Physics and Chemistry of the Earth 94 (2016) 167–179

Contents lists available at ScienceDirect

Physics and Chemistry of the Earth

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

Review and discussion of homogenisation methods for climate data

S. Ribeiro*, J. Caineta, A.C. Costa

NOVA IMS, Universidade Nova de Lisboa, Portugal

A R T I C L E I N F O

Article history:
Received 20 March 2015
Received in revised form
24 June 2015
Accepted 11 August 2015
Available online 28 August 2015

Keywords:
Methods comparison
Data quality
Irregularities
Trends
Homogenization

A B S T R A C T

The quality of climate data is of extreme relevance, since these data are used in many different contexts. However, few climate time series are free from non-natural irregularities. These inhomogeneities are related to the process of collecting, digitising, processing, transferring, storing and transmitting climate data series. For instance, they can be caused by changes of measuring instrumentation, observing practices or relocation of weather stations. In order to avoid errors and bias in the results of analysis that use those data, it is particularly important to detect and remove those non-natural irregularities prior to their use. Moreover, due to the increase of storage capacity, the recent gathering of massive amounts of weather data implies also a toilsome effort to guarantee its quality. The process of detection and correction of irregularities is named homogenisation. A comprehensive summary and description of the available homogenisation methods is critical to climatologists and other experts, who are looking for a homogenisation method wholly considered as the best. The effectiveness of homogenisation methods depends on the type, temporal resolution and spatial variability of the climatic variable. Several comparison studies have been published so far. However, due to the absence of time series where irregularities are known, only a few of those comparisons indicate the level of success of the homogenisation methods. This article reviews the characteristics of the most important procedures used in the homogenisation of climatic variables based on a thorough literature research. It also summarises many methods applications in order to illustrate their applicability, which may help climatologists and other experts to identify adequate method(s) for their particular needs. This review study also describes comparison studies, which evaluated the efficiency of homogenisation methods, and provides a summary of conclusions and lessons learned regarding good practices for the use of homogenisation methods.

Contents

* Corresponding author at: NOVA IMS, Universidade Nova de Lisboa, Campus de Campolide, 1070-312 Lisbon, Portugal.
E-mail addresses: sribeiro@novaims.unl.pt (S. Ribeiro), jcaineta@novaims.unl.pt (J. Caineta), ccosta@novaims.unl.pt (A.C. Costa).

http://dx.doi.org/10.1016/j.pce.2015.08.007
1474-7065/

1. Introduction

Success in atmospheric modelling, weather forecasting or climate change monitoring depends on the quality of climate data used as input. Long time series without artificial discontinuities in their statistical characteristics are rare (Alexandersson and Moberg, 1997). Those irregularities can be due to climatic factors, or can be related to facts that happened during the process of collecting or recording climate data. Examples of climatic factors are the eruption of a volcano and the emission of its gases and ashes to the atmosphere contributing to the decrease of solar radiation, or the effect of the North Atlantic Oscillation in extreme temperature and precipitation records across Europe (Gaffen et al., 2000).

Non-climatic factors may introduce abrupt or gradual changes in the time series (Alexandersson and Moberg, 1997). Examples of the former are changes in the method of measuring and calculating climate values, such as the use of different daily times in the calculation of daily mean temperature (Peterson et al., 1998), change of measurement units (K, °C and °F for temperature) without any notice (Aguilar et al., 2003), changes in the formula for calculation of the variable's average (Puglisi et al., 2010), relocation of a station (Venema et al., 2013), or its repositioning to a different height (Auer et al., 2005). Gradual and soft changes can be exemplified by the presence of a tree or bush growing nearby the weather station, or the development of an urban area on its surroundings – the increasing of nocturnal temperature called the "Urban Heat Island Effect" (Brunet et al., 2006; Li et al., 2004; Sahin and Cigizoglu, 2010). A high number of non-natural irregularities are also introduced during the process of collecting, digitising, processing, transferring, storing and transmitting climate data series (Brunet and Jones, 2011).

These non-climatic factors may introduce artificial discontinuities, or inhomogeneities, in the time series. Such discontinuities can lead to misinterpretations of the studied climate. In order to avoid errors and obtain homogeneous climate time series, non-natural irregularities in climate data series must be detected and removed prior to its use.

Three main types of inhomogeneities can be distinguished: point errors (coming from the observation to transmission and mechanisation processes); breakpoints corresponding to change-points or shifts in the mean (changes of location, instrumentation, observing practices or land use of the surroundings); and trends (sensor decalibration or urban growth) (Guijarro, 2006). Breakpoints are the most frequent form of inhomogeneities, since most technical changes happen abruptly (Domonkos, 2011a). Trend inhomogeneities are generally more difficult to detect, because they may be superimposed on a true climate trend (Easterling and Peterson, 1995).

Homogenisation is known as the process of detecting and correcting inhomogeneities (Aguilar et al., 2003). Another definition is provided by Štěpánek et al. (2006), where homogenisation includes the following steps: detection, verification and possible correction of outliers, creation of reference series, homogeneity testing (various homogeneity tests), determination of inhomogeneities in the light of test results and metadata, adjustment of inhomogeneities and filling in missing values. Mathematics, software and metadata are referred by Szentimrey (2011) as indispensable for homogenisation of climate data.

Recently, the importance of studying extremes of weather and climate required the development of homogenisation methods for climate data series with higher temporal resolution (e.g., daily data) (Brunetti et al., 2012). In case of precipitation, this task became a challenge due to its great variability (Rustemeier et al., 2011). This variability also results in great uncertainty in homogenisation. True climatic fluctuations in daily precipitation may be interpreted as change-points and removed from time series as inhomogeneities. Moreover, the magnitude of inhomogeneities may differ with varying weather situations (Nemec et al., 2013). Another problem is associated with errors linked to the measuring process, particularly during extreme weather events. For example, larger adjustments are likely to be required for precipitation as its recording is strongly affected by wind strength (Auer et al., 2005). Systematic underestimation of snowfall is also a serious problem in areas where a substantial part of precipitation is collected by rain gauges as snow (Auer et al., 2005; Eccel et al., 2012). To overcome these issues, daily homogenisation methods require complex techniques or the improvement of homogenisation methods previously used for monthly and annual climate series. Those homogenisation methods are of paramount importance as those series are the basis for political decisions with socio-economic consequences (Venema et al., 2013).

The present review provides a description and discussion of homogenisation methods for climate data series, and summarises the conclusions of some comparison studies undertaken to assess their efficiency. Section 2 addresses the classification of homogenisation methods, Section 3 comprises a review of the available homogenisation methods, and Section 4 presents several homogenisation software packages. Comparison studies are briefly described in Section 5, where it is also given focus to the HOME project (COST Action ES0601). Finally, some conclusions are drawn in Section 6.

2. Approaches for detecting and correcting inhomogeneities

Homogenisation methods may have different characteristics, depending on the use of metadata, the subjectivity involved, the use of additional climate time series, the capability of detecting multiple breakpoints, etc. Those characteristics are discussed in the following subsections.