Benchmarking Homogenization Algorithms for Monthly Data

This is the author's manuscript

Original Citation:

Availability:
This version is available http://hdl.handle.net/2318/138704 since

Publisher:
AIP Publishing LLC

Published version:
DOI:10.1063/1.4819690

Terms of use:
Open Access
Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)
Benchmarking Homogenization Algorithms for Monthly Data


1 Meteorological institute of the University of Bonn, Germany
2 Meteo France, Ecole Nationale de la Meteorologie, Toulouse, France
3 Center on Climate Change (C3), Universitat Rovira i Virgili, Tarragona, Spain
4 Zentralanstalt für Meteorologie und Geodynamik, Wien, Austria
5 Agencia Estatal de Meteorologia, Palma de Mallorca, Spain
6 Slovenian Environment Agency, Ljubljana, Slovenia
7 Hungarian Meteorological Service, Budapest, Hungary
8 Czech Hydrometeorological Institute, Brno, Czech Republic
9 Global Change Research Centre AS CR, v.v.i., Brno, Czech Republic
10 Deutscher Wetterdienst, Offenbach, Germany
11 NOAA/National Climatic Data Center, USA
12 Meteorological and hydrological service, Zagreb, Croatia
13 Laboratory of Atmospheric Physics, University of Patras, Greece
14 National Institute of Meteorology and Hydrology -- BAS, Sofia, Bulgaria
15 Norwegian Meteorological Institute, Oslo, Norway
16 Department of Earth Science, University of Turin, Italy
17 National Meteorological Administration, Bucharest, Romania
18 National Institute for R&D in Environmental Protection, Bucharest, Romania
19 Institute of Atmospheric Sciences and Climate (ISAC-CNR), Bologna, Italy
20 Grup de Climatologia, Universitat de Barcelona, Spain
21 Meteorological Service of Catalonia, Area of Climatology, Barcelona, Catalonia, Spain
22 Centre d’Estudis de la Neu i de la Muntanya d’Andorra (CENMA-IEA) Andorra
23 Royal Netherlands Meteorological Institute, De Bilt, The Netherlands
24 Met Office Hadley Centre, Exeter, United Kingdom

Abstract. The COST (European Cooperation in Science and Technology) Action ES0601: Advances in homogenization methods of climate series: an integrated approach (HOME) has executed a blind intercomparison and validation study for monthly homogenization algorithms. Time series of monthly temperature and precipitation were evaluated because of their importance for climate studies. The algorithms were validated against a realistic benchmark dataset. Participants provided 25 separate homogenized contributions as part of the blind study as well as 22 additional solutions submitted after the details of the imposed inhomogeneities were revealed. These homogenized datasets were assessed by a number of performance metrics including i) the centered root mean square error relative to the true homogeneous values at various averaging scales, ii) the error in linear trend estimates and iii) traditional contingency skill scores. The metrics were computed both using the individual station series as well as the network average regional series. The performance of the contributions depends significantly on the error metric considered. Although relative homogenization algorithms typically improve the homogeneity of temperature data, only the best ones improve precipitation data. Moreover, state-of-the-art relative homogenization algorithms developed to work with an inhomogeneous reference are shown to perform best. The study showed that currently automatic algorithms can perform as well as manual ones.

Keywords: Surface climate network, instrumental climate records; monthly temperature records; monthly precipitation records; surface stations; homogenization, benchmarking; blind validation study; surrogate data.
INTRODUCTION

To study climate variability the original surface observations are indispensable, but these have to be treated with care. Long observational records always contain changes due to non-climatic factors as well. Such inhomogeneities can be either sudden jumps (breaks) or gradual trends in one station. In a recent paper [1], the methods to remove these factors were blindly tested on a benchmark temperature and precipitation dataset with inserted inhomogeneities. This abstract will present the main ideas from this study. All the most common and the most developed algorithms were tested.

Most surface stations are not operated for climatic purposes, but rather to meet the needs of weather forecasting, agriculture and hydrology [2]. Consequently, the average period between detected breaks in Western instrumental records is 15 to 20 years. The typical size of the breaks is of the same order as the climate change signal during the 20th century [3-7]. Specific breaks or gradual trends in one station. In a recent paper [1], the methods to remove these factors were blindly tested on a benchmark temperature and precipitation dataset with inserted inhomogeneities. This abstract will present the main ideas from this study. All the most common and the most developed algorithms were tested.

Inhomogeneities caused by changes in the screens that protect the instruments for radiation and wetting are also common [10]. In 19th century Europe it was common to use a metal screen in front of a window on a North facing wall. However, the building may warm the screen leading to higher temperature measurements. When this problem was realized the cotton region shelter was introduced. Other typical causes of inhomogeneities are changes in the surrounding environment, e.g., land use change and building activity. In recent times, the most important inhomogeneity is the change over to automatic weather stations [11].

HOMOGENIZATION

 Ideally, the date of a change of instruments, locations or observing practices would be documented and parallel measurements made with the original and the new set-up for several years [12]. By making parallel measurement with replicas of historical instruments, screens, etc., the influence of some historical inhomogeneities can still be studied today. Because you are never sure that your metadata is complete, statistical homogenization is necessary as well. The most commonly used principle to detect and remove the artificial changes is relative homogenization [13]. This assumes that nearby stations are exposed to almost the same climate signal, but not any non-climatic changes. By looking at the difference between nearby stations, the year to year variability of the climate is removed, as well as the regional climatic trend. In such a difference time series, a clear and persistent jump can easily be detected and can only be due to changes in the measurement conditions.

If there is a jump (break) in a difference time series of a pair of stations, it is not yet clear which of the two stations it belongs to. Furthermore, time series typically have more than just one jump. These two features make statistical homogenization a challenging and beautiful statistical problem. Homogenization algorithms typically differ in how they solve these two fundamental problems.

Traditionally, this first fundamental problem is solved in relative homogenization by comparing a candidate series with a composite reference time series computed from its neighboring stations. This composite reference is assumed to be homogeneous due to averaging, which is only approximately true. The main research impetus for the last two decades has been the development of so-called direct homogenization algorithms that also function with an inhomogeneous reference time series.

Sometimes there are no other stations in the same climate region. In this case, sometimes absolute homogenization is applied and the inhomogeneities are detected in the time series of one station, i.e. without using a reference [14]. If there is a clear and large break at a certain date, such a break may be removed reasonably accurately, but smaller jumps and gradually occurring inhomogeneities (for instance due to the urban heat island or growing vegetation) cannot be distinguished from real natural variability and climate change. Data homogenized this way does not have the quality one may expect and should be used with much care.
BENCHMARKING

The benchmark dataset mimics station networks and their data problems with unprecedented realism. Homogeneous data was generated using the so-called surrogate data approach, which reproduces the cross- and auto-correlation functions, as well as the non-Gaussian distribution of climate observations [15, 16]. To this data random break-type inhomogeneities, as well as breaks that occur simultaneously in multiple stations, were added. Furthermore, local trends, which either continue at the end (to model for instance the urban heat island effect) or go back to baseline (to model growing vegetation that is cut back at the end) were inserted. The sizes of the breaks and local trends follow a normal distribution with a width of 0.8 °C for temperature [17], and 15 % for precipitation. Further, a stochastic nonlinear network-wide trend was added. The main novelty was that the test was blind. Furthermore, the benchmark was generated and the analysis of results was performed by independent researchers, who did not homogenize the data themselves. Everyone was invited to homogenize the data; 25 homogenized blind contributions were returned.

RESULTS

For a clear presentation of all results, the homogenization methods used to homogenize the 25 contributions would have to be explained, this is not possible in the limited space of this abstract.

To get a flavor of the results, see Figure 1. It shows the root mean square error of the centered monthly data. The time series are centered by subtracting the mean because homogenization aims to improve the temporal consistency of the data, not the absolute level. The first four contributions in Figure 1 — ACMANT (Spain) [18], PRODIGE monthly (Meteo France) [19], USHCN main (NOAA, USA) [4] and MASH main (Hungarian weather service) [20, 21] — are all direct homogenization algorithms and clearly perform better than the traditionally used SNHT method [22], here exemplified by CSNHT.

The USHCN contribution is unique in that it has almost no stations with a higher error after homogenization, the contribution also has many values exactly on the bisect (no changes performed) and it made only small changes to the network without any inserted breaks (values on the ordinate). It should be noted that
all but one relative method did improve the station trends. Many inhomogeneities are documented. Had such documentation been provided for part of the breaks in the benchmark data, the errors would have been lower.

The second main conclusion is that direct relative homogenization algorithms are clearly better than traditional ones. It needed a realistic benchmark dataset to see this difference with such clarity. With mathematical argumentation, climatological reasoning and the benchmark metrics all pointing in the same direction, we thus strongly recommend the use of direct homogenization algorithms.

The performance ranking of the homogenization methods depends on the error metric considered, on whether the root mean square error is computed on the monthly, yearly or decadal data and on whether it is computed on the station data or on the network mean climate signal. These rankings also do not correlate strongly with the error in the linear trend estimates (or break detection scores). In other words, it is difficult to compute one error metric that would signify the remaining error after homogenization for all climatic purposes. The computation and communication of the remaining uncertainties of homogenized data should be one of the research priorities for the coming years.

We feel that benchmarking has helped the homogenization community to mature [24]. The discussions on the properties of the benchmark, the nature of inhomogeneities in the various regions and on homogenization methods, as well as the joint work on the same dataset helped to bring scientists closer together in a way that writing individual papers cannot. The International Surface Temperature Initiative has started a follow-up benchmarking program for homogenization algorithms [25]. This benchmark will be global and be even more realistic, especially due to the inclusion of metadata, biased inhomogeneities and random missing data.

Everyone is invited to download and analyze the benchmark dataset. The homogeneous, inhomogeneous and homogenized datasets are published in the internet. Another offspring of the Action is HOMER, an open-source state-of-the-art homogenization package based on the best methods. The package is written in R and also performs basic quality control. Furthermore a mailing list for researchers working on homogenization has been started. All these resources can be accessed via the webpage of the Action, http://www.homogenisation.org, which will be kept running for the coming years and which also contains an extensive bibliography.

With advanced and well-validated statistical methods, the homogenization of annual and monthly station data is a mature field. The homogenization of daily data is still in its infancy, however. Daily data are essen-
tial for studying extremes of weather and climate and therefore the basis for important political decisions with huge socio-economic consequences. For such studies the complete distribution needs to be homogenized. Looking at the physical causes of inhomogeneities, one would expect that many of them especially affect the tails of the distribution of the daily data. Likewise the IPCC AR4 report warns that changes in the mean of the distribution of the daily data. A better understanding of the nature of daily inhomogeneities and better tools to correct them will be the main challenge for the coming years.

REFERENCES


