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

Reconstructing sunshine duration and solar radiation long-term evolution for Italy: a challenge for quality control and homogenization procedures

Veronica Manara[1], Michele Brunetti[2], Maurizio Maugeri[1,2]

[1]Department of Physics, Università degli Studi di Milano, Milan, Italy, veronica.manara@unimi.it
[2]Institute of Atmospheric Sciences and Climate, CNR, Bologna, Italy

Abstract – **In the last two decades, the scientific community has become aware of the fact that the real climate signal in original series of meteorological data is generally hidden behind non-climatic noise caused by a number of factors. Time series of meteorological data can therefore not be used for climate research without facing this issue. In this context, we have recently set up a database of Italian sunshine duration and solar radiation daily records and we have subjected them to a detailed quality check and homogenization procedure. Moreover, as the records are rather sparse and a significant fraction of them have wide gaps, we completed the data and set up a procedure in order to obtain at first a gridded version of the dataset and then average records that are representative of the entire Italian territory. The paper will highlight the main steps of the methodology that allowed us to get these average quality-checked and homogenized records and will discuss some open issues.**

I. 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 meteorological data is generally hidden behind non-climatic noise. The non-climatic signals can be caused by station relocations, changes in instruments and instrument screens, changes in observation times, observers, and observing regulations, algorithms for the calculation of means and so on [1], [2]. 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 homogeneity has very large consent [1], [2].

The need of facing the homogeneity issue is important for all meteorological variables. However, until now, most of the efforts have been dedicated to temperature and precipitation homogenization and few works about other variables are available in the scientific literature. Therefore, it is imperative to improve our knowledge about the effect of inhomogeneities in masking the climate change signal in other variables. Among them it is of fundamental importance to improve the homogeneity of variables such as sunshine duration (SD) and surface solar radiation (SSR), which are very important to evaluate the temporal change and variability in the energy balance of the climate system and whose measure is subjected to a number of problems [3], [4], [5] which can hidden the real signal. An additional issue concerning the Italian records is that, even though a rather high number of stations measures SD and SSR since the end of the 1950s, a significant fraction of them covers only a minor time interval. Therefore, beside to a careful homogenization, the study of the temporal evolution of these meteorological variables over Italy requires also the application of methodologies that allow managing the presence of records, which have often more gaps than data. This problem is particularly evident for SD that is here presented for the 1936-2013 period.

Within this context, the goals of this paper are: (a) to present the SD and SSR records available for Italy (Section II), (b) to discuss how we managed the homogeneity issue (Section III), (c) to present the methodology we setup to get monthly, seasonal and annual SD (SSR) records for the 1936-2013 (1959-2013) period which are representative for the entire Italian territory (Section IV) and (d) to present and discuss the SD long-term trend and evolution as an example of the results obtained applying these methodologies to the original series (Section V).

II. DATA

The SD station network considered in this paper is presented in [6]. It is based on records from (a) the paper archive of the former Italian Central Office for Meteorology (24 records -1936-2013), which is managed by CREA (Council for Agricultural Research and Agricultural Economy Analysis), (b) the database of Italian Air Force synoptic stations (AM - Aeronautica Militare) (47 records – 1958-2013), and (c) the National Agrometeorological database (BDAN - Banca Dati Agrometeorologica Nazionale) (59 records – 1994-2013,

some of them allow updating corresponding AM records), which is also managed by CREA. In addition, it includes four records (Pontremoli, Varese, Modena, and Trieste) from university and local observatories. The dataset encompasses 104 daily records (Fig. 1) with data for the 1936-2013 period.
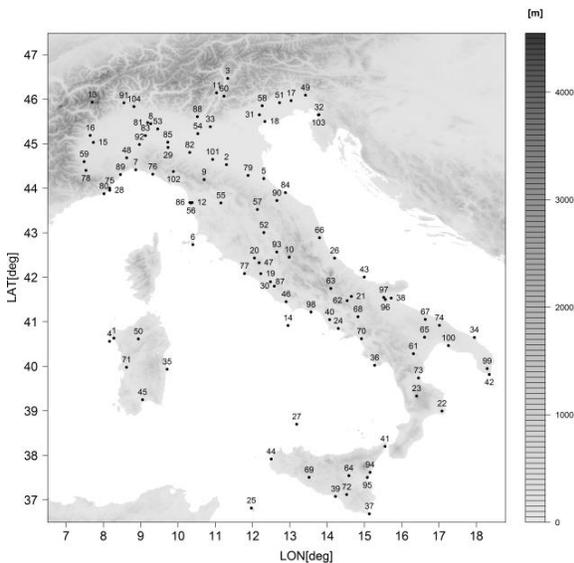

Fig. 1. Spatial distribution of the SD stations. The stations have the same order as in Table 1 reported by [6]. The figure shows the orography of the region too.
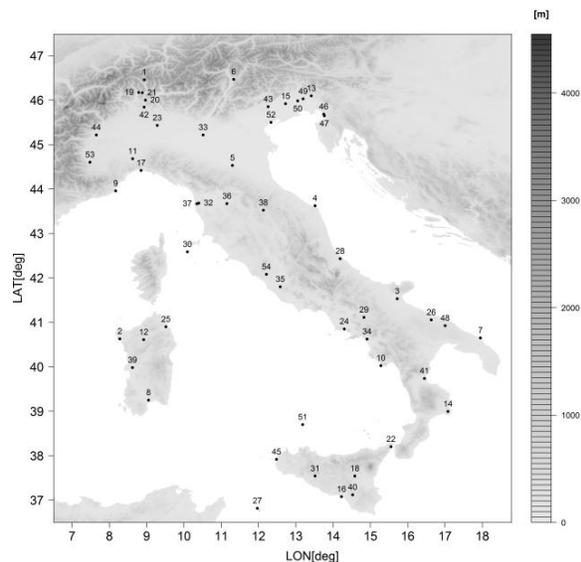

Fig. 2. Spatial distribution of the SSR stations. The stations have the same order as in Table 1 reported by [7]. The figure shows the orography of the region too.

The SSR station network considered in this paper is presented in [7]. It is mainly based on AM (29 records –

1959-2013 period) and BDAN stations (19 records – 1994-2013 period). Moreover, it includes the observatory of Trieste (1971-2013 period) and five Swiss stations close to the Italian border, from the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss – 1981-2013 period). The data set encompasses 54 daily records (Fig. 2) with data for the 1959-2013 period.

We calculated the monthly series for SD and for SSR when the fraction of missing data did not exceed 10% and 20% respectively.

III. QUALITY CHECK AND HOMOGENIZATION

At first, the metadata of the records were recovered and the coordinates were checked for consistency, controlling elevation in relation to position. Metadata allow to build the station history that is useful to anticipate and preview which problems could affect the records.

All daily records were checked in order to find out and correct gross errors (negative values and values exceeding the SD/SSR at the top of the atmosphere) [1].

We subjected then all our monthly records to a relative homogeneity test [2], [8]. In this procedure, each series (test series) is tested against 10 other series (reference series) by means of the Craddock test [9]. The reference series are chosen in the same geographical region of 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, the 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 for 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 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 monthly mean and the total over the year. When a break is identified, the preceding portion of the series is corrected, leaving the most recent portion of the series unchanged in order to update the records with data yet to be obtained.

The ability of the procedure to detect breaks depends on the spatial coherence of the data and on the density of the record network. Some information on such issues is given by [6] and [7] for SD and SSR respectively.

The importance of the homogenization procedure to get realiable single station series is evident from Fig. 3 where the Italian average annual SD anomaly series (relative anomalies with respect to the period 1984-2013) before

and after the homogenization is shown with corresponding Gaussian low-pass-filters (11-year window; 3-year standard deviation). These filters allow a better visualization of the decadal variability and long-term trend.
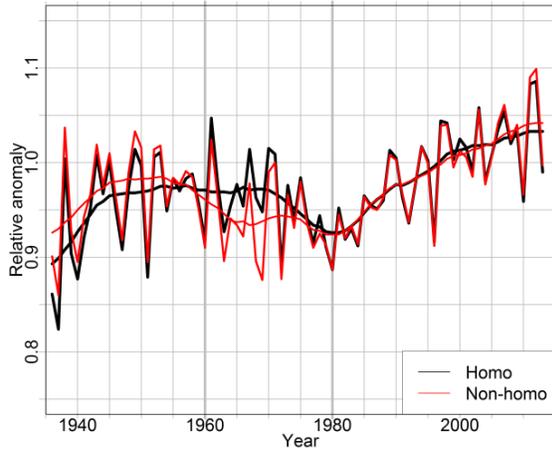

Fig. 3. Average annual Italian SD anomaly series plotted together with an 11 year window-3-year standard deviation Gaussian low-pass filter before (red line) and after (black line) the homogenization procedure. The anomalies are expressed with respect to 1984-2013 period.

The average annual records before and after the homogenization procedure show a different decadal variability during the periods 1936-1959 (average adjustment: 0.981) and 1960-1979 (average adjustment: 1.033) [6]. On the contrary, during the period 1980-2013, only small correcting factors have been applied to the original series (average adjustment: 1.004) [6]. The necessity of reducing SD before 1960 to get homogeneous series may be due to the strong urbanization which occurred in Italy in the following decades, causing a reduction in the sky-view factor especially for some urban observatories [6].

A similar result is obtained also for SSR especially before 1980s when many instrument changes and recalibrations occurred. More details are given by [7].

IV. FROM THE STATION RECORDS TO THE AVERAGE REGIONAL RECORDS

Together with the homogeneity issue, the main problem of the SD and SSR datasets is that most of the records cover only a short time interval with respect to the period covered by the corresponding data set. This problem is clearly highlighted by dividing the total number of available monthly data (32443 for SD and 18059 for SSR) for the number of months in the period covered by the two datasets. The result gives evidence that the number of available SD data are only 33.4 % of the ones that would be necessary to cover without gaps the entire

1936-2013 period. The situation is better for SSR. However, also for this variable we have only 51.6 % of the data that would be necessary to cover without gaps the entire 1959-2013 period.

Another problem is that there are periods in the time intervals covered by the two datasets with particularly low data availability, as for example the period before the end of 1950s for SD and the period at the beginning of 1990s for SSR.

In order to better highlight all these problems, we show in Fig. 4 and 5 the data availability for the two datasets.
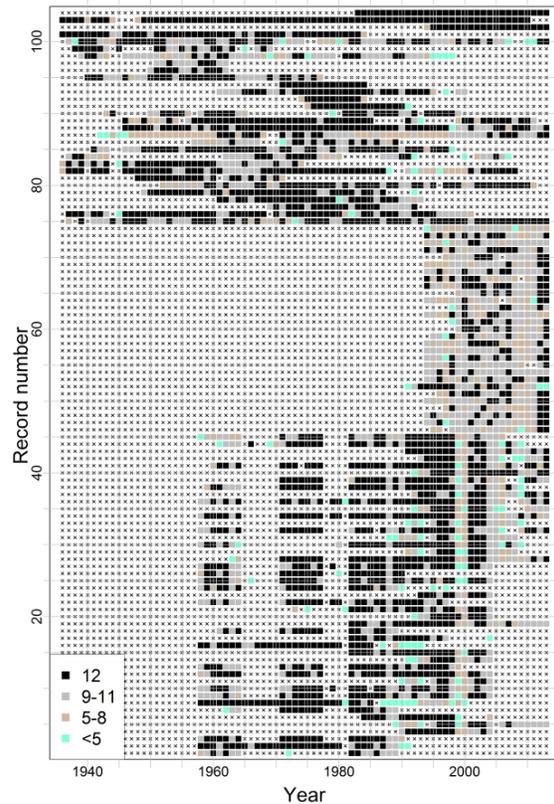

Fig. 4. Number of available SD months per year for each record. The record number on the y-axis correspond to the number reported in Fig. 1. A year is indicated with a black square if the available monthly values are 12, with a grey square if they are comprised between 9 and 11, with a brown square if they are comprised between 5 and 8 and with a blue square if they are less than 5. A year is indicated with a small cross when no months are available.

Here the rows correspond to the record numbers as indicated in Fig. 1 and Fig. 2, whereas the columns correspond to the years. A year is indicated with a black square if the available monthly values are 12, with a grey square if they are comprised between 9 and 11, with a brown square if they are comprised between 5 and 8 and with a blue square if they are less than 5. A year is

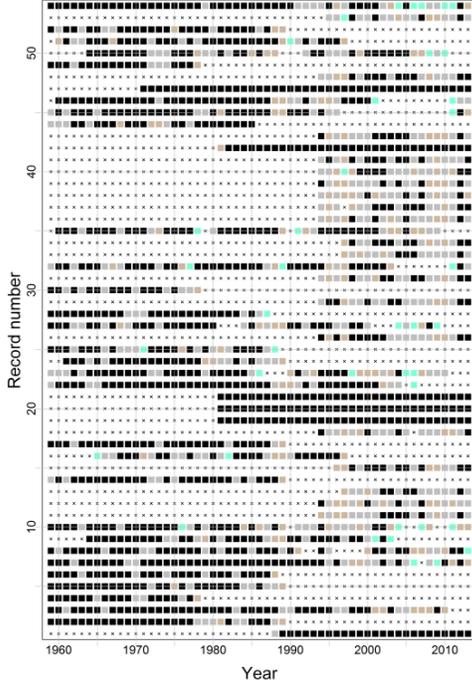indicated with a small cross when no months are available.

Fig. 5. Number of available SSR months per year for each record. The record number on the y-axis correspond to the number in Fig. 2. The data are represented as in Fig. 4.

The many gaps shown in Fig. 4 and Fig. 5 make a direct estimation of average Italy SD and SSR records impossible. Such records would in fact be highly vulnerable to fluctuations in spatial coverage, with positive bias when more sunny stations are available and negative bias in the opposite case. For this reason, we transformed the records into anomaly records before averaging them over Italy. The advantage of the anomalies with respect to the absolute values is that their variations take place on a much larger spatial scale. Such a greater spatial coherence causes the anomalies to be less vulnerable to missing data and more suitable for the calculation of regional average or gridded series.

In order to transform the absolute value records into anomaly ones, we had to fill the gaps in the monthly series, at least in the period we used as reference to convert the data into anomalies.

Specifically in [6], each missing SD datum was estimated by means of the most correlated reference record. The selection of the reference record to use for the estimation of the missing datum was performed considering only the records fulfilling three conditions: i) distance within 500 km from the record under analysis; ii) availability of at least 10 monthly values in commune with it in the month of the gap and iii) the availability of

the datum that is missing in the record under analysis. If no records fulfilled these conditions, the missing datum was not estimated. The reference records fulfilling these conditions were ranked based on a weighting factor defined as the product of two weighting functions depending on the horizontal and vertical distances and the reference record with the highest rank was considered for the estimation of the missing datum. These weighting factors are based on Gaussian functions and they are defined as [10]:

$$w_{i,j}^{\mathrm{var}} = \exp\left[-\left(\left(\Delta_{i,j}^{\mathrm{var}}\right)^2 \middle/ c_{\mathrm{var}}\right)\right] \tag{1}$$

where i is the series to fill and j the reference series. The weight is calculated for two variables (*var* – position and elevation), $\Delta_{i,j}^{\mathrm{var}}$ is the absolute value of the horizontal or vertical distance between the reference series and the series to be completed. c_{var} is a coefficient which regulates the decrease of the weighting factor. c_{var} can be expressed in terms of the distance $\Delta_{1/2}^{\mathrm{var}}$ for which the weight should be 0.5 [10]:

$$c_{\mathrm{var}} = -\left[\left(\Delta_{1/2}^{\mathrm{var}}\right)^2 \middle/ \ln 0.5\right] \tag{2}$$

For the horizontal distance, we used a factor equal to 0.5 at 300 km, whereas for the vertical distance we applied the following relation:

$$\Delta_{1/2}^{h} = \begin{cases} 250m & h \leq 500m \\ h/2\,m & h > 500m \end{cases} \tag{3}$$

The introduction of a difference vertical distance factor allowed us using the few high-level sites in our database as much as possible to estimate missing data in high-level records. This is especially useful in winter in northern Italy, where SD/SSR can have significantly different behavior at low and high levels.

Once the reference series was selected, the gap was filled under the assumption of the constancy of the ratio between incomplete and reference series. This procedure has been checked with a leave-one-out approach. Specifically, the procedure was applied to all available data, i.e. each available value of each SD record was considered as missing and estimated by means of the procedure. Then, when original data and the re-estimated ones were compared, about 50% of the ratios between the original values and the estimated ones resulted between

16

0.95 and 1.05.

This gap filling method may however give rise to outliers resulting from peculiar climatic conditions of the reference series. In order to avoid this problem, the median of a set of five estimated values, corresponding to the five highest correlated reference records, is here selected instead of only the one with the highest rank. The difference between this new procedure and the one we used in [6] is evident for example observing the comparison between the series of the relative SD for the month of January for Milano Linate filled with the old method and the new one (Fig. 6).
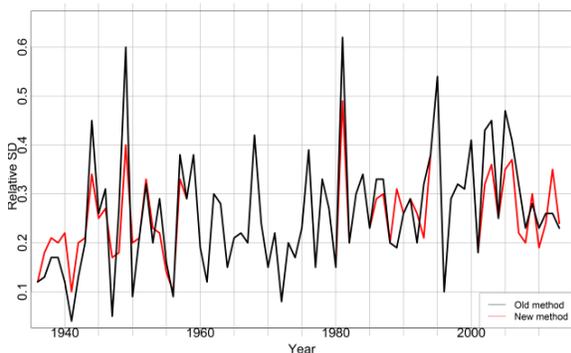

Fig 6. The black line represents the original series filled with the old method while the red line represents the original series filled with the new one.

The black line represents the original series filled with the old method while the red line represents the original series filled with the new method. The means over the whole period for the original series and the gap-filled series with the two methods do not change while variability of the series completed with the new method (deviation standard: 0.10) is less than the variability of the series completed with the old one (deviation standard: 0.12). In this paper, we use the new method to complete the two database. We used this new method also in [7] for SSR.

For SSR a reference series was considered when it was distant less than 500 km from the record under analysis, it had at least 6 monthly values in common with it in the month of the gap and it had the datum that is missing in the record under analysis. When less than five reference records fulfilling the requested conditions were available, the median was calculated with the available reference series.

After the gap filling, we deleted the series with less than 90% of available data over the 1984-2013 (1976-2005) period for the SD (SSR) database. This reduced the SD dataset to 95 records while for the SSR dataset, all series had at least 99% of available data during the 1976-2005 and so all the records were considered for the following analysis. The periods 1984-2013 and 1976-2005 were selected as reference period to calculate the relative monthly, seasonal and annual anomaly series,

respectively, for SD and SSR because they are those with the highest data availability after the gap-filling procedure. Then, starting from the gap-filled series we generated a 1°x1° gridded version of the dataset in order to balance the contribution of areas with a higher number of stations with those that have a lower station coverage. We followed the technique described by [2] and [11], which is based on an Inverse Distance Weighting approach with the addition of an angular term to take into account the anisotropy in spatial distribution of stations. The resulting grid spans from 7° to 19°E and from 37° to 47°N with 68 points over the Italian territory for SD [6] and with 58 points for SSR [7].

Finally, we calculated the Italian average monthly, seasonal and annual anomaly series from the gridded dataset simply by averaging all grid point anomaly records.

V.   LONG-TERM EVOLUTION OVER ITALY

As validation of the techniques applied to the original series, in Fig. 7 the average Italian seasonal and annual SD anomaly records is shown together with the corresponding low-pass filter (11 year window-3-year standard deviation Gaussian low-pass filter).

As already reported by [6] for northern and southern Italy, the curves indicate an increase starting from the 1980s (known in literature as "Brightening period" [12]). This signal concerns all seasons with the only exception of the autumn, that shows an increasing tendency only from the beginning of 1990s. Before this brightening signal, there is a decreasing tendency (known in literature as "Global dimming" [12]) that is however less intense than the following increase. It concerns especially spring and summer and the period comprised between the middle of the 1960s and the beginning of the 1980s. In the early period, from the 1930s to the 1950s, there is some evidence of an increasing tendency (known in literature as "Early brightening" [12]). This signal concerns, however, a period in which data availability is low (Fig. 4), causing a greater uncertainty in the average records.

The overall picture of Italian SD trend is, in good agreement with the trend observed in many areas of the world [12]. In particular, the resulting trend as reported by [6] turn out to be in good agreement with corresponding records from neighboring areas, as for example the southern part of the Greater Alpine Region [13] and eastern Spain [11].

SSR (Figure not shown) shows a similar evolution with a significant dimming from the beginning of the series to the mid-1980s and a brightening until the end of the series as reported for northern and southern Italy by [7].

VI.   CONCLUSIONS

Considering the importance of solar radiation not only for scientific interest a new database of SD and SSR has

been set up over the Italian territory for the 1936-2013 and 1959-2013 period, respectively.

In this paper, the main steps to obtain reliable mean series over the Italian territory starting from the raw series have been illustrated. In particular, it has been shown how the homogenization procedure is important, especially for variables like SD and SSR, underling how, at regional level, systematic biases in the original records can hide a significant part of the long term-trend. Beside the quality control, particular emphasis is placed also to show how the problem of the gaps in the series should be considered in order to calculate the regional series. This is true especially over areas with a complex orography like Italy.
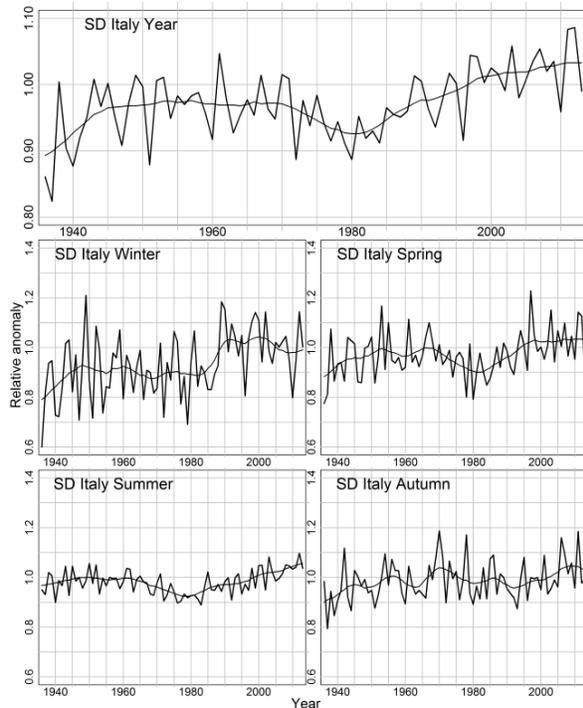

Fig. 7. Average annual and seasonal Italian SD anomaly series plotted together with the corresponding filter (11 year window-3-year standard deviation Gaussian low-pass filter). The anomalies are calculated with respect to the 1984-2013 period.

Finally, considering the length of the Italian SD records, seasonal and annual SD anomaly series over the area are presented as example of the results obtained applying the presented techniques to the original series.

REFERENCES

[1] E. Aguilar, I. Auer, M. Brunet, T. C. Peterson, J. Wieringa, "Guidelines on climate metadata and homogenization", World Climate Programme Data and Monitoring, WCDMP-No.53, WMO-TD No.1186, Geneva, 2003.

[2] M. Brunetti, M. Maugeri, F. Monti, T. Nanni, "Temperature and precipitation variability in Italy in the last two centuries from homogenised instrumental time series", Int. J. Climatol, 2006, vol.26, pp.345-381.

[3] WMO-No.8, "Solar radiation and sunshine duration", Guide to meteorological instruments and observing practices, chap.9, 3rd ed, World Meteorological Organization, Geneva, Switzerland, 1969.

[4] WMO-No.8, "Measurements of radiation", Guide to meteorological instruments and observing practices, chap.7, World Meteorological Organization, Geneva, Switzerland, 2008.

[5] WMO-No.8, "Measurements of sunshine duration", Guide to meteorological instruments and observing practices, chap.8, World Meteorological Organization, Geneva, Switzerland, 2008.

[6] V. Manara, M. C. Beltrano, M. Brunetti, M. Maugeri, A. Sanchez-Lorenzo, C. Simolo, S. Sorrenti, "Sunshine duration variability and trends in Italy from homogenized instrumental time series (1936-2013)", J. Geophys. Res. Atmos., 2015, vol.1, No.120, pp.3622-3641.

[7] V. Manara, M. Brunetti, A. Celozzi, M. Maugeri, A. Sanchez-Lorenzo, M. Wild, "Detection of dimming/brightening in Italy from homogenized all-sky and clear-sky surface solar radiation records and underlying causes (1959-2013)", Atmos. Chem. Phys. Discussion, 2016, doi:10.5194/acp-2016-206, in review.

[8] V. K. C. Venema et al., "Benchmarking homogenization algorithms for monthly data", Clim. Past, 2012, vol.8, pp.89-115, doi:10.5194/cp-8-89-2012.

[9] J. M. Craddock, "Methods of comparing annual rainfall records for climatic purposes", Weather, 1979, vol.34, pp.332-346.

[10] M. Brunetti, M. Maugeri, T. Nanni, C. Simolo, J. Spinoni, "High-resolution temperature climatology for Italy:Interpolation method intercomparison", Int. J. Climatol. 2013, vol.34, pp.1278-1296, doi: 10.1002/joc.3764.

[11] A. Sanchez-Lorenzo, M. Brunetti, M. Wild, J. Calbò, J. Martin-Vide, "Recent spatial and temporal variability and trends of sunshine duration over the Iberian Peninsula from a homogenized data set", J Geophys. Res., 2007, vol.112, D20115, doi:10.1029/2007JD008677.

[12] M. Wild, "Global dimming and brightening: A review", J. Geophys. Res., 2009, vol.114, D00D16, doi:10.1029/2008JD011470.

[13] I. Auer et al., "HISTALP-historical instrumental climatological surface time series of the Greater Alpine Region", Int. J. Climatol., 2007, vol.27, pp.17-46, doi:10.1002/joc.1377.