INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. **31**: 630–632 (2011)
DOI: 10.1002/joc.2088

Royal Meteorological Society

Short Communication

A note on the use of the standard normal homogeneity test to detect inhomogeneities in climatic time series

A. Toreti,[a,b]* F. G. Kuglitsch,[a] E. Xoplaki,[a,c] P. M. Della-Marta,[d] E. Aguilar,[e] M. Prohom[f] and J. Luterbacher[g]

[a] *Oeschger Centre for Climate Change Research (OCCR) and Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland*
[b] *Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), Rome, Italy*
[c] *The Cyprus Institute, Energy, Environment and Water Research Center (EEWRC), Nicosia, Cyprus*
[d] *Federal Office for Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland*
[e] *Climate Change Research Group, Geography Unit, Universitat Rovira i Virgili de Tarragona, Tarragona, Spain*
[f] *Meteorological Service of Catalonia, Area of Climatology, Generalitat de Catalunya, Barcelona, Catalonia, Spain*
[g] *Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus-Liebig University of Giessen, Giessen, Germany*

ABSTRACT: Homogenization methods are developed to reduce the impact of non-climatic factors on climate series. Martínex *et al.* (2009), (*International Journal of Climatology*, Doi 10.1002/joc.1884) applied a set of homogenization procedures to available Spanish temperature series. In this report, we address critical issues of that paper concerning a specific property of the standard normal homogeneity test and the application scheme of the homogenization tests. We conclude with some important recommendations on the application of homogenization methodologies. Copyright © 2010 Royal Meteorological Society

KEY WORDS homogenization; SNHT; temperature

Received 22 June 2009; Revised 26 November 2009; Accepted 18 December 2009

Martínez *et al.* (2009) recently published an article on daily maximum and minimum temperature ($T_{\max}$, $T_{\min}$) and diurnal temperature range (DTR) recorded at 37 meteorological stations in Catalonia (NE Spain) covering the period 1975–2004. After a reliable quality control procedure, they carried out a set of four different homogeneity tests: (1) standard normal homogeneity test (SNHT; Alexandersson, 1986; Alexandersson and Moberg, 1997a), (2) Buishand range (Buishand, 1982), (3) Pettitt test (Pettitt, 1979) and (4) Von Neumann ratio test (Von Neumann, 1941). The authors did not correct series but decided to reject them in case of inhomogeneity detection. In the description of the homogenization procedure, there are a few misleading points that we would like to discuss in this short comment.

The thorough understanding of the behaviour of homogeneity tests and their correct application to climatic time series preserve the climatic signal and eliminate or reduce the influence of non-climatic factors. The removal of false detected inhomogeneities and the acceptance of inhomogeneous series affect each subsequent analysis (e.g. trend assessments, extreme analysis). Therefore, it is of major importance that homogenization procedures are carefully applied. Moreover, as recommended by Aguilar *et al.* (2003), a complete description of the stations (names, code and exact location) should be provided.

Martínez *et al.* (2009) state that *SNHT detects breaks more easily at the beginning and the end of the temperature series, whereas the Buishand range and the Pettitt tests are more sensitive to breaks located in the middle of the series.*

In the following paragraphs, a brief description of SNHT behaviour is provided showing that SNHT performance decays for breaks located at the beginning and the end of series.

There are several studies that have investigated the strengths and weakness of break detection algorithms. For instance, Alexandersson and Moberg (1997a), since because the exact distribution of the test statistic under the null hypothesis is unknown, reported on critical levels of the SNHT statistic for series with a number of values from 10 to 250. Khaliq and Ouarda (2007) extended these critical values from 10 to 50 000. Furthermore, Alexandersson and Moberg (1997b) avoided the application of SNHT to segments with a length less than ten values. Ducré-Robitaille *et al.* (2003) analysed the behaviour of eight techniques for break detection with simulated

---

*Correspondence to: A. Toreti, Institute of Geography, Climatology and Meteorology, University of Bern, Bern, Switzerland.
E-mail: toreti@giub.unibe.ch

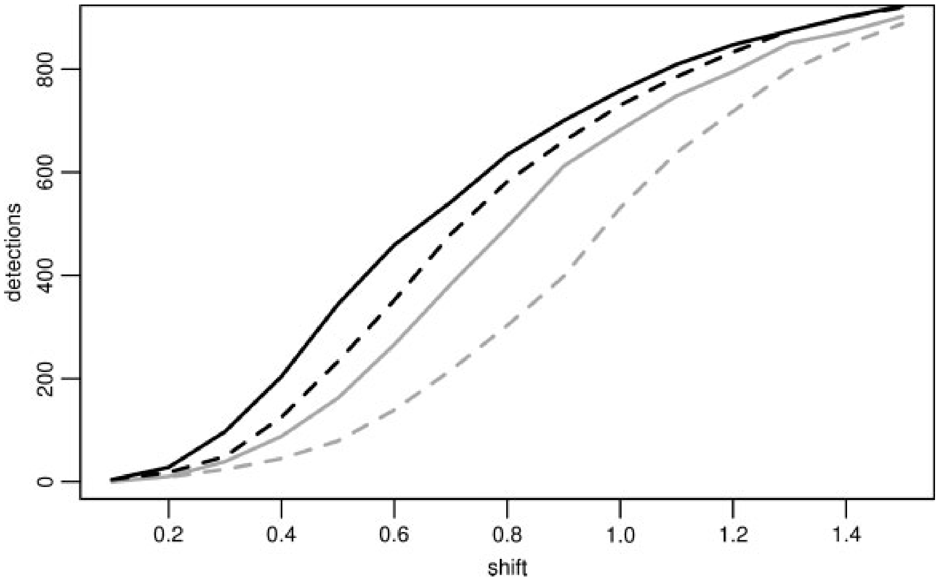

Figure 1. Results from the simulation study with a shift of the mean between 0.1σ and 1.5σ. The grey-dashed (solid) line shows the performance of SNHT with an inhomogeneity located at the fifth (tenth) value. The black-dashed line shows the number of detected break points in the 1000 tests with an inhomogeneity located at the 15th (50th) value.

annual series. They found that SNHT false break detection (a set of homogeneous series was tested) increases at the beginning and the end of the series even if the total number of false detection is quite low (8.6%). Adding artificial shifts (from 0.25 to 2 °C) to their set of simulated series they found poor performances of SNHT with breaks (characterized by a magnitude less than 0.75 °C) located at the beginning of the series. DeGaetano (2006) found, in his analysis of break detection methods, a similar behaviour of SNHT in the case of false detected breaks (using 1000 homogeneous series). He reported also a sharp decline in the number of identified discontinuities (a single shift was added to their simulated series) in series shorter than 21 years.

We carried out a simulation of 1000 series (with a length of 100 years, Figure 1) and found agreement with Ducré-Robitaille *et al.* (2003). An artificial shift was added after 5, 10, 15 and 50 values, with a magnitude between 0.1σ and 1.5σ (where sigma refers to the candidate series). Figure 1 shows the decreasing performance of SNHT in the detections of break points (with a tolerance of 1 year) close to the beginning of series. This is evident especially for breaks located at the fifth and tenth position with a magnitude less than 1.2σ. All the previous studies point out that the detection of inhomogeneities (break points) either located at the beginning (end) of a series or based on a small amount of values is difficult and should be confirmed by metadata.

Martínez *et al.* (2009) also provide a description of homogenization results but fail to explain how the SNHT was applied. We assume that their series were tested in an absolute way, that is, without reference series and without calculating difference series (which is a standard procedure in homogenizing data series; e.g. Peterson and Easterling, 1994; Aguilar *et al.*, 2003). This assumption is confirmed by their results, that is, most of the detected breaks are in 1984–1986 and 1992–1994, and by the fact that they did not consider data before 1975 due to the well-known slope change in temperature record *that could mask other lacks of homogeneity* (Section 2). However, this approach is not able to distinguish a climate shift from an inhomogeneity. Moreover, it leads to the acceptance of series that could be affected by real inhomogeneity. Indeed, some of the temperature series in their analysis were also used by Prohom *et al.* (2008); applying three different homogenization approaches, they showed that ten series out of 16 have showed at least one break point within the same period (1975–2007), all of them confirmed by the available metadata (i.e. relocation and/or change from manual stations to automatic weather stations).

SNHT (and other similar techniques) should be only applied to a standardized difference (in the case of temperature) series (Y-R, where Y denotes the so-called *candidate* and R the *reference series*) as recommended by Aguilar *et al.* (2003). This procedure can be carried out using either a single reference series (e.g. Peterson and Easterling, 1994; Alexandersson and Moberg, 1997a, 1997b) or multiple reference series using highly correlated series (e.g. Caussinus and Mestre, 2004). Even if these methods are based on reference series, breaks due to common network history (e.g. simultaneous introduction of new instruments) are not easily detectable. One way to get over this situation is to use series from different meteorological networks. This increases the probability to detect breaks affecting the entire network.

Martínez *et al.* (2009) attributed the identified breaks to two volcanic eruptions (El Chichón, Mexico, 1982; Mount Pinatubo, Philippines, 1991). As stated by Aguilar *et al.* (2003), homogenization aims to identify and eliminate (or reduce) the effects on non-climatic factors such as changes in instruments, observing practices, station relocations and station environments. Natural phenomena such as volcanic eruptions are not included in this list because they belong to the set of 'climatic factors', which can cause change points in a series, that should not be considered as inhomogeneities. Therefore, we point to the fact that a change point is the effect of one or more factors (climatic and non-climatic) on a series, whereas a break point or inhomogeneity is a change point caused by non-climatic factors.

In order to identify non-artificial change points affecting a series (e.g. piecewise trend; Seidel and Lanzante, 2004), techniques developed for this specific issue (e.g. Bai and Perron, 2003) must be applied after the homogenization process.

To summarize, the homogenization of climatic series is a difficult task that must be performed carefully, especially when metadata is not available. The aim of homogenization is the removal of non-climatic factors, whereas the climatic signal must be preserved. SNHT, as other techniques, must be applied comparing series with a *reference* to get reliable break points. Moreover, the limits of this test should be taken into account during the homogenization procedure. An incorrect application of homogenization procedures to climate data could subsequently lead to unreliable climate analysis. The comparison of detected break points by several tests is a good strategy, especially when metadata is not available. It gives robust results avoiding overestimation and correction of false inhomogeneities. Furthermore, methods for correcting daily temperature series (e.g. Della-Marta and Wanner, 2006) have been available for a couple of years and should be used instead of rejecting series affected by too many break points.

Unfortunately, Martínez *et al.* (2009) did not provide us data to calculate the influence of their approach on trend analysis.

References

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J. 2003. Guidance on metadata and homogenization. WMO TD N. 1186 (WCDMP N. 53); 51 pp.

Alexandersson H. 1986. A homogeneity test applied to precipitation data. *Journal of Climatology* **6**: 661–675.

Alexandersson H, Moberg A. 1997a. Homogenization of Swedish temperature data. Part I: homogeneity test for linear trends. *International Journal of Climatology* **17**: 25–34.

Alexandersson H, Moberg A. 1997b. Homogenization of Swedish temperature data. Part II: homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *International Journal of Climatology* **17**: 35–54.

Bai J, Perron P. 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* **18**: 1–22.

Buishand TA. 1982. Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology* **58**: 11–27.

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

DeGaetano AT. 2006. Attributes of several methods for detecting discontinuities in mean temperature series. *Journal of Climate* **19**: 838–853.

Della-Marta PM, Wanner H. 2006. A method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate* **19**: 4179–4197.

Ducré-Robitaille JF, Vincent LA, Boulet G. 2003. Comparison of techniques for detection of discontinuities in temperature series. *International Journal of Climatology* **23**: 1087–1101.

Khaliq MN, Ouarda TBMJ. 2007. On the critical values of the standard normal homogeneity test (SNHT). *International Journal of Climatology* **27**: 681–687.

Martínez MD, Serra C, Burgueño A, Lana X. 2009. Time trends of daily maximum and minimum temperatures in Catatonia (NE Spain) for the period 1975–2004. *International Journal of Climatology* **30**(2): 267–290, DOI 10.1002/joc.1884.

Peterson TC, Easterling DR. 1994. Creation of homogeneous composite climatological reference series. *International Journal of Climatology* **14**: 671–679.

Pettitt AN. 1979. A non-parametric approach to the change-point problem. *Journal of Applied Statistics* **28**: 126–135.

Prohom MJ, Esteban P, Herrero M, Mestre O, Aguilar E, Kuglitsch FG. 2008. Analysis of Catalan, Andorran and French temperature series from the early 20th century to the present using different homogenisation approaches. *Geophysical Research Abstracts* **10**: EGU General Assembly 2008, Vienna.

Seidel DJ, Lanzante JR. 2004. An assessment of three alternatives to linear trends for characterizing global atmospheric temperature changes. *Journal of Geophysical Research* **109**: D14108, DOI 10.1029/2003JD00414.

Von Neumann J. 1941. Distribution of the ratio of the mean square successive difference to the variance. *Annals of Mathematical Statistics* **12**: 367–395.