending on their characteristics (Aguilar et al., 2003): objective/subjective, direct/indirect and absolute/relative. Relative methods make use of data from neighbouring stations (called reference stations) for comparison with data series from the candidate station (the station to be homogenised). Absolute methods only consider the data from the candidate station in the detection of inhomogeneities.

Recently, the European initiative (COST Action ES0601) 'HOME' (Advances in homogenisation methods of climate series: an integrated approach), evaluated the performance of a set of statistical homogenisation methods, using a benchmark data set of temperature and precipitation. Due to their excellent performance, the algorithms ACMANT, Craddock, MASH, PRODIGE and USHCN are strongly recommended by Venema et al. (2012). These authors also refer the need to give priority to the homogenisation of precipitation, due to the less good results presented by the contributions for precipitation. Moreover, Domonkos et al. (2012) mention the need of further tests to better understand the performance of homogenisation methods. Due to the diversity of the characteristics of climatic time series, it is essential to perform more tests with different data set properties. These authors provide a thorough literature review on the methodological evolution of the homogenisation methods for temperature. Ribeiro et al. (2015) compare homogenisation methods based on literature reviews and discuss their advantages and disadvantages.

Craddock test (Craddock, 1979) accumulates the normalised differences between the test series and the homogeneous reference series in order to find inhomogeneities. This author applied the method to precipitation time series and concluded that best results were obtained by the use of station pairs with the minimum coefficient of variation of the ratio of the two series. This test is part of the homogenisation package THOMAS, from the Federal Office of Meteorology and Climatology in Switzerland (Begert et al., 2005; Michael Begert, 2015, personal communication).

MASH, Multiple Analysis of Series for Homogenisation (Szentimrey, 1999, 2007) is a homogenisation method originally developed for monthly series. This relative method does not assume reference series as homogeneous. It is a multiple breakpoint detection algorithm that increases its performance taking the problem of significance and efficiency in account. Metadata is used automatically, in particular the possible dates of breakpoints. The algorithm also includes a procedure for the evaluation of the homogenisation results. In the version of the MASH algorithm for daily data, the estimation of daily inhomogeneities is based on the monthly inhomogeneities calculated (Lakatos et al., 2008).

Caussinus and Mestre (2004) introduced a new methodology for the detection of inhomogeneities, which included pairwise comparison, step function fitting, the Caussinus and Lyazrhi (1997) algorithm, and variance optimisation. This method, later named PRODIGE, is based on the idea that a series is homogeneous between two change points. Pairwise comparisons are then obtained between the candidate series and the other reference series, creating a series of differences. These series are tested against the Caussinus and Lyazrhi technique. If a common breakpoint is detected in all the difference series, it is attributed to the candidate station. The overall detection and correction are performed by moving neighbourhoods. The correction estimation is based on ANOVA.

ACMANT, Adapted Caussinus–Mestre Algorithm for Networks of Temperature Series (Domonkos, 2011; Domonkos et al., 2011), is a fully automated and relative homogenisation method, which uses the core of the detection and adjustment methods of the PRODIGE (step function fitting and ANOVA correction segments). It applies a bivariate-test for detecting change points that uses the annual mean and the summer–winter difference.

The USHCN homogenisation method is another automatic homogenisation method applied to the United States Historical Climatology Network (Menne and Williams, 2009). The detection part of this method is composed by an early version of SNHT, the cutting algorithm, a Bayesian-based decision about the form of the inhomogeneities (trend-like inhomogeneities can be detected), and a special purpose significance test. Pairwise comparisons are made in an automated way, and metadata can also be used automatically.

The present study provides a follow-up of a previous study (Costa and Soares, 2009b), where a new detection methodology based on direct sequential simulation (DSS) was tested with very auspicious results. However, due to technology and time limitations, a small number of simulations were performed at that time and the number of candidate series was limited to four. In this study, the number of simulations is increased, some sensitivity experiments are performed, and some conclusions are drawn regarding those analyses. For comparison purposes, the same data set was used, which is composed of 66 stations located in the south of Portugal. The analysed climate variable is the annual number of wet days (threshold of 1 mm), calculated from the measured daily value of precipitation, at each weather station, per year. Two sets of candidate stations are used in different stages of the study: the first set, composed of 4 stations, is used for the sensitivity analysis of the DSS parameters; the second set, comprising 10 stations, is used for the sensitivity analysis of the number of neighbour nodes used in the simulation of each node.

The results of the analysis of both sets of candidate stations are compared with the results achieved by Costa and Soares (2009b) through the Standard normality homogenisation test (SNHT, Alexandersson, 1986), the Buishand range test (Buishand, 1982) and the Pettitt test (Pettitt, 1979). These techniques are commonly used and generally accepted for the detection of inhomogeneities (e.g., Sahin and Cigizoglu (2010); Santos and Fragoso, 2013; Wijngaard et al., 2003). Pandžić and Likso (2010) indicate SNHT as one of the most popular methods. Wijngaard et al. (2003) make a brief description of the advantages and disadvantages of those three tests.

Section 2 details the network used in this study. Section 3 briefly describes the methodological framework, particularly the DSS process and the sensitivity analysis methodology. Results are presented in Section 4. Finally, some conclusions and future work are stated in Section 5.

2. Data and study background

The inhomogeneity detection methods were applied to precipitation data from 66 monitoring stations located in the south of Portugal (Fig. 1). The annual number of wet days between 1980 and 2001 was used as the studied variable, which was calculated from the daily values of precipitation measured at each station, with a threshold of 1 mm defining a wet day. The annual wet day count was used because it is expected to be representative of important characteristics of variation at the daily scale (Wijngaard et al., 2003). This is one of the extreme climate indices defined by the joint CCl/CLIVAR/JCOMM Expert Team (ET) on Climate Change Detection and Indices (ETCCDI), which may contribute to gain a uniform perspective on observed changes in climate extremes (e.g., Klein Tank et al., 2009). The analysis of changes in climate extremes usually requires daily resolution data, but well-established statistical methods for homogeneity testing daily precipitation data are lacking. According to Wijngaard et al. (2003), this variable generally has a lower variability than the annual amounts, particularly in areas with a large contribution from