AIP Conference Proceedings

Homogenization of a surface solar radiation dataset over Italy

Cite as: AIP Conference Proceedings **1810**, 090004 (2017); https://doi.org/10.1063/1.4975544
Published Online: 22 February 2017

Veronica Manara, Michele Brunetti, Maurizio Maugeri, Arturo Sanchez-Lorenzo, and Martin Wild

View Online

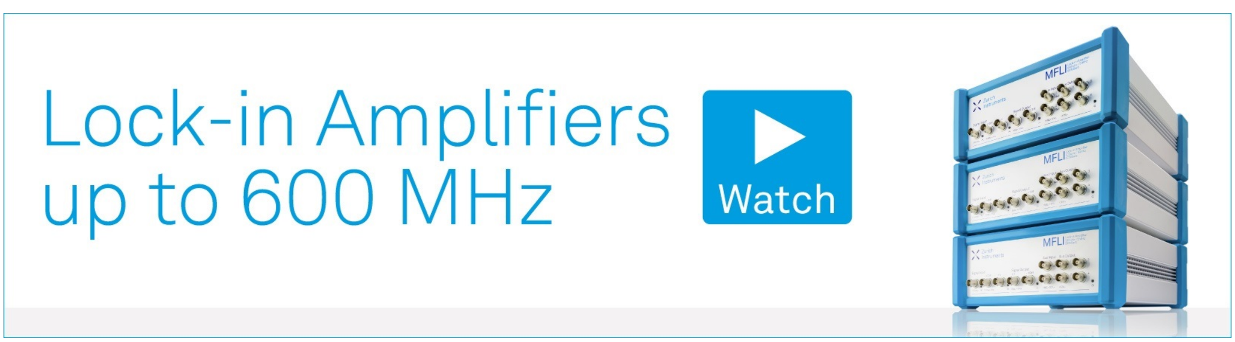
Export Citation

ARTICLES YOU MAY BE INTERESTED IN

The Global Energy Balance Archive (GEBA): A database for the worldwide measured surface energy fluxes
AIP Conference Proceedings **1810**, 090013 (2017); https://doi.org/10.1063/1.4975553

Cloud cover estimation optical package: New facility, algorithms and techniques
AIP Conference Proceedings **1810**, 080009 (2017); https://doi.org/10.1063/1.4975540

Cloud effects on atmospheric solar absorption in light of most recent surface and satellite measurements
AIP Conference Proceedings **1810**, 090003 (2017); https://doi.org/10.1063/1.4975543



AIP Publishing

© 2017 Author(s).

Homogenization of a Surface Solar Radiation Dataset over Italy

Veronica Manara[1, a)], Michele Brunetti[2, b)], Maurizio Maugeri[1, 2, c)], Arturo Sanchez-Lorenzo[3, d)] and Martin Wild[4, e)]

[1]Department of Physics, Università degli Studi di Milano, Milan, Italy.
[2]Institute of Atmospheric Sciences and Climate, CNR, Bologna, Italy.
[3]Instituto Pirenaico de Ecología, Consejo Superior de Investigaciones Científicas (IPE-CSIC), Zaragoza, Spain.
[4]ETH Zürich, Institute for Atmospheric and Climate Science, Zürich, Switzerland.

[a)]Corresponding author: veronica.manara@unimi.it
[b)]m.brunetti@isac.cnr.it
[c)]maurizio.maugeri@unimi.it
[d)]arturo.sanchez@ipe.csic.es
[e)]martin.wild@env.ethz.ch

Abstract. Observational data cannot be used for climate research without a clear knowledge about the state of the data in terms of temporal homogeneity. The main steps and results of the homogenization procedure applied to a surface solar radiation dataset over the Italian territory for the period 1959-2013 are discussed.

INTRODUCTION

In the last two decades, the scientific community has become aware of the fact that the real climate signal in the original series of observational data can be hidden behind non-climatic noise caused by station relocations, changes in instruments, observation times, observers and so on [1]. This issue is very important for all meteorological variables [1] and is even more important for variables such as surface solar radiation (SSR), whose measurements are subjected to a number of problems [2]. Therefore, the study of decadal variability and long-term trend of SSR has a strong need of facing the temporal homogeneity issue.

Manara et al. in a recent paper discussed the main points required to set up a SSR database over Italy and to obtain regional average SSR records for Northern and Southern Italy in the 1959-2013 period [3]. That paper, however, mainly focuses on the temporal evolution of the considered records, whereas data pre-processing is presented in a rather concise way. Within this context, the aim of the present paper is to discuss in more detail the homogenization of this dataset.

Meteorological series can be tested for homogeneity and homogenized both by direct and indirect methods. The first approach is based on objective information that can be extracted from the station history (metadata), the latter uses statistical methods, generally based on comparison with other series. The direct methodology has the advantage of providing detailed information about the time location of the inhomogeneities and the sources that caused them, but they are not always available and complete. Moreover, it is generally difficult to convert them into quantitative values useful to correct the discontinuities. On the other hand, the indirect methodology is more suitable to calculate correcting factors to eliminate the breaks. However, the identification of the inhomogeneities is not always easy and unambiguous, as (i) inhomogeneities and errors are present in almost all observational series, making it difficult to objectively assign the breaks to one or another of them and (ii) correlation among data series depends on various factors (e.g., regional patterns, climate elements under analysis, time resolution of data and so on) and when the

Radiation Processes in the Atmosphere and Ocean (IRS2016)
AIP Conf. Proc. 1810, 090004-1–090004-4; doi: 10.1063/1.4975544
Published by AIP Publishing. 978-0-7354-1478-5/$30.00

common variance between the test series and the reference series is too low, the potential discontinuity signal in a homogeneity test disappears into statistical noise [4].

The approach we use is based on the indirect methodology. We give however great importance to the metadata as they are very important to help in deciding whether a possible break indicated by the statistical test has actually to be corrected [4]. In particular, our procedure consists in subjecting the records to a relative homogeneity test [4, 5], where each series (test series) is tested against 10 other series (reference series) by means of the Craddock test [6]. The reference series are chosen in the same geographical region as the test series in order to avoid extrapolating the signal of a region to another one. When a break (and so an inhomogeneity) is identified in the test series, those reference series that prove to be homogeneous in a sufficiently long period centered on the break are chosen to estimate the adjustment factors to apply to the test series. Several series are used in order to better identify the break and to get adjustments that are more reliable. For each break, the monthly adjustments are calculated from each reference series. Then, the correcting factors are fitted with a trigonometric function in order to smooth the noise in the adjustment annual cycle and extract only the physical signal. The final set of the monthly adjustments is then calculated by averaging all the yearly cycles. When a clear yearly cycle is not evident, the adjustments used to correct the monthly data are chosen constant through the year and they are calculated as the weighted average among the monthly values, where the weights are the ratios between the monthly SSR mean and the total SSR over the year. When a break is identified, the preceding or the following portion of the series is corrected, leaving when possible the most recent portion of the series unchanged in order to update the records with data yet to be obtained. The entire procedure is generally applied to monthly records. It can however be easily extended also to daily ones under the condition that each monthly homogenized value obtained averaging the daily homogenized values is equal to the corresponding monthly homogenized value obtained directly from the homogenization of the monthly series.

ITALIAN SSR DATA AND METADATA

The SSR station network considered in this paper is mainly based on the Italian Air Force synoptic stations (AM - Aeronautica Militare - 29 records) and on the National Agrometeorological database (BDAN - Banca Dati Agrometeorologica Nazionale - 19 records). Moreover, it includes the observatory of Trieste and five Swiss stations from the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss) close to the Italian border. The data set encompasses 54 records with daily resolution covering the 1959-2013 period. The data that come from AM and from Trieste were recorded with the Robitzsch bimetallic actinograph until the 1980s. This instrument was then replaced with the CM11 Kipp & Zonen pyranometer in the first case and with different types of CM Kipp and Zonen pyranometers in the second one. The CM11 Kipp and Zonen pyranometer was also used in the stations included in the BDAN database with the only exception of two stations that use the EP07 Middleton Instruments pyranometer. Finally, the data that come from MeteoSwiss were measured with a CM21 Kipp & Zonen pyranometer. Full details on data and metadata availability are reported in [3] and [7].

HOMOGENIZATION OF THE ITALIAN SSR RECORDS

The homogenization of the Italian SSR records was not easy because of three main problems: the low spatial density of the stations, the large fraction of missing data [7] and the presence of a high number of inhomogeneities. The first two problems are enhanced by a fairly low spatial coherence of the records, which causes the common variance between two stations to fall to about 50% at a distance of 150 km (Fig. 1). This makes the selection of the reference series more difficult, considering that they should not only be located close to the test series but must also have available data over the period to homogenize.

Hereinafter we show, as an example, the different steps of the homogenization procedure for the Messina station, located in the Southern part of Italy (38.201°N, 15.553°E, 59 m) and with available data for the period 1959-2006.

Figure 2(a), shows the Craddock test applied to Messina with respect to 10 reference stations. Each curve shows the cumulative normalized differences (s) between the test and the reference series according to the formula:

$$s_i = s_{i-1} + a_i \left({}^{b_m}\!/_{a_m} \right) - b_i \tag{1}$$

where a is the reference series, b is test series while a_m and b_m are the mean series over the period selected as reference period to calculate the adjustment factors.

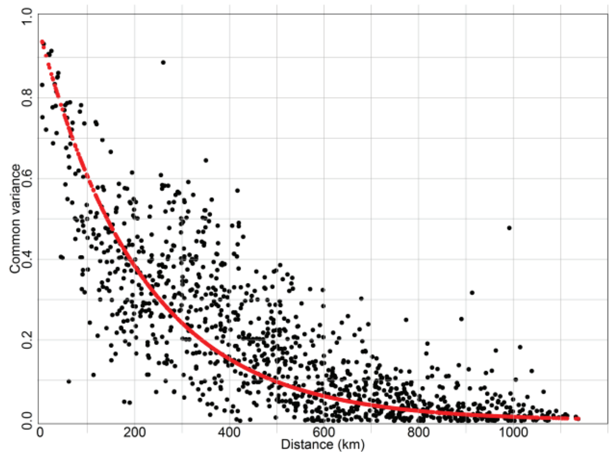
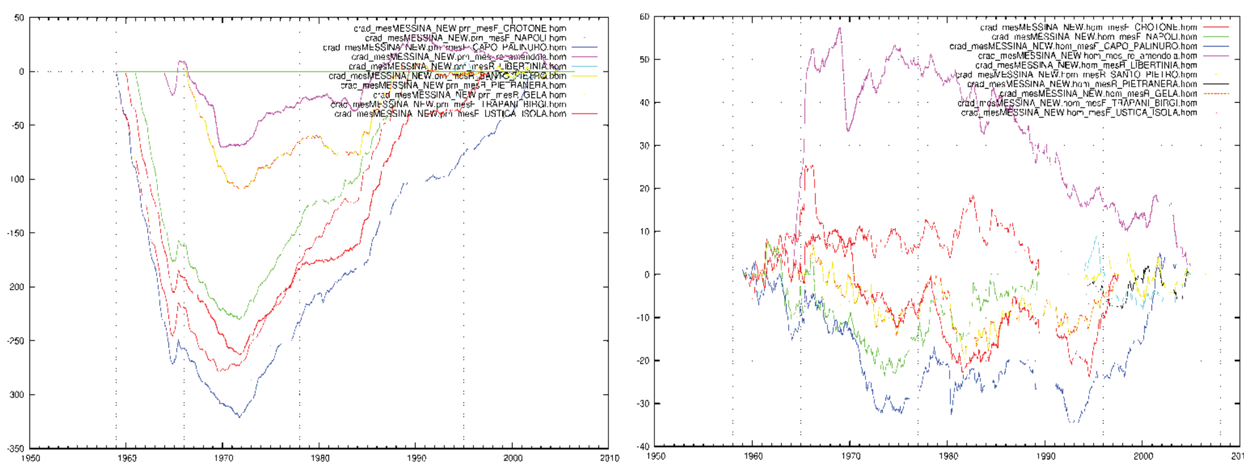


FIGURE 1. Common variance versus distance of monthly SSR records. In order to remove the effect of the annual cycle, the common variance has been calculated after deseasonalization of the records.

This test identifies the breaks by means of changes in the slope of the test curves [6]. The most important breaks seem therefore to be located in the years 1964, 1965, 1969, 1971, 1973, 1979 and 1989. Most of these breaks correspond to changes highlighted by the metadata. Specifically, the breaks of the years 1964 and 1971 correspond to recalibrations of the instrument while the break of the year 1989 corresponds to a change in the instrument (from the Robitzsch bimetallic actinography to the CM11 Kipp & Zonen pyranometer). The first break we corrected is the most recent one (1989). We adjusted therefore all data between the selected break and the preceding one by means of monthly correcting factors (0.91 for the period 1984-1989). They allow to make the data of the selected period homogeneous with those after the break used as reference period to calculate the adjustment factors. The resulting record was obviously not homogeneous as there were other breaks before the one we considered in the first step. We therefore applied the same procedure to all the identified breaks until we got the homogenized record.

Figure 2(b) shows the Craddock test applied to this final homogeneous record. The Craddock curves of Messina with respect to the different reference series show changes that are amenable more to statistical noise than to non-homogeneity in the test series. The change that is evident in the pink line around the mid of the 1960s is probably due to a break in the reference series and not in the test series considering that this curve is the only one that shows a pronounced variation during these years. The same procedure we applied to the Messina record was applied to all 54 stations of our data set.



(a) (b)

FIGURE 2. Craddock test applied to Messina with respect to 10 reference series: for the original series (a) and for the final homogeneous series (b). The scale of the y-axis is different to better highlight the signal contained in the series.

Some statistics on the breaks we corrected are shown in Fig. 3, which shows the number of corrected breaks vs time in absolute and relative terms (with respect to the number of available series). This figure shows that the Italian SSR records had a rather high number of breaks even if some of them concern only few years or small adjustments. The relative number of breaks is high especially until the 1980s due to the high number of recalibration and instrument changes that occurred in this period.

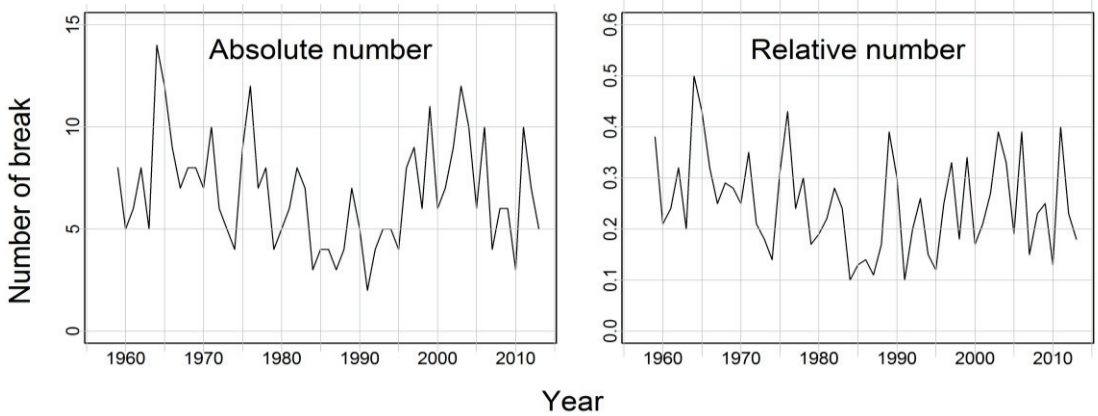


FIGURE 3. Number of detected breaks per year in the SSR data set (left column) and number of detected breaks normalized by the number of available series (right column).

A very interesting issue in relation to data homogenization is to check whether the adjustment factors applied to the station series give raise to a random noise that disappears considering the average over a large number of stations (regional average record) or, on the contrary, they produce a long-term signal that masks the real climate tendency. In the first case the ratio between the regional average records calculated from the original and from the homogenized data set would be around one for the entire record length. For the Italian SSR record this situation concerns only the period after the 1980s where at National scale this ratio is 1.008, whereas before 1980 the effect of the inhomogeneities does not produce a random noise (see Fig. 3 in [3]).

CONCLUSIONS

The real climate signal in the original series of observational data can be hidden behind non-climatic noise. So, at present, the statement that time series of meteorological data cannot be used for climate research without a clear knowledge about the state of the data in terms of temporal homogeneity has very large consent, especially when variables such as surface solar radiation (SSR) are considered. In this paper, the main steps of the homogenization procedure applied to an Italian SSR dataset for the period 1959-2013 are described. The procedure adopted is based on an indirect methodology (Craddock test) supported by metadata. The result give evidence that homogenization is a key issue for the study of the temporal evolution of SSR.

REFERENCES

1. E. Aguilar et al., “Guidelines on climate metadata and homogenization”, World Climate Programme Data and Monitoring, WCDMP-No.53, WMO-TD No.1186, Geneva, (2003).
2. WMO-No.8, “Measurements of radiation”, Guide to meteorological instruments and observing practices, Chap. 7, World Meteorological Organization, Geneva, Switzerland, (2008)
3. V. Manara et al., *Atmos. Chem. Phys. Discussion*, doi:10.5194/acp-2016-206, in review, (2016).
4. M. Brunetti et al., *J. Climatol*, **26**, 345-381, (2006).
5. V. K. C. Venema et al., *Clim. Past*, **8**, 89-115, doi:10.5194/cp-8-89-2012, (2012).
6. J. M. Craddock, *Weather*, **34**, 332-346, (1979).
7. V. Manara et al., “Reconstructing sunshine duration and solar radiation long-term evolution for Italy: a challenge for quality control and homogenization procedures”. Proceedings of the 14th IMEKO TC10 Workshop Technical Diagnostics - New Perspectives in Measurements, Tools and Techniques for system’s reliability, maintainability and safety. Milan, Italy, (June 27-28, 2016).