Solar Energy 158 (2017) 49–62

Contents lists available at ScienceDirect

Solar Energy

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

Quality control of global solar radiation data with satellite-based products

Ruben Urraca[a], Ana M. Gracia-Amillo[b,*], Thomas Huld[b], Francisco Javier Martinez-de-Pison[a], Jörg Trentmann[e], Anders V. Lindfors[f], Aku Riihelä[f], Andres Sanz-Garcia[c,d]

[a] EDMANS Group, Department of Mechanical Engineering, University of La Rioja, Logroño, Spain
[b] European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy
[c] University of Helsinki, Viikinkaari, 5 E, P.O. Box 56, 00014 Helsinki, Finland
[d] Institute of Advanced Biomedical Engineering and Science, Tokyo Women's Medical University, 8-1 Kawada-cho, Shinjuku-ku, Tokyo 162-8666, Japan
[e] Deutscher Wetterdienst, 63067 Offenbach, Germany
[f] Finnish Meteorological Institute, P.O. Box 503, FI-00101 Helsinki, Finland

ARTICLE INFO

Keywords:
Satellite-based models
Global horizontal irradiance
Quality control (QC)
Ground radiation records
Solar radiation

ABSTRACT

Several quality control (QC) procedures are available to detect errors in ground records of solar radiation, mainly range tests, model comparison and graphical analysis, but most of them are ineffective in detecting common problems that generate errors within the physical and statistical acceptance ranges. Herein, we present a novel QC method to detect small deviations from the real irradiance profile. The proposed method compares ground records with estimates from three independent radiation products, mainly satellite-based datasets, and flags periods of consecutive days where the daily deviation of the three products differs from the historical values for that time of the year and region. The confidence intervals of historical values are obtained using robust statistics and errors are subsequently detected with a window function that goes along the whole time series. The method is supplemented with a graphical analysis tool to ease the detection of false alarms.

The proposed QC was validated in a dataset of 313 ground stations. Faulty records were detected in 31 stations, even though the dataset had passed the Baseline Surface Radiation Network (BSRN) range tests. The graphical analysis tool facilitated the identification of the most likely causes of these errors, which were classified into operational errors (snow over the sensor, soiling, shading, time shifts, large errors) and equipment errors (miscalibration and sensor replacements), and it also eased the detection of false alarms (16 stations). These results prove that our QC method can overcome the limitations of existing QC tests by detecting common errors that create small deviations in the records and by providing a graphical analysis tool that facilitates and accelerates the inspection of flagged values.

1. Introduction

Solar radiation has been historically recorded at ground level by different meteorological agencies in order to provide reliable data for the assessment of the solar resource. These records are not only the most accurate source of solar radiation data, but are also crucial for validating satellite-based models, which are swiftly becoming the most widely used option to obtain spatial estimates of solar radiation (McArthur, 2005; Polo et al., 2016; Sengupta et al., 2015). Different parameters can be used to measure the amount and type of solar radiation reaching the Earth. The most common one is the global horizontal irradiance (G), i.e. the total shortwave incoming radiation received by a horizontal surface. More specialized monitoring stations can also measure the radiation components, i.e. the direct normal irradiance (B_N) and diffuse horizontal irradiance (D), providing more information about the type of radiation being received. Other parameters also recorded are the longwave radiation (upwelling and downwelling) and the sunshine duration, a parameter historically used to indirectly estimate G. However, in most cases the parameter normally recorded is only the G and the B_N and D are derived using decomposition models (Gueymard and Ruiz-Arias, 2015).

Measuring G is more prone to errors than other meteorological variables (Moradi, 2009). Younes et al. (2005) proposed the classification of these errors into two broad groups: equipment and operational errors. Equipment errors are inherent to the type of pyranometer used and the calibration applied, and include the zenithal error (cosine error), azimuthal error, stability, non-linearity, temperature dependence and spectral response. Highest-quality records are obtained with thermopile pyranometers, which are based on the thermoelectric effect. Within thermopiles, three levels of quality are established by the ISO

* Corresponding author.
E-mail address: ana.gracia-amillo@ec.europa.eu (A.M. Gracia-Amillo).

http://dx.doi.org/10.1016/j.solener.2017.09.032
Received 5 May 2017; Received in revised form 17 August 2017; Accepted 15 September 2017
Available online 22 September 2017

Nomenclature

CM-SAF	Satellite Application Facility on Climate Monitoring
B	beam/direct surface irradiance received on a horizontal plane
B_N	beam/direct surface irradiance received on a plane always normal to sun rays
BSRN	Baseline Surface Radiation Network
CAMS	Copernicus Atmosphere Monitoring System
CI	Confidence Interval
D	diffuse surface irradiance received on a horizontal plane
E_{0N}	solar constant adjusted to Earth - Sun distance
E_0	extraterrestrial irradiance received on a horizontal plane
ECMWF	European Center for Medium-range Weather Forecast
G	global surface irradiance received on a horizontal plane
K	diffuse ratio
KN	beam transmittance
KT	clearness index
MAD	Median Absolute Deviation
MAE	Mean Absolute Error
MFG	Meteosat First Generation
MSG	Meteosat Second Generation
n	parameter to adjust the level of restriction (width) of the CIs
PV	photovoltaic
QC	quality control
w	window width, i.e. number of consecutive days analyzed at once
WMO	World Meteorological Organization

Greek letters

δ	deviation (estimated - observed)
θ_S	solar zenith angle

Subscripts

d	day
g	group of stations - spatial group
h	hour
m	total months of the time series
m′	twelve months of the year (Jan. to Dec.)
s	station

Superscript

p	product

9060:1990 (ISO, 1990) and World Meteorological Organization (WMO) (WMO, 2008): (i) Secondary Standard or High quality, (ii) First Class or Medium quality and (iii) Second Class or Low Quality. A low-cost option to record solar radiation is the use of radiometers based on the photovoltaic effect, such as silicon-based photodiodes and solar-reference cells. However, they are considered not compliant with the quality rules of the ISO 9060:1990 due to the limited spectral response of the silicon (400–1000 nm). On the other hand, operational errors are independent of the type of sensor and involve different factors such as shading by nearby objects, dew, frost, snow or dust (soiling) covering the dome of the pyranometer, incorrect leveling, station shut-downs, electric fields in the vicinity of cables or a malfunction in the datalogger, among others. An adequate selection of the place to install the pyranometers, as well as a regular maintenance, can prevent most of these operational errors. Another classification proposed by Zahumenský (2004) distinguishes between random errors, which are symmetrically distributed around zero, systematic errors, asymmetrically distributed, large errors mainly caused by malfunctions of the devices and errors in data processing, and micrometeorological errors, which are incoherences of the ground records compared to the surrounding regions. Overall, all types of error introduce a certain degree of uncertainty in the radiation measurements and applying a quality control (QC) procedure becomes an essential step before using ground datasets.

Many QC methods have been proposed by meteorological agencies and independent researchers. Some well-known examples are the QC tests from the Baseline Surface Radiation Network (BSRN) (Long and Dutton, 2002), the MESOR recommendations (Hoyer-Klick et al., 2008), the NREL SERI QC procedure (NREL, 1993), the QCRad methodology (Long and Shi, 2008) or the web-based services from MINES ParisTech (Geiger et al., 2002) and AQC test (Molineaux and Ineichen, 2003). These QC procedures flag those samples identified out of the normal ranges of data and usually leave the decision of removing flagged cases to the user. They can be classified in four broad categories (Ohmura et al., 1998): range tests (physically and extremely rare limits), across-quantities relationships, model comparisons and graphical analysis. The least restrictive level of most QC procedures is a range test based on the physically possible limits, with the upper limit being equal to the extraterrestrial irradiation (E) and a lower limit lying within −4 and 0 W/m^2 (Long and Dutton, 2002; Hoyer-Klick et al., 2008; Long and Shi, 2008). In a second step, the physical ranges are narrowed imposing more strict conditions. The upper limit is usually reduced with estimations from a clear-sky model (Geiger et al., 2002; Journée and Bertrand, 2011; Hoyer-Klick et al., 2008; Younes et al., 2005) such as the ESRA clear-sky model (Rigollier et al., 2000), the Page model (Page and Lebens, 1986) or the Bird clear-sky model (Bird and Hulstrom, 1980, 1981). Simulation with these clear-sky models are usually carried out under clean atmospheric conditions (aerosols and water vapor set to 0) instead of using estimated or climatological values. The lower limit is increased up to the level of extremely overcast conditions. This is typically imposed with the dimensionless clearness index ($KT = G/E$), with values around 0.03 (Geiger et al., 2002), or with the modified clearness index (Perez et al., 1990). Other approaches to reduce the acceptance ranges are to use climatological values (Long and Shi, 2008), to interpolate records from nearby locations or to use estimations from meteorological variables (Tang et al., 2010), mainly sunshine duration (Journée and Bertrand, 2011; Moradi, 2009; Muneer and Fairooz, 2002). Other QC tests also check the stability of the time series generated, analyzing the step between consecutive samples (Journée and Bertrand, 2011). Besides, some authors have proposed QC methods tailored to detect time shifts by analyzing the symmetry between morning and afternoon records (Ineichen, 2013) and by using graphical analysis (Moreno-Tejada et al., 2015). When the diffuse or direct components are also available, coherence or consistency tests are commonly imposed (Long and Dutton, 2002). Some authors have proposed the use of envelope tests in a dimensionless k-space consisting of the clearness index, the direct-beam clearness index ($KN = B/E$) and the global-to-diffuse ratio ($K = G/D$). These envelope tests are based on setting empirical or statistical limits either on the KT−K space or the KT−KN one (Younes et al., 2005; Journée and Bertrand, 2011; NREL, 1993; Pashiardis and Kalogirou, 2016) and the subsequent graphical analysis of the envelopes obtained.

The majority of these tests are designed to detect only large deviations in ground records, while some of the most common errors just introduce small deviations from the real irradiance profile. These faulty records pass most tests because they are acceptable from a statistical or a physical perspective, but they still have a negative influence on the assessment of solar radiation. Our aim is to develop a new QC

procedure to detect these types of small but common deviations by taking advantage of the temporal uniformity of radiation products, mainly satellite-based estimations. This approach is already used in some available applications such as the Copernicus Atmosphere Monitoring System (CAMS), which quality controls their operational satellite-based estimations by comparing new estimations against the historical values of the time series (Schroedter-Homscheidt et al., 2016). However, to the authors knowledge, no QC implementation based on the comparison of ground records against several radiation products exists. Our proposal analyzes groups of consecutive days at a time with a window function, and flags those periods where the daily deviations of three independent products differs from the historical values for that time of the year and region. This approach narrows the acceptance ranges enabling the detection of small but constant deviations (systematic errors) in the irradiance profiles. The method focuses on the daily G because it is the parameter available in most ground stations and radiation products, and it is supplemented with a graphical analysis tool to help the user decide whether flagged values by the proposed QC method have to be removed from the dataset. The QC method has been validated with three radiation products in a ground dataset composed by 313 stations over Europe with records from 2005 to 2015. The study is divided into four parts: a sensitivity analysis to determine the best configuration for the QC method (Supplementary Material 1); a description and classification of the errors found in the ground dataset; a brief analysis of the influence of these periods on a validation of a satellite-based product; and a discussion of the limitations of the method.

The organization of the paper is as follows. Section 2 introduces the ground dataset along with the radiation products used in the QC procedure. In Section 3, the QC procedure is presented. Section 4 describes the steps followed to validate the QC method. In Section 5, the results obtained are shown and discussed. Finally, in Section 6 the main conclusions and remarks are drawn.

2. Data

The validation dataset is composed of ground records from 313 monitoring stations located over Europe (Fig. 1). The stations belong to different meteorological and agricultural networks (Table 1). In all cases, the variable retrieved is the global horizontal irradiance from 2005 to 2015 at the highest temporal resolution available.

The type of pyranometer used to obtain the ground records differs between networks, between stations of the same network, and even within the same station due to the replacement of old sensors by newer ones. The majority of stations have Secondary Standard thermopile pyranometers from Kipp & Zonen (CM11/CMP11, CM21, CM22/CMP22), which generate the highest quality records. This is the case for BSRN stations (BSRN, 2016) and most meteorological networks such as the Finnish Meteorological Institute (FMI) (FMI, 2016), the Swedish Meteorological and Hydrological Institute (SMHI) (SMHI, 2016), the Deutscher Wetterdienst (DWD) (DWD, 2016), Meteo-France (Meteo France, 2016) and Met Office (Met Office, 2017). Most of these pyranometers are ventilated, which prevents the accumulation of snow and frost over the pyranometer outer dome. Secondary Standard pyranometers are also installed in most stations from the Landbruksmeteorologisk Tjeneste (LMT) (NIBIO-LMT, 2016), a project run by the Norwegian Institute of Bioeconomy Research (NIBIO), and in the JRC in Ispra (Italy) (JRC, 2016). Some records from Second Class sensors are still present on Met Office and LMT networks. Finally, the Spanish agricultural network, i.e. Servicio de Informacion Agroclimatica para el Regadio (SIAR) (SIAR, 2015), uses silicon-based photodiodes, the SP1110 [Skye Instruments].

The radiation products used are two satellite-based products, SARAH-JRC and CLARA-A2, and one reanalysis dataset, ERA-Interim. The two satellite-based products are Climate Data Records (CDRs) developed by the Satellite Application Facility on Climate Monitoring (CM-SAF). The SARAH-JRC is a variation of the product available at CM SAF webpage, SARAH-1 (Muller et al., 2015), and it is used by the Joint Research Centre (JRC) in PVGIS (PVGIS, 2016), an on-line tool to asses the PV output power. The product processes images from the Meteosat First and Second Generation (MFG and MSG) to provide hourly estimations within the METEOSAT disk since 1983 with a spatial resolution of $0.05° \times 0.05°$. While SARAH-1 takes half-hourly images and then averages the estimations to generate hourly values, SARAH-JRC hourly estimates are obtained by evaluating just one satellite image per hour. CLARA-A2 (Karlsson et al., in press) is the second edition of CM SAF dataset based on records from AVHHR instruments onboard polar orbiting satellites NOAA- and METOP- and has a spatial resolution of 0.25°. The product provides daily estimates with global resolution since 1982, but data is not available over snow-covered surfaces. The ERA-Interim reanalysis (Dee et al., 2011) has been developed by the European Center for Medium-range Weather Forecast (ECMWF) and provides global coverage with 3-hourly resolution in a regular longitude-latitude grid of $0.75° \times 0.75°$. ERA-Interim surface irradiance product has lower spatial resolution and accuracy than the two satellite-based products (Bojanowski et al., 2014; Urraca et al., in press), and shows a significant positive bias caused by the overestimation of clear-sky conditions (Jones et al., in press). The opposite effect is also true but less pronounced, so it just partly mitigates the positive bias (Boilley and Wald, 2015). Despite this, ERA-Interim is included in the QC method because it has global coverage with no gaps, making it the only available estimation in places or periods where the two satellite products are missing.

3. Quality control procedure

The proposed QC method is a semi-automatic procedure based on the statistical analysis of the daily deviations between a set of radiation products and ground records. First, the daily deviations of each product

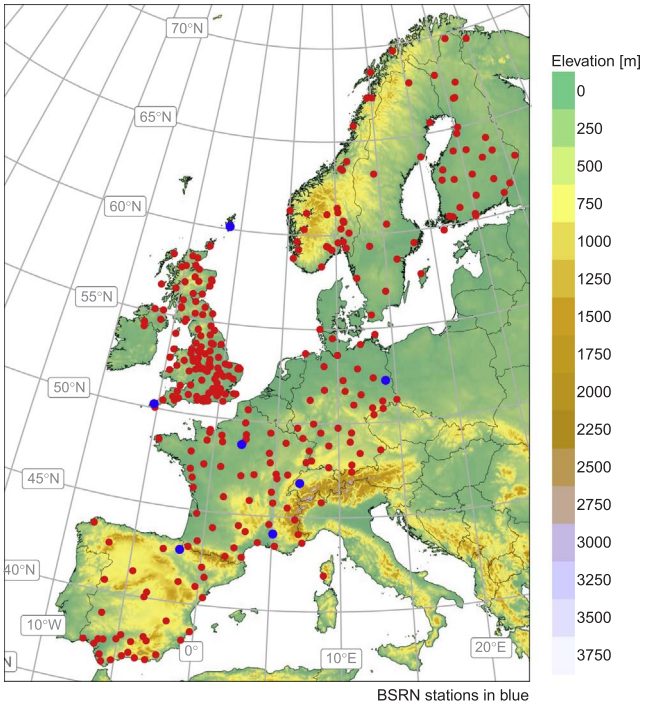

Fig. 1. Locations of the 313 ground stations.

Table 1
Description of the networks included in the ground dataset.

Network	Type	Stations	Resolution	Type of pyranometer			
				Secondary standard	Second class	Photodiode	Not reported
Met Office [UK]	Meteo	121	hourly	85	9	–	27
LMT [NOR]	Agro	29	hourly	28	1	–	–
SMHI [SWE]	Meteo	12	hourly	12	–	–	–
FMI [FIN]	Meteo	27	hourly	27	–	–	–
DWD [GER]	Meteo	34	hourly	34	–	–	–
Meteo France [FRA]	Meteo	49	hourly	49	–	–	–
SIAR [SPA]	Agro	33	half-hourly	–	1	32	–
BSRN	–	7	minutely	7	–	–	–
JRC-Ispra [IT]	–	1	minutely	1	–	–	–
Total		313		243	11	32	27

are characterized, obtaining the confidence intervals (CIs) which these deviations typically lie at each location and time of year. Second, a window function that goes along the whole time series flags groups of consecutive days where the deviation of all products is over or under the CI limits. Third, two plots are automatically generated to visually inspect the flagged days and determine whether they are true errors or false alarms. These plots are the time series of the daily deviations between products and ground records (daily deviation plot), and the comparison of hourly irradiance profiles of ground stations and products (hourly irradiance plot). All products selected must be uniform in time to enable a proper characterization of the bias for a specific location and time of the year. The method is designed to work with daily averaged ground records, because most of these products are available at daily resolution.

3.1. Definition of the confidence intervals

The first step is to characterize the deviations between estimations and measurements in space and time. Deviations (δ) are calculated by subtracting ground records from the products estimates (Eq. (1)).

$$\delta_t = y_t - o_t \tag{1}$$

where y are the estimations, o the measurements (observed values) and t the temporal resolution. The characterization is handled from a statistical perspective by defining the CIs within which the daily deviations of each product lie. These CIs are calculated on a monthly basis (temporal averaging) for groups of stations with similar characteristics (spatial averaging). The averaging is performed in two steps in order to increase the robustness of the CIs against noise. Initially the bias with respect to the median ($\widehat{Bias}$) of the daily deviations is calculated for each month and year of the time series (m) at every location s (Eq. (2)).

$$\widehat{Bias}^{\,p}_{m,s} = median^{\,p}_{m,s}(\delta^{\,p}_{d,s}) \qquad 1 \leqslant m \leqslant M,\ 1 \leqslant s \leqslant S,\ p \in products \tag{2}$$

where M are the total months of the time series and S the number of stations. Subsequently, these values are again averaged by grouping the twelve months of the year m' (temporal averaging) and stations within the same spatial group g (spatial averaging), resulting in a unique set of twelve values per group and product (Eq. (3)).

$$\begin{aligned} \widehat{Bias}^{\,p}_{m',g} &= median^{\,p}_{m',g}(\widehat{Bias}^{\,p}_{m,s}) \qquad m' \in (\text{Jan},\ldots,\text{Dec}),\quad 1 \leqslant g \leqslant GR,\ p \in products \\ MAD^{\,p}_{m',g} &= 1.4826 \times median^{\,p}_{m',g}(|\widehat{Bias}^{\,p}_{m,s}|) \qquad m' \in (\text{Jan},\ldots,\text{Dec}),\quad 1 \leqslant g \leqslant GR,\ p \in products \end{aligned} \tag{3}$$

where GR is the number of spatial groups. The median is preferred instead of the mean as the measure of central tendency, while the Mean Absolute Deviation (MAD) is the measure of dispersion used. The MAD

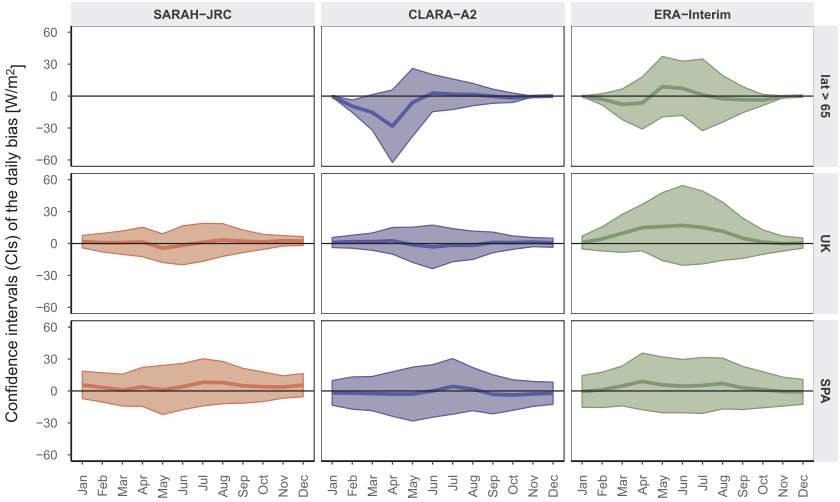

Fig. 2. Example of the CIs within which the historical bias of the three radiation products lies. The CIs shown were calculated with $n = 2$ for the following three groups of stations, from top to bottom: latitude above 65°N, UK, and Spain. Note that SARAH has no coverage at latitudes above 65°N.

includes a constant scale factor of 1.4286 to ensure the consistency of estimates for different sample sizes. The use of this constant value and the median makes this statistic more independent of the sample size and a more robust measure of variability than the standard deviation (Leys et al., 2013). Finally, the CIs are obtained with Eq. (4).

$$CI^{p}_{m',g} = \widehat{Bias}^{p}_{m',g} \pm n \times MAD^{p}_{m',g} \qquad m' \in (\text{Jan},...,\text{Dec}), \quad 1 \leqslant g \leqslant GR,\ p \in products \tag{4}$$

where n is the coefficient that weights the MAD in order to adjust the restriction level of the QC procedure.

Seven spatial groups are made ($GR = 7$): Nordic region above 65°N ($lat > 65$, 14 stations), Norway below 65°N ($NOR_{lat<65}$, 23 stations), Finland and Sweden below 65°N ($SWE{-}FIN_{lat<65}$, 31 stations), Germany (GER, 35 stations), France along with the stations in North Italy and Switzerland (CENTRAL, 53 stations), United Kingdom (UK, 123 stations) and Spain (SPA, 34 stations). One group is roughly made for each country, making additional splits only if a significant number of stations within a country is under special climatological conditions. This is the case of the region above 65°N, which presents strong intra-annual variations with low solar elevation angles in winter, seasonal snow and low viewing angle of satellites, causing the failure of most radiation products. Besides, a minimum number of stations per group is required to use robust statistics. This is why stations from Finland and Sweden below 65°N were grouped together. An additional group could have been made for mountain regions due to the particular performance of radiation products there, but the number of stations was too low (5 stations in the Alps and Pyrenees). These stations are kept with the rest of the country (1 Germany, 4 France), and they are treated as false alarms if flagged by the QC method.

The CIs of each group should be ideally calculated only with high-quality records, but in this study records from all type of sensors are used because in the regions with low-quality sensors these are the only records available. The CIs are initially calculated with all stations except the five high-elevation locations, and the QC is run to detect the most significant error. Then, the CIs are recalculated excluding these large errors, being this the CI version used for operational purposes throughout the study. Some stations with minor errors or particular climatological conditions can still be present in the dataset used to obtain the CIs, but the robust statistics described above are able to ignore these cases and gather the general trend of the spatial group. Fig. 2 shows the CIs for the three radiation products and three of the spatial groups using $n = 2$.

3.2. Flag biased periods: window function

Once the CIs for the daily bias are defined, a window function (Algorithm 1) is run through the time series to flag periods of consecutive days where the daily bias is predominantly over or under the CI limits. The number of days analyzed is set with the window width parameter (w), which is considered an input for the algorithm. The window start date increases in steps of five days (fast moving filter), due to the large overlap between two consecutive windows, speeding up the whole process. For each group of days, the percentage of valid samples ($dvalid$), samples over the upper limit ($dover$) and samples under the lower limit ($dunder$) are computed for each product. Days with an absolute deviation under 5 W/m^2 or an absolute relative deviation under 5% are not accounted for calculating $dover$ and $dunder$. Then, the algorithm checks that there is enough data in the period analyzed, removing products with more than 80% of days missing and checking that at least one product spans almost the whole period (less than 20% of missing days). In that case, the function flags the whole period as potentially problematic if more than 80% of the valid samples in this period are over the CI_{up}, or if more than 80% of the samples are under the CI_{low}.

Algorithm 1. Window function for a specific station

inputs:
- δ^{p}_{d} = time series of the daily deviations for each product.
- $r\delta^{p}_{d}$ = time series of the daily relative deviations for each product.
- $CI^{p}_{m'}(n)$ = confidence intervals of δ_d for each product, spatial group and month.
- $CI^{p}_{up,m'}(n)$ = upper limit of CI.
- $CI^{p}_{low,m'}(n)$ = lower limit of CI.
- w = window width, i.e. the number of consecutive days analyzed at a time.
- n = parameter to adjust the level of restriction of the CIs based on the number of days analyzed (w).

variables:
- p = radiation product, with $p\in$ (SARAH-JRC, CLARA-A2, ERA-Interim).
- m' = month of the year, with $m' \in (Jan,...,Dec)$.
- d = day of the time series.
- d_{ini} = first day of the window.
- d_{end} = last day of the window.
- d_{valid} = days with no missing δ_d in %.
- d_{over} = days with δ_d over CI_{up} in %.
- d_{under} = days with δ_d under CI_{down} in %.

constants:
- $step = 5\ days$.
- $r\delta_{d,min} = 5\%$.
- $\delta_{d,min} = 5$ W/m^2.
- $dvalid_{min} = 20\%$.
- $dvalid_{complete} = 80\%$.
- $maxout = 80\%$.

Begin
  for $d = 1$ to length of the time series **with** $step$ **do**
    $d_{ini} \leftarrow d$
    $d_{end} \leftarrow d + w{-}1$
    for all $p \in products$ **do**
      $dvalid^{p} = 100\frac{\sum_{d=d_{ini}}^{d_{end}} if\ (\delta^{p}_{d}\ !=\ VOID)}{w}$
      $dover^{p} = 100\frac{\sum_{d=d_{ini}}^{d_{end}} \delta^{p}_{d} > CI^{p}_{up,m'}(n)\ \textbf{with}\ |\delta^{p}_{d}| > \delta_{d,min}\ \textbf{AND}\ |r\delta^{p}_{d}| > r\delta_{d,min}}{w}$
      $dunder^{p} = 100\frac{\sum_{d=d_{ini}}^{d_{end}} \delta^{p}_{d} < CI^{p}_{low,m'}(n)\ \textbf{with}\ |\delta^{p}_{d}| > \delta_{d,min}\ \textbf{AND}\ |r\delta^{p}_{d}| > r\delta_{d,min}}{w}$
    end for
    if $dvalid^{p} < dvalid_{min}$ **then**
      Discard product p
    end if
    if $any(dvalid^{p}) > dvalid_{complete}$ **then**
      if $mean(dover) \geqslant maxout$ **OR** $mean(dunder) \geqslant maxout$ **then**
        Flag period (d_{ini}: d_{end})
      end if
    end if
  end for
End

3.3. Graphical analysis

The majority of QC methods (Younes et al., 2005) flag samples with a high probability of error, but leave to the user the decision of removing the samples from the dataset. Unlike these procedures, our QC

method includes a graphical analysis tool to verify that flagged values are days with real issues in the pyranometer data. This tool is composed of two graphical outputs. The first plot is the time series of the daily deviation between products and ground records (daily deviation plot), while the second one compares the hourly irradiance profile of ground sensors against the ones of satellite-based products (hourly irradiance plot). In both plots, the ground records flagged by the QC method are shaded. These graphical outputs enable the detection of false alarms, and the hourly irradiance plot allows finding the cause of the error. Note that this second plot can only be generated if one of the radiation products has hourly resolution. Hence, it is convenient to include at least one product with hourly resolution, despite the fact that the QC method works with daily means. Section 5.1 presents more details on how the two plots look like for the most common types of error.

4. Implementation

Initially, the BSRN range tests (BSRN, 2016), namely the "extremely rare limits" and the "physically possible limits", are applied to the G records at its original temporal resolution (Eq. (5)).

$$\begin{aligned} &\textit{Physically possible limits}: \ -4\ \mathrm{W/m^2} < G < E_{0N}\cdot 1.5\cdot\cos(\theta_S)^{1.2} \\ &\qquad\qquad + 100\ \mathrm{W/m^2} \\ &\textit{Extremely rare limits}: \ -2\ \mathrm{W/m^2} < G < E_{0N}\cdot 1.2\cdot\cos(\theta_S)^{1.2} + 50\ \mathrm{W/m^2} \end{aligned} \tag{5}$$

where θ_S is the solar zenith angle and E_{0N} the solar constant adjusted to Earth - Sun distance. This removes isolated samples falling outside the normal irradiance range. The number of samples flagged with the BSRN tests is normally small and they are automatically eliminated due to the large size of the database. These tests can be only applied if ground records have sub-daily resolution, otherwise this step is skipped. Ground records are then integrated from its original resolution (minute resolution, half-hourly, hourly) to daily means. The missing values are kept in the dataset and no gap-filling method is used in order not to mask potential errors in the recording process. In the case of radiation products with sub-daily resolution (SARAH-JRC and ERA-Interim), the daily means are calculated by summing the hourly and three-hourly values respectively.

The most important setup parameters of the QC method are w and n. The optimum value of both parameters is determined by a sensitivity analysis (see Supplementary Material 1). The final configuration consists in running the window-function two times (Fig. 3) with the following settings:

- Run 1: w = 20 days and n = 2.4. Search of problems that cause large bias in ground records. These samples are represented in orange in the graphical analysis.
- Run 2: w = 90 days and n = 0.4. Search of quasi-permanent problems in the sensors that cause small deviations in the records, such as the case of calibration issues. These samples are shown in gray.

The QC method is applied to the 313 ground stations using the configuration based on two runs, and the different errors found are classified. Example plots illustrate the visual inspection phase for each error. Finally, the influence of faulty records in the validation of radiation products is quantified. This analysis is conducted in terms of the bias and the mean absolute error (MAE), given by Eqs. (6) and (7) respectively.

$$Bias = \frac{1}{N}\sum_{t=1}^{N}(y_t - o_t) \tag{6}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_t - o_t| \tag{7}$$

All computations were implemented in the freely distributed statistical software `R` (R Core Team, 2014). The core work of data manipulation and visualization was made with the set of packages from `tidyverse` (Wickham, 2016). Time series were handled with `lubridate` (Grolemund and Wickham, 2011) package, while spatial objects were manipulated with `sp` (Pebesma and Bivand, 2005), `raster` (Hijmans, 2015) and `rgdal` (Bivand et al., 2016). Finally, solar position calculations were performed with the functions of the `solaR` (Perpiñán, 2012) packages.

5. Results and discussion

5.1. Type of errors identified

The QC method detected faulty periods in 47 out of the total 313 stations (Table 2). After a visual inspection, it was confirmed that problems existed in 31 stations while the periods of the remaining 16 stations were false alarms. In addition, the QC method was not able to detect faulty records in 3 stations. Note that all ground records were unlabeled when retrieved from the different monitoring networks. In order to quantify the accuracy of the method, the samples were classified between acceptable or faulty data by inspecting the two plots included in the quick inspection phase for all stations. All samples labeled as faulty presented a clear defect, but still some additional errors

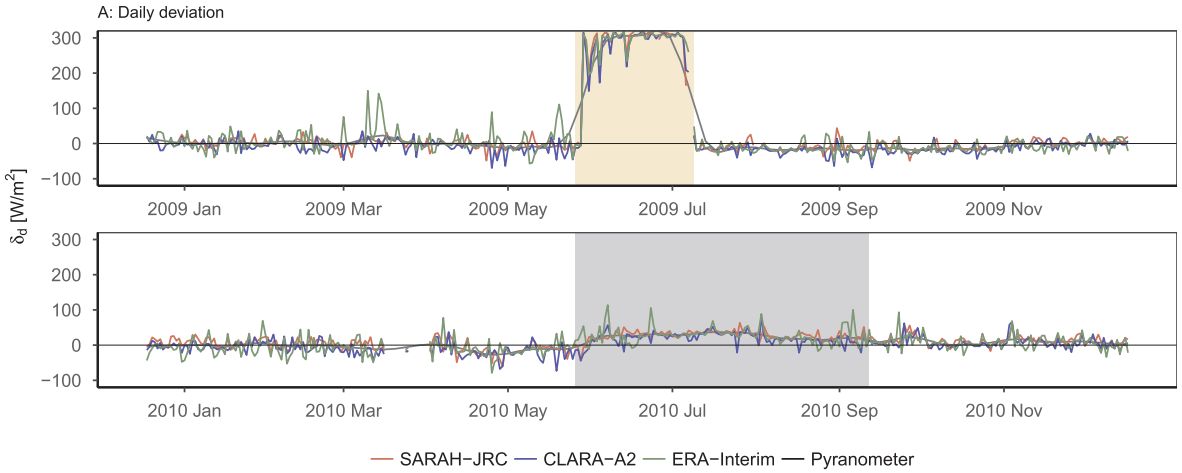

Fig. 3. Example of the periods flagged with the QC method. The orange area depicts the run looking for large short-lived biases (w = 20 days, n = 2.4), while the gray area shows the run looking for small permanent biases (w = 90 days, n = 0.4). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Description of periods flagged by the QC method.

		Type of issue	n [days]	Station [days]
Detected errors	Operational	Snow/Frost	4 [112]	NOR-124 [28], FIN-180 [39], FIN-185 [28], FRA-267 [17]
		Soiling	7 [2358]	UK-20 [42], UK-72 [225], FIN-181 [95], SPA-251 [1799] SPA-256 [100], FRA-295 [47], FRA-301 [50]
		Shading	11 [2619]	UK-55 [221], UK-112 [365], NOR-138 [83], NOR-142 [730] SPA-228 [18], SPA-232 [20], ITA-264 [49], FRA-265 [730] FRA-285 [123], NOR-147 [250], FIN-185 [30]
		Time shift	3 [966]	SPA-230 [885], SPA-236 [40], SPA-247 [41]
		Zero periods	2 [32]	UK-43 [18], UK-71 [14]
		Large errors	1 [162]	NOR-148 [162]
	Equipment	Miscalibration	2 [2035]	GER-207 [577], SPA-234 [1458]
		Sensor replacement	1 [1418]	UK-19 [1418]
Undetected errors	Operational	Shading	1 [1784]	UK-105 [1784]
	Equipment	Sensor replacement	2 [1462]	UK-31 [671], UK-36 [791]
False alarms		Snow-covered surface	3 [845]	FIN-191 [20], GER-224 [655], FRA-269 [170],
		Uncertainties - photodiodes	11 [1410]	SPA-229 [405], SPA-231 [215], SPA-238 [115], SPA-239 [140] SPA-243 [140], SPA-249 [110], SPA-252 [20], SPA-254 [20] SPA-255 [20], SPA-256 [105], SPA-257 [120]
		Uncertainties - slight deviations	2 [145]	FRA-280 [20], FRA-283 [125]

could exist in the false alarm group and within the assumed acceptable samples.

The detected errors were classified in two groups, operational and equipment errors, and within each group, the most likely cause of the error was proposed. False alarms are kept in the datasets, and most of them were caused by the uncertainties in the photodiodes and because the deviations observed were not strong enough.

Most of the detected faulty records were operational errors. The deviations caused by these errors are typically larger that the ones caused by equipment errors, and hence they can be detected more easily. The most common cause is that the sun is being blocked, either by an external object (shading) or by something over the sensor (snow/frost, soiling). Another common issue is the presence of days with some hourly slots set to zero, which is probably caused by a failure in the data logger as the irradiance profile during the rest of the day is acceptable. Data loggers are also the most probable cause of time shifts, which was a defect detected in three stations. The last type of error are large errors in the sensor or during data processing. Despite involving the sensor, this is not considered an equipment error because the deviations are random, short in time, and the sensor works optimally prior and after the faulty records.

Equipment errors are more difficult to detect than operational errors because they result in small systematic deviations that can span over the whole time series. If this deviation affects only a portion of the time series and the sensor works well prior or after this period, a re-calibration or a replacement of the sensor are the most common causes. This happened with three UK stations, where Second Class pyranometer were replaced by Secondary Standards, and GER-207 [577] and SPA-234 [1458], where the deviation was corrected without replacing the sensor implying a re-calibration of the sensor. Checking the historical station metadata is essential to determine the cause of an equipment error, even though this information is usually limited to the replacement dates of the sensor and it rarely includes the schedules of calibration routines.

We have classified as false alarms the samples flagged by the QC procedure but not removed from the database. These comprise actual false alarms, periods where the deviations observed are not caused by a faulty record (3 stations), and uncertain cases (13 stations), where an error might exist but after visual inspection we cannot guarantee that it is indeed an actual error and hence cannot be removed. False alarms due to the design of the QC method appear only at high elevation stations and in locations with seasonal snow (3 stations). The QC method flags the underestimation of most products during the snow season, because the CIs have been calculated with a group of stations mostly located in low-lying areas where this problem does not appear. Concerning the cases where a deviation exists but is not strong enough,

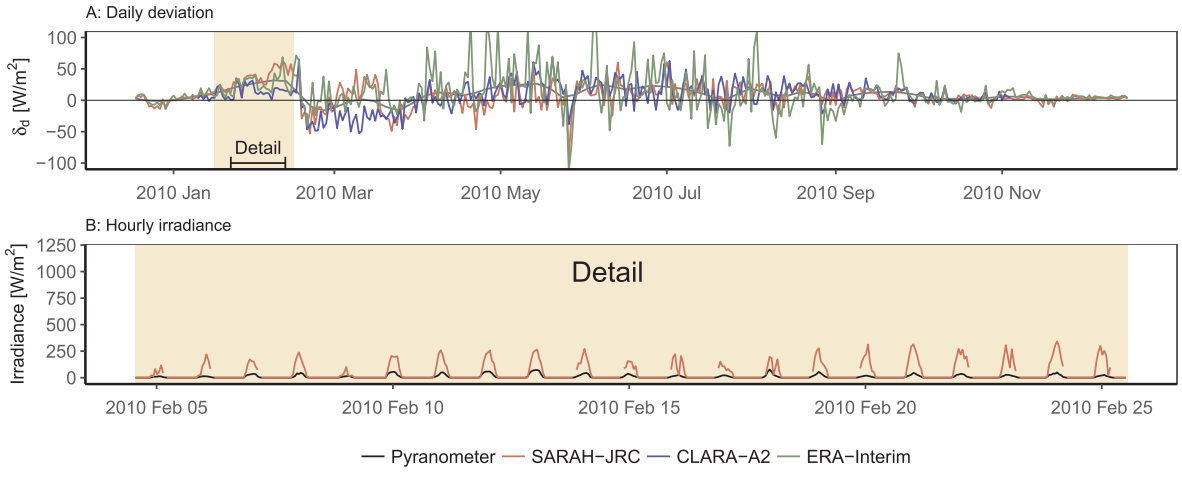

Fig. 4. Operational error: snow/frost covering the sensor (Apelsvoll - LMT). The orange area depicts the run looking for large short-lived biases ($w = 20$ days, $n = 2.4$). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

most of them come from the instabilities of the silicon-based photodiodes installed in the Spanish agricultural network (11 stations). Deviations observed are significant but are randomly distributed in space and time, and higher quality records from Secondary Standard pyranometers would be essential to determine whether these periods are true errors or false alarms.

The typical outcome of QC methods is a numeric flag indicating that the record is potentially erroneous. This flag usually takes different values depending on the level of testing where the sample has failed (Long and Dutton, 2002; Long and Shi, 2008). However, the interpretation of the flags along with the removal of the sample from the dataset is left to the user. This is not a trivial task as it requires further data processing. Our method makes this task easier by including a graphical analysis tool that automatically generates two plots when faulty records are found in a station. These plots are the time series of daily deviations and the hourly irradiance profiles, and in both plots faulty records are highlighted. Hereafter we depict and describe how these two plots look like for the different types of errors found in the case of study (Figs. 4–10).

Snow or frost covering the dome of pyranometers is a common defect in regions with seasonal snow, such is the case of high-latitude regions and high-elevation mountains. Most stations in Finland, Norway and Sweden install ventilated pyranometers to improve the evaporation of water and snow, but the accumulation of snow cannot be completely prevented. This type of error is found during January and February (Fig. 4A), and it is identified as a period in which the radiation products are larger than the ground records. The difference is low, around 50 W/m^2, due to the low irradiance, and this defect typically lasts around one month. This issue is rarely detected by the traditional QC tests, either empirical or statistical, because the sensor produces low records but still plausible under cloudy conditions.

Other substances typically referred to in the literature as soiling, such as dust or bird droppings, can also cover the sensor. The consequence is an underestimation of the irradiance by the sensor similar to snow, but the pattern observed is quite different. The deviations can appear in any month of the year and they prevail for longer periods, until the sensor is cleaned. Besides, this daily deviation increases from 20 up to 100 W/m^2 (Fig. 5A), because the dust steadily accumulates on the dome of the pyranometer until it is cleaned, either by the maintenance responsible or by the rain. This hinders the detection of this error using the traditional QC tests because the records obtained resemble the ones of a cloudy day. Care must be taken when the daily deviation observed is low, because even analyzing the intra-daily profiles it is difficult to conclude if the positive deviation is caused by soiling or by an actual overestimation of the radiation product. Moreover, most radiation products typically overestimate under high-aerosol conditions (Müller and Träger-Chatterjee, 2014), which also leads to a higher probability of soiling being accumulated on the sensor. In these cases, the length of the overestimation period can be used to identify the most likely cause, because high-aerosol periods are usually shorter than soiling problems. However, there are still some cases in which it was not possible to determine with certainty the cause of the error, as it occurred with 11 stations of the SIAR network. The overestimations observed were small, randomly distributed in time while they appeared in places with high probability of periods with high aerosol content. Besides, these records were obtained with silicon-based photodiodes installed in agricultural stations, so the quality of the sensor is low, and the location of the stations as well as the maintenance procedures might not be the most adequate ones. Consequently, these cases were not removed, and high quality records from a well-maintained meteorological network would be required to give a more detailed analysis of these periods.

Shading is another effect that leads to an underestimation of irradiance, in this case because the sun is being blocked by artificial or natural objects. This defect can affect the whole time series (poor selection of the location), or emerge during the operational period (poor site maintenance). Shading is differentiated from snow or soiling because the underestimation typically follows a regular pattern in time being more pronounced in months with low solar elevation angles Fig. 6. The shades are clearly identified in the hourly irradiance plot (Fig. 6B), because the sensor works properly around noon hours while a systematic underestimation appears at sunrise or sunset. The defect gets masked when working with daily or monthly aggregated values because the records resemble the ones of a low radiation day, passing most QC methods.

Note that there are some cases where the shadows can affect not only the pyranometer but also the surrounding region that has to be characterized with the ground station. This can happen at high-latitude regions or deep valleys. Here, the user has to decide whether to keep or remove the shadows depending on the end-use of the records. For instance, shadows should be kept in the datasets for an evaluation of the solar power capacity of the region, while they should be removed for the validation of radiation products, because they are generally focused on the modeling of global atmospheric processes.

The existence of some hourly slots set to zero is a defect observed in two stations. The pyranometer produces acceptable records during the rest of the day, so the most probable cause of the error should be the data logger, which integrates and saves the instantaneous records. This error results in an abrupt and random underestimation in the daily bias plot (Fig. 7A), and the cause of this underestimation is easily identified

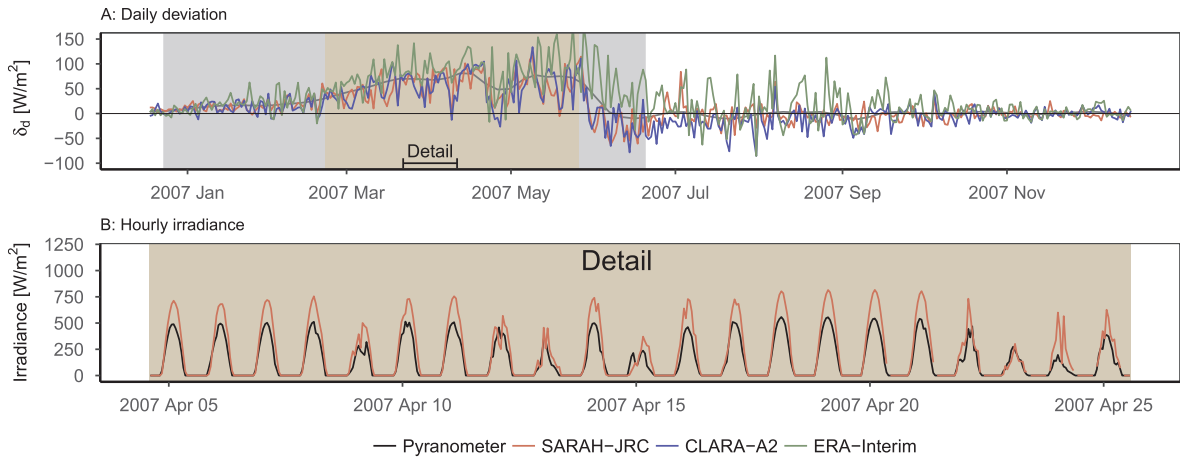

Fig. 5. Operational errors: soiling (UK-1352 - Met Office). The orange area depicts the run looking for large short-lived biases (w = 20 days, n = 2.4), while the gray area shows the run looking for small permanent biases (w = 90 days, n = 0.4). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

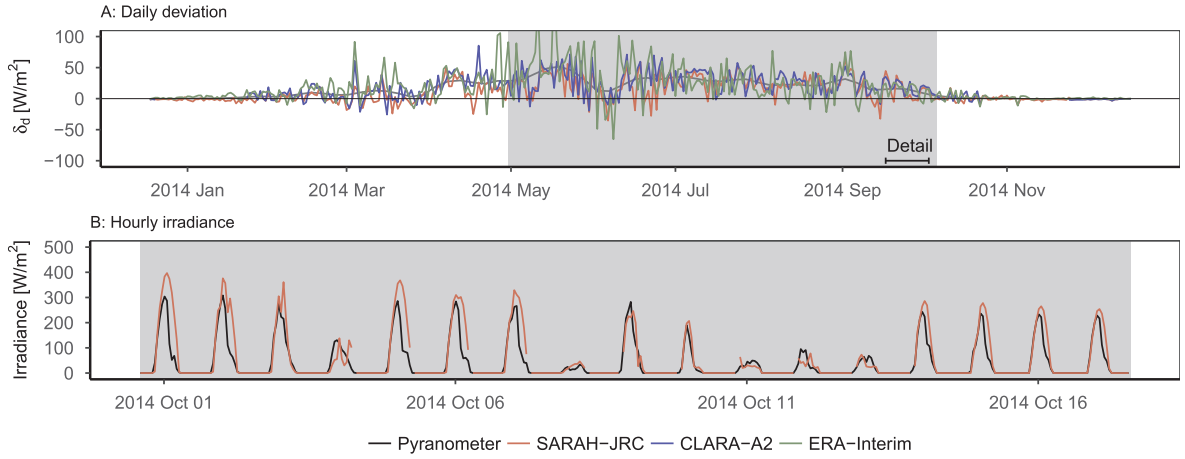

Fig. 6. Operational error: shading (Mære - LMT). The orange area depicts the run looking for large short-lived biases ($w = 20$ days, $n = 2.4$). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

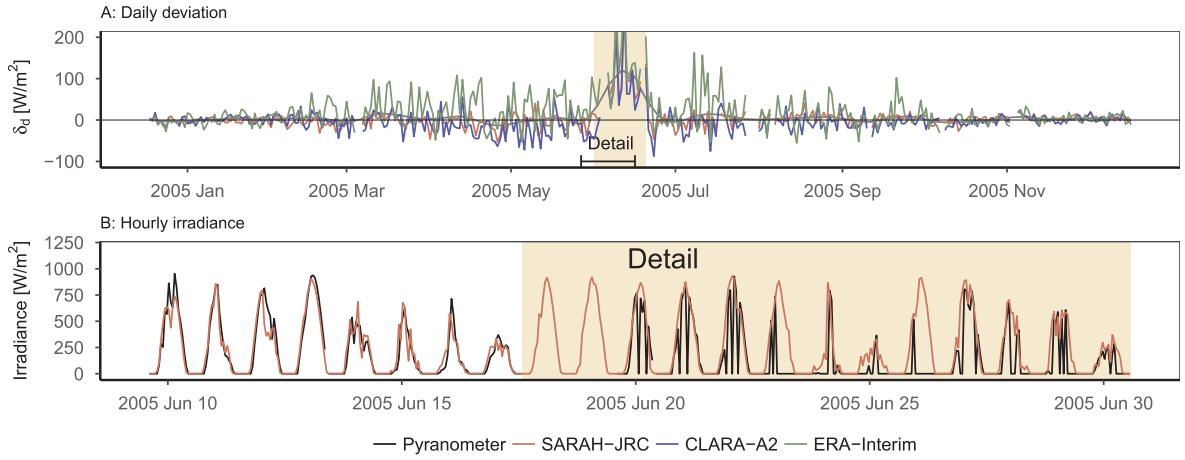

Fig. 7. Operational error: hourly slots set to zero (UK 1302 - Met Office). The orange area depicts the run looking for large short-lived biases ($w = 20$ days, $n = 2.4$). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

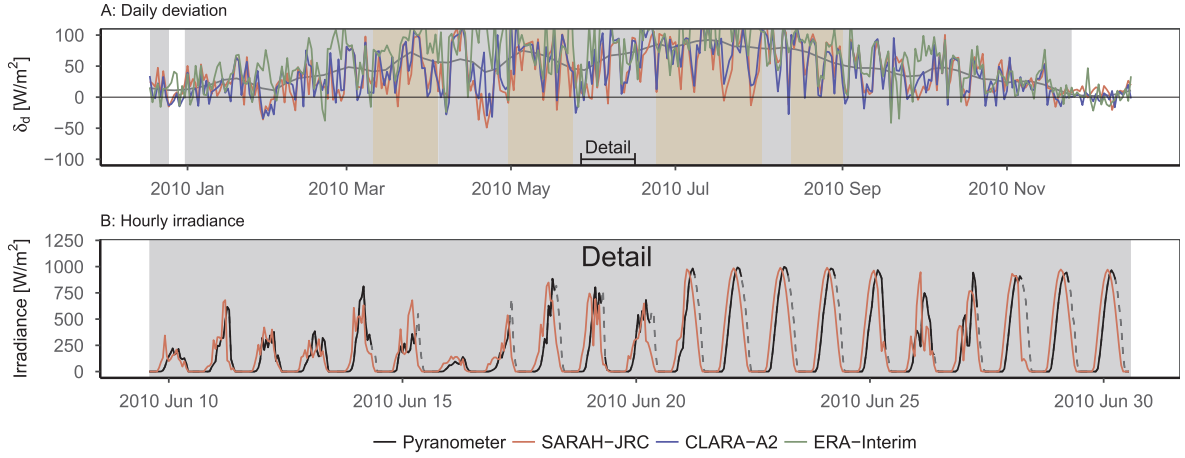

Fig. 8. Operational error: time shift (BU102 - SIAR). The orange area depicts the run looking for large short-lived biases ($w = 20$ days, $n = 2.4$), while the gray area shows the run looking for small permanent biases ($w = 90$ days, $n = 0.4$). The gray line is a smoothed version of the bias of the three products. The dashed line depicts the values removed with the BSRN range tests. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in hourly irradiance plot (Fig. 7B). The defect passes most QC range test with the lower limit for G set to 0 (Long and Dutton, 2002). In the case of tests that work with daily resolution, the effect gets masked during the aggregation process as it also occurred with the shading effect.

Time shifts is another defect that may be related to a failure of the data loggers, although other causes such as a misinterpretation of the time system (local, UTC or solar) could be behind this issue. This error can be partly detected by hourly QC range tests, which will flag either morning (forward time shift) or afternoon slots (backwards time shift) because the records obtained are outside the physical limits. Tests based

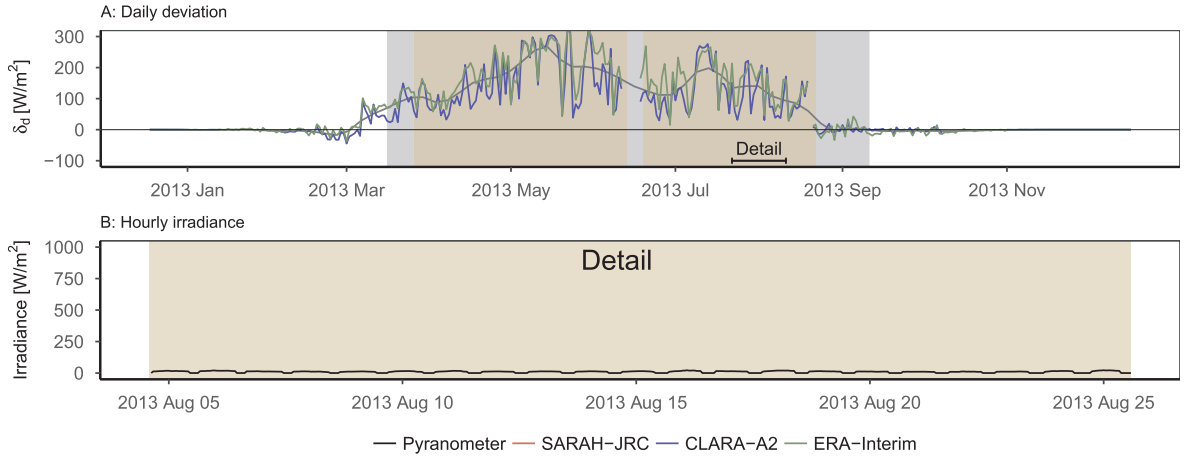

Fig. 9. Operational error: large error (Sortland - LMT). The orange area depicts the run looking for large short-lived biases (w = 20 days, n = 2.4), while the gray area shows the run looking for small permanent biases (w = 90 days, n = 0.4). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

on the symmetry of the irradiance profiles have been also tailored to detect this error (Ineichen, 2013). However, this defect gets completely masked after the integration from hourly to daily. In the present study, our QC is able to detect time shift because the BSRN range test have been run before running the QC method. For instance, afternoon records were removed with the BSRN tests in the station shown in Fig. 8. This results in an underestimation of the daily mean irradiance (Fig. 8A) and it enables the detection of the failure with the proposed QC method. Despite time shifts are not an issue when working with daily values, the combination of the BSRN tests and the proposed QC method allows to detect this type of common error in regional and agricultural networks when sub-daily values are provided.

The last type of operational error found are large errors, which can be caused by malfunctions on the sensor and errors in the data processing stage. These problems typically generate records out of the physical possible limits and they could be detected with some of the QC tests described in Section 1. However, values that are within the acceptance ranges may be sometimes obtained. This is the case of the station shown in Fig. 9 whose records were around 2–3 W/m^2 during the whole year and had passed the BSRN range tests. The error was easily detected with a strong positive overestimation in the daily deviation plot (Fig. 9A).

Equipment errors due to a low-quality sensor or a miscalibration are detected in the daily deviation plot as periods with a low permanent bias that can be either positive or negative. This defect is more evident if the deviation starts or ends within the limits of the period under study. For example, Fig. 10A shows a station with a constant negative deviation from 2005 to 2008 that disappears in 2009. The station metadata reveals that the same sensor was used from 2005 to 2015, so the extinction of the bias was probably due to a re-calibration of the sensor. Equipment errors are rarely detected by traditional QC methods due to the small deviation of the records from the actual irradiance profile. The key point in the detection of this type of errors is the configuration used in the second run of the window function. The analysis of 90 days at a time with the window function enabled the use of narrower acceptance limits, which results in the detection of these small deviations. In addition, the importance of spatial averaging (grouping stations) in the computation of the CIs is evident, because these type of permanent errors wouldn't been detected if the CIs had been calculated with data from a unique station.

Overall, the diversity of the errors identified proves the ability of the proposed QC method to identify faulty records with small deviations from the real irradiance profile. These errors are rarely detected by traditional QC range tests because the values are possible from a physical and a statistical perspective. The best evidence is that all samples from the dataset had previously passed the BSRN range tests (Long and Dutton, 2002) and only a portion of the day in the case of time shifts were flagged by this procedure. Besides, the above description of the different errors demonstrates the benefits of the graphical analysis tool, in opposition to the most used QC methods (Long and Shi, 2008; Hoyer-Klick et al., 2008; NREL, 1993) that just generate a numerical flag and leave to the user the interpretation of the data. The two plots ease the identification of false alarms, making the inspection phase quasi-automatic, quicker and providing more information to identify the most probable source of the error.

5.2. Influence of faulty records on the validation of satellite-based products

The QC method uses radiation products as reference data to detect errors in ground records. However, the process is generally the other way around. High quality ground data are used to validate the newest versions of radiation products. If the ground dataset contains faulty records, the results obtained with these values can be mistaken for actual defects on the radiation products preventing an accurate validation of the product.

Fig. 11 analyzes the influence of the faulty records found with the QC method in the validation of one of the satellite products, SARAH-JRC. The statistics are computed on a yearly basis in order to prevent short periods of faulty data from getting masked in a long time series of 11 years. Results showed a high influence of the periods with errors in the validation metrics. A significant reduction up to 25 W/m^2 in the Bias and MAE was generally observed between the years containing errors and years with acceptable data. The presence of these faulty records was even able to change the sign of the Bias, as it occurred in stations SPA-234 and FRA-265. This fact demonstrates that using raw data, or even data that has passed some of the basic tests (BSRN range tests), may still lead to an incorrect interpretation of the quality of the product analyzed. Moreover, Fig. 11 demonstrates that short periods with errors no longer than 20 days had a notable influence on the yearly aggregated metrics. For instance, differences up to 20 W/m^2 and 15 W/m^2 for the yearly Bias and MAE, respectively, were observed in stations in which just 18 days were removed.

5.3. Limitations of the QC method

The detection of faulty records based on comparing ground records against estimations from radiation products presents some critical aspects. First, the estimations of all products must be stable in time. In this case, the three radiation products used have been designed for climate monitoring purposes. Both satellite products included are CDRs

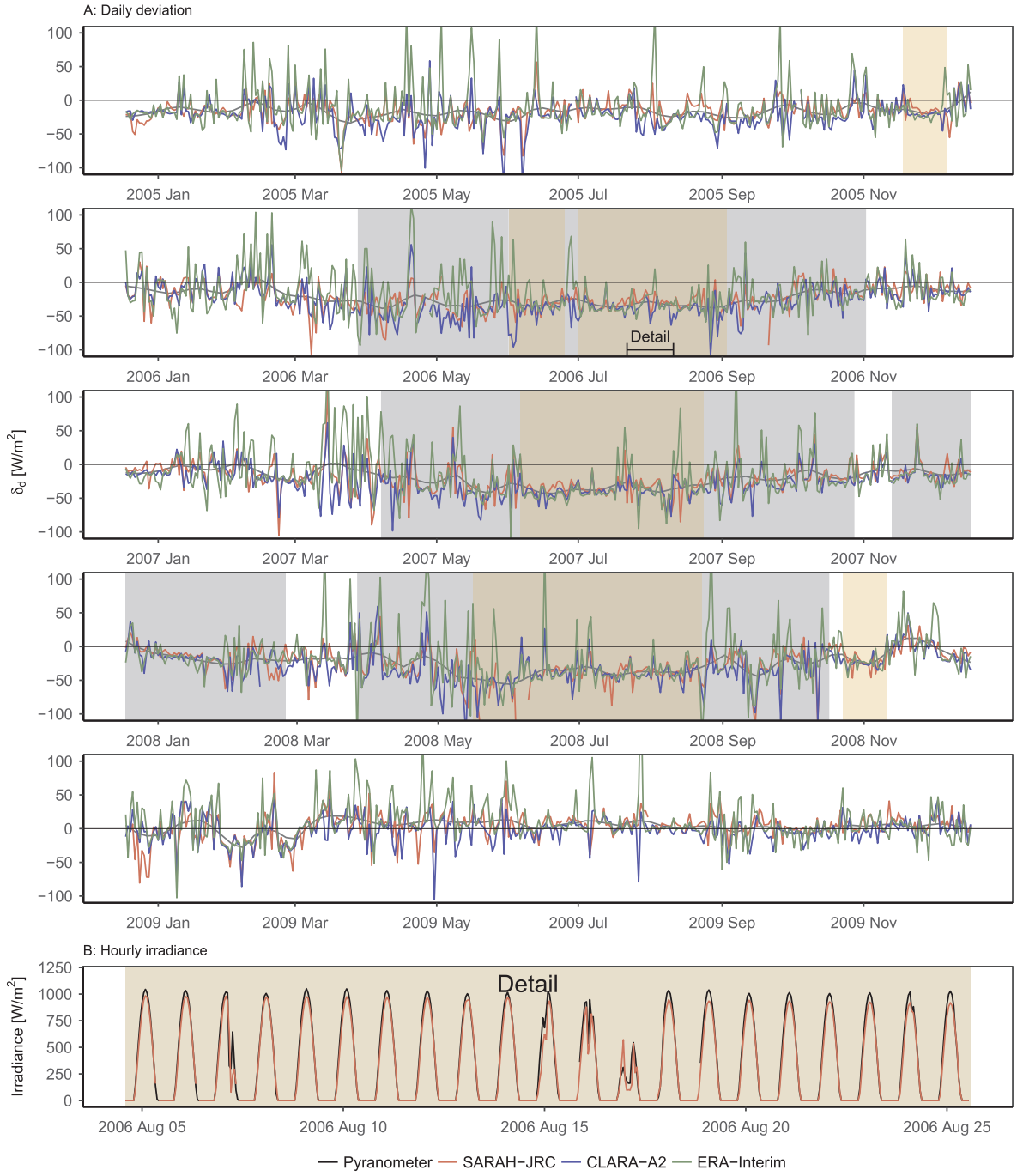

Fig. 10. Equipment error: re-calibration of a sensor (J01 - SIAR). The orange area depicts the run looking for large short-lived biases (w = 20 days, n = 2.4), while the gray area shows the run looking for small permanent biases (w = 90 days, n = 0.4). The gray line is a smoothed version of the bias of the three products. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

normally used for trend analysis, although this does not fully guarantee the stability of the records. CDRs are obtained by combining images from different satellite generations, in which even the type of sensor is sometimes replaced, resulting in a potential introduction of artificial trends in the times series. Besides, satellite instruments degrade in time due to the aging of the components, so the calibration constants have to be constantly corrected to mitigate these aging effects (Decoster et al., 2014).

Nevertheless, the magnitude of these temporal instabilities is generally within the uncertainty of pyranometers and considerably smaller than the deviations caused by the equipment and operational errors, so they do not interfere in the QC method. In the case of SARAH-1, the decadal gradient is −1.1 W/m^2/decade (Müller and Träger-Chatterjee, 2014), while for CLARA-A2, a good agreement was observed between the decadal trend of CLARA-A2 and the one from a set of BSRN stations (Karlsson et al., in press). In respect to ERA-Interim reanalysis, the same NWP model has been used to derive the whole dataset and it includes a bias correction scheme for long-term drifts and calibration errors (Beyer et al., 2011). In addition, the analysis of the inter-annual stability of the three products in the period studied (2005–2015) showed an acceptable temporal stability of the three products (data not shown), and none of the false alarms identified in the 313 ground stations could be attributed to an artificial degradation of the products.

Another critical point is the minimization of the variance of the

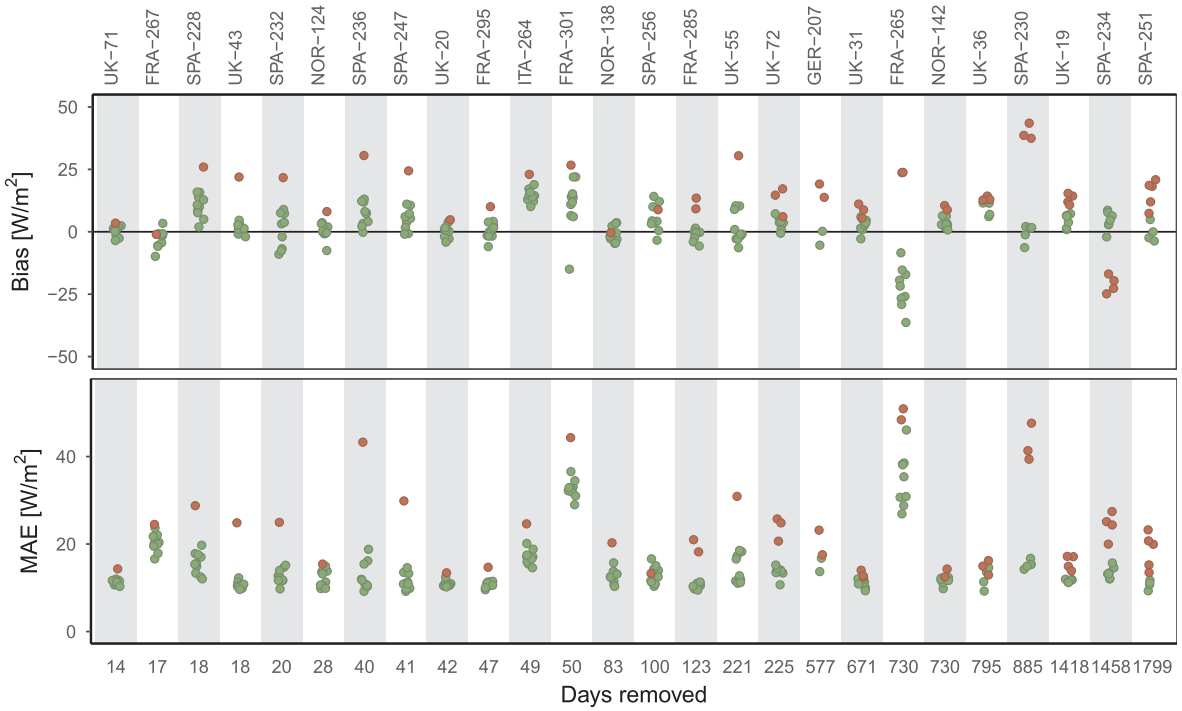

Fig. 11. Yearly aggregated Bias and MAE in W/m^2 of SARAH-JRC in the stations where periods with problematic data were found. Red dots depict the years with faulty records, while green dots represent the years without problems. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimations. One product can overestimate or underestimate solar radiation as long as this is done consistently, because systematic errors on the products are accounted by the CIs of historical values. However, the variance of the estimations defines the width of the CIs, and hence it limits the magnitude of the errors detected by the QC method. These random errors can be due to the limited quality of the products themselves (low spatial and temporal resolution, low-quality ancilliary products, simple atmospheric models) and to the particular conditions of the region analyzed (mountains, snow covered regions, coastlines, etc.) (Súri and Cebecauer, 2014). For instance, the high variance of the estimations at Nordic countries results in wider CIs than the ones in South Europe. This implies that small errors such as shades during sunrise or sunset are more easily detected in South Europe than in the Nordic region. This problem is mitigated in the calculation of the CIs on a monthly basis (temporal averaging) and grouping stations that share similar characteristics (spatial averaging). This spatio-temporal averaging along with the use of robust statistics makes the CIs more restrictive enabling a better detection of small errors.

False alarms may be generated due to the design of the QC procedure. Our method flags days with large daily deviations compared to the historical values of the bias, but deviations observed could result from a failure on both ground or gridded datasets. The use of three independent products derived from different models increases the likelihood that large deviations observed are caused by errors in the recording process. Combining different products (instead of using one single product) also enhances the detection capacity of the QC method (see Supplementary Material 1). However, all radiation products may still fail at the same time. This is especially alarming when the failure is not consistent in space and time. A side effect of computing the CIs in a monthly basis (temporal averaging) for groups of stations (spatial averaging) is that inconsistent failures of the products will be also flagged by the QC method. This is the case of the presence of one month with snow cover in a location where snow rarely falls (temporal inconsistency), or the case of a few high-elevation stations grouped with the rest of stations in a predominantly flat country (spatial inconsistency). Despite this, false alarms due to inconsistent deviations were observed only in 3 of the 313 ground stations (Table 2), all of them high-elevation stations or locations with seasonal snow. Besides, the number of false alarms decreased after combining the three independent products (see Supplementary Material 1), and those alarms still prevailing could be rapidly detected in the two plots generated for the visual inspection phase.

6. Conclusions

This paper presents a new QC method to detect faulty records on ground stations. The method exploits the advances in the development of solar radiation products and it flags periods of consecutive days where the daily deviations of several independent radiation products is over or under the historical bias for the products in that location and period of the year. The QC method also includes a graphical analysis tool to analyze the flagged records, helping users in the decision to keep or remove them from the dataset.

The QC method was tested on a large database comprised by 313 ground stations, and it detected several operational errors (28 stations) such as snow or soiling over the sensor, shading, time shifts or large errors in the data logger and in the sensor, and equipment errors (3 stations) such as a miscalibration and low quality sensors. All these faulty samples had already passed the BSRN range tests, proving that our QC method can identify slight deviations not detectable with some of the most common QC methods. Besides, we included a graphical analysis tool based on the automatic generation of two plots that facilitates the detection of false alarms and the identification of the most probable cause of the error. Finally, we quantified the influence of these errors in the validation on a satellite-based product. The presence of faulty periods (around one month) enlarged the yearly Bias and MAE up to 20 W/m^2 and in some cases even changed the sign of the bias, highlighting the importance of using quality controlled data for the correct validation of radiation products.

The method we have presented depends on the availability of independently derived time series of solar radiation data. The data used here are available for free from the respective data providers:

ERA-interim European Center for Medium-Range Weather Forecast (ECMWF) www.ecmwf.int.
CLARA and SARAH Climate Monitoring Satellite Application Facility (CM-SAF) www.cmsaf.eu.

CLARA and ERA-interim have global coverage while SARAH covers Europe, Africa, and most of Asia. If the QC procedure was to be applied to ground measurements from America, data for most of the Americas are available from the National Solar Radiation Database (NSRDB) (NREL, 2016). NSDRB data contains its own biases but as long as the biases are consistent in time it shouldn't prevent its use in the QC method. These data sets are generally organized as spatial maps for single times or days, so the data handling may be somewhat cumbersome. NSRDB data are also available as time series for single locations. In the near future, hourly time series from SARAH will also be made available via the PVGIS web application.

Acknowledgements

The authors would like to thank the different national networks for providing their high-quality data free of charge. We specially appreciate the collaboration of Sandra Anderson from the SHMI, Halvard Loe from NIBIO, Virginie Gorjux from Meteo France. We also appreciate the help from Elena Koumpli and Diana Palmer in accessing to the CEDA catalogue. Besides, we thank the DWD, SIAR and BSRN for their on-line open services. We would also like to acknowledge the CM SAF project for providing the two satellite-based datasets, and the ECMWF for the ERA-Interim database.

R. Urraca is funded by the fellowship FPI-UR-2014 granted by the University of La Rioja. A.V. Lindfors and A. Riihelä were supported by the Academy of Finland, decision 284536. A. Sanz-Garcia is funded by Academy of Finland, decision 273689 and wishes to acknowledge CSC - IT Center for Science, Finland, for computational resources. This work used the Beronia cluster (Universidad de La Rioja), which is supported by FEDERMINECO Grant No. UNLR-094E-2C-225.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.solener.2017.09.032.

References

Beyer, H., Costanzo, C., Heinemann, D., 2011. Evaluation and assessment of the CM-SAF surface solar radiation climate data records. EUMETSAT Satelite Application Facility on Climate Monitoring, CDOP AS Study No 18 < http://www.cmsaf.eu/EN/Documentation/Reports/2007-now/CLM_AS10_P05.pdf?_blob=publicationFile&v=3 > .

Bird, R.E., Hulstrom, R.L., 1980. Direct Insolation Models. Solar Energy Research Institute (Now NREL), Golden, CO, SERI/TR-335-344 < http://www.nrel.gov/docs/legosti/old/344.pdf > .

Bird, R.E., Hulstrom, R.L., 1981. A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces. Solar Energy Research Institute (now NREL), Golden, CO, SERI/TR-642-761 < http://www.nrel.gov/docs/legosti/old/761.pdf > .

Bivand, R., Keitt, T., Rowlingson, B., 2016. RGDAL: Bindings for the Geospatial Data Abstraction Library < https://CRAN.R-project.org/package=rgdal > . R package version 1.1-10.

Boilley, A., Wald, L., 2015. Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface. Renew. Energy 75, 135–143. http://dx.doi.org/10.1016/j.renene.2014.09.042.

Bojanowski, J.S., Vrieling, A., Skidmore, A.K., 2014. A comparison of data sources for creating a long-term time series of daily gridded solar radiation for Europe. Solar Energy 99, 152–171. http://dx.doi.org/10.1016/j.solener.2013.11.007.

BSRN, 2016. Baseline Surface Radiation Network < http://bsrn.awi.de > (accessed 10.10.16).

Decoster, I., Clerbaux, N., Baudrez, E., Dewitte, S., Ipe, A., Nevens, S., Velazquez Blazquez, A., Cornelis, J., 2014. Spectral aging model applied to Meteosat First Generation visible band. Rem. Sens. 6, 2534–2571. http://dx.doi.org/10.3390/rs6032534.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Quart. J. R. Meteorol. Soc. 137, 553–597. http://dx.doi.org/10.1002/qj.828.

DWD, 2016. Deutscher Wetterdienst < http://www.dwd.de/DE/Home/home_node.html > (accessed 10.10.16).

FMI, 2016. Finish Meteorological Institute < http://en.ilmatieteenlaitos.fi > (accessed 10.10.16).

Geiger, M., Diabatè, L., Mènard, L., Wald, L., 2002. Controlling the quality of solar irradiation data by means of a web service.

Grolemund, G., Wickham, H., 2011. Dates and times made easy with lubridate. J. Statist. Softw. 40, 1–25.. <http://www.jstatsoft.org/v40/i03/> .

Gueymard, C., Ruiz-Arias, J., 2015. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. Solar Energy 128, 1–30. http://dx.doi.org/10.1016/j.solener.2015.10.010.

Hijmans, R.J., 2015. RASTER: Geographic Data Analysis and Modeling < http://CRAN.R-project.org/package=raster > . r package version 2.4-20.

Hoyer-Klick, C., Beyer, H.G., DUmortier, D., Schroedter-Homscheidt, M., Wald, L., Martinoli, M., Schilings, C., Gschwind, B., Menard, L., Gaboardi, E., Ramirez-Santigosa, L., Polo, J., Cebecauer, T., Huld, T., Súri, M., de Blas, M., Lorenz, E., Pfatischer, R., Remund, J., Ineichen, P., Tsvetkov, A., Hofierka, J., 2008. MESOR. Management and exploitation of solar resource knowledge. In: Proceeding of the EUROSUN 2008, 1st International Conference on Solar Heating, Cooling and Buildings, Lisbon, Portugal, October 2008, pp. 7–10.

Ineichen, P., 2013. Solar Radiation Resource in Geneva: Measurements, Modeling, Data Quality Control, Format and Accessibility, pp. 333.7–333.9 < http://archive-ouverte.unige.ch/unige:29599 > (iD: unige:29599).

ISO, 1990. ISO 9060:1990: Specification and Classification of Instruments for Measuring Hemispherical Solar and Direct Solar Radiation. Geneva, Switzerland.

Jones, P.D., Harpham, C., Troccoli, A., Gschwind, B., Ranchin, T., Wald, L., Goodess, C. M., Dorling, S., 2017. Using ERA-Interim Reanalysis output for creating datasets of energy- relevant climate variables. Earth Syst. Sci. Data Discuss., 1–31. doi:http://dx.doi.org/10.5194/essd-2016-67 (in press).

Journée, M., Bertrand, C., 2011. Quality control of solar radiation data within the RMIB solar measurements network. Solar Energy 85, 72–86. http://dx.doi.org/10.1016/j.solener.2010.10.021.

JRC, 2016. Joint Research Centre, Ispra, ESTI Meteo Tower < http://iamest.jrc.it/meteo/meteo.php > (accessed 10.10.16).

Karlsson, K., Anttila, K., Trentmann, J., Stengel, M., Merinik, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G., Schludnt, C., Stein, D., Finkensieper, S., Håkansson, N., Hollmann, R., 2016. CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. Atmos. Chem. Phys. Disc., 1–41, doi:http://dx.doi.org/10.5194/acp-2016-935 (in press).

Leys, C., Ley, C., Klein, O., Bernard, P., Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. 49, 764–766. http://dx.doi.org/10.1016/j.jesp.2013.03.013.

Long, C.N., Dutton, E.G., 2002. BSRN Global Network Recommended QC Tests, V2.0. BSRN Technical Report < http://ezksun3.ethz.ch/bsrn/admin/dokus/qualitycheck.pdf > .

Long, C.N., Shi, Y., 2008. An automated quality assessment and control algorithm for surface radiation measurements. Open Atmos. Sci. J. 2, 23–27. http://dx.doi.org/10.2174/1874282300802010023.

McArthur, L.J.B., 2005. World Climate Research Programme - Baseline Surface Radiation Network (BSRN) - Operations Manual Version 2.1. Experimental Studies Division, Atmospheric Environment Service, Downsview, Ontario, Canada.

Met Office, 2017. MIDAS: Global Radiation Observations. NCAS British Atmospheric Data Centre < http://catalogue.ceda.ac.uk/uuid/b4c028814a666a651f52f2b37a97c7c7 > (accessed 25.04.17).

Meteo France, 2016. Meteo France < http://www.meteofrance.fr > (accessed 10.10.16).

Molineaux, B., Ineichen, P., 2003. Automatic Quality Control of Daylight Measurement: Software for IDMP Stations, Technical Report. Vaulx en Velin, France: International Daylight Measurement Programme, École National des Travaux Publics < http://idmp.ente.fr > .

Moradi, I., 2009. Quality control of global solar radiation using sunshine hours. Energy 34, 1–6. http://dx.doi.org/10.1016/j.energy.2008.09.006.

Moreno-Tejada, S., Ramírez-Santigosa, L., Silva-Pérez, M.A., 2015. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. Renew. Energy 78, 531–537. http://dx.doi.org/10.1016/j.renene.2015.01.031.

Müller, R., Träger-Chatterjee, C., 2014. Brief accuracy assessment of aerosol climatologies for the retrieval of solar surface radiation. Atmosphere 5, 959–972. http://dx.doi.org/10.3390/atmos5040959.

Muller, R., Uwe, P., Trager-Chatterjee, C., Cremer, R., Trentmann, J., Hollmann, R., 2015. Surface Solar Radiation Data Set - Heliosat (SARAH) - Edition 1. Satellite Application Facility on Climate Monitoring (CM SAF), doi:http://dx.doi.org/10.5676/EUM_SAF_CM/SARAH/V001.

Muneer, T., Fairooz, F., 2002. Quality control of solar radiation and sunshine measurements - lessons learnt from processing worldwide databases. Build. Serv. Eng. Res. Technol. 23, 151–166. http://dx.doi.org/10.1191/0143624402bt038oa.

NIBIO-LMT, 2016. Norwegian Institute of Bioceonomy Research) - Landbruksmeteorologisk Tjeneste < http://lmt.nibio.no/ > (accessed 10.10.16).

NREL, 1993. Users Manual for SERI QC Software - Assessing the Quality of Solar Radiation Data, Technical Report. National Renewable Energy Laboratory (NREL), Golden, CO < http://www.nrel.gov/docs/legosti/old/5608.pdf > .

NREL, 2016. National Solar Radiation Database (NSRDB) < https://nsrdb.nrel.gov > .
Ohmura, A., Dutton, E.G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., Heimo, A., König-Langlo, G., McArthur, B., Müller, G., Philipona, R., Pinker, R., Whitlock, C.H., Dehne, K., Wild, M., 1998. Baseline Surface Radiation Network (BSRN/WCRPP): new precision radiometry for climate research. Bull. Am. Meteorol. Soc. 79, 2115–2136. http://dx.doi.org/10.1175/1520-0477(1998)079%3C2115:BSRNBW%3E2.0.CO;2.
Page, J., Lebens, R., 1986. Climate in the United Kingdom: A Handbook of Solar Radiation, Temperature and Other Data for Thirteen Principal Cities and Towns. HMSO, London.
Pashiardis, S., Kalogirou, S.A., 2016. Quality control of solar shortwave and terrestrial longwave radiation for surface radiation measurements at two sites in Cyprus. Renew. Energy 96, 1015–1033. http://dx.doi.org/10.1016/j.renene.2016.04.001.
Pebesma, E.J., Bivand, R.S., 2005. Classes and Methods for Spatial Data in R. R News 5 < http://cran.r-project.org/doc/Rnews/ > .
Perez, R., Ineichen, P., Seals, R., Zelenka, A., 1990. Making full use of the clearness index for parameterizing hourly insolation conditions. Solar Energy 45, 111–114. http://dx.doi.org/10.1016/0038-092X(90)90036-C.
Perpiñán, O., 2012. Solar radiation and photovoltaic systems with R. J. Statist. Softw. 50, 1–32.. <https://cran.r-project.org/web/packages/solaR/index.html> .
Polo, J., Wilbert, S., Ruiz-Arias, J.A., Meyer, R., Gueymard, C., Súri, M., Martín, L., Mieslinger, T., Blanc, P., Grant, I., Boland, J., Ineichen, P., Remund, J., Escobar, R., Troccoli, A., Sengupta, M., Nielsen, K.P., Renne, D., Geuder, N., Cebecauer, T., 2016. Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets. Solar Energy 132, 25–37. http://dx.doi.org/10.1016/j.solener.2016.03.001.
PVGIS, 2016. Photovoltaic Geographical Information System < http://re.jrc.ec.europa.eu/pvgis/ > (accessed 20.10.16).
R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria < http://www.R-project.org/ > .
Rigollier, C., Bauer, O., Wald, L., 2000. On the clear sky model of the ESRA - European Solar Radiation Atlas with respect to the Heliosat method. Solar Energy 68, 33–48. http://dx.doi.org/10.1016/S0038-092X(99)00055-9.
Schroedter-Homscheidt, M., Hoyer-Klick, C., Killius, N., Lefèvre, M., Wald, L., Wey, E., Saboret, L., 2016. User's Guide to the CAMS Radiation Service. Status December 2016 < https://atmosphere.copernicus.eu/sites/default/files/FileRepository/Resources/Documentation/Radiation/CAMS72_2015SC1_D72.11.3.1_201612_UserGuide_v2.pdf > .
Sengupta, M., Habte, A., Kurtz, S., Dobos, A., Wilbert, S., Lorenz, E., Stoffel, T., Renné, D., Gueymard, C., Myers, D., Wilcox, S., Blanc, P., Perez, R., 2015. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications, NREL Technical Report < http://www.nrel.gov/docs/fy15osti/63112.pdf > .
SIAR, 2015. Servicio de Información Agroclimática para el Regadío < http://eportal.magrama.gob.es/websiar/Inicio.aspx > (accessed 10.10.16).
SMHI, 2016. Swedish Meteorological and Hydrological Institute < http://www.smhi.se/en > (accessed 10.10.16).
Súri, M., Cebecauer, T., 2014. Satellite-based solar resource data: model validation statistics versus user's uncertainty. In: ASES SOLAR 2014 Conference, San Francisco, 7–9 July 2014.
Tang, W., Yung, K., He, J., Qin, J., 2010. Quality control and estimation of global solar radiation in China. Solar Energy 84, 466–475. http://dx.doi.org/10.1016/j.solener.2010.01.006.
Urraca, R., Martinez-de Pison, E., Sanz-Garcia, A., Antonanzas, J., Antonanzas-Torres, F., 2016. Estimation methods for global solar radiation: case study evaluation of different approaches in central Spain. Renew. Sustain. Energy Rev., doi:http://dx.doi.org/10.1016/j.rser.2016.11.222 (in press).
Wickham, H., 2016. tidyverse: Easily Install and Load 'Tidyverse' Packages < https://CRAN.R-project.org/package=tidyverse > . R package version 1.0.0.
WMO, 2008. WMO Guide to Meteorological Instruments and Methods of Observation: WMO No. 8. 7th ed. Geneva, Switzerland < http://www.wmo.int/pages/prog/gcos/documents/gruanmanuals/CIMO/CIMO_Guide-7th_Edition-2008.pdf > .
Younes, S., Claywell, R., Muneer, T., 2005. Quality control of solar radiation data: present status and proposed new approaches. Energy 30, 1533–1549. http://dx.doi.org/10.1016/j.energy.2004.04.031.
Zahumenský, I., 2004. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. World Meteorological Organization. Geneva, Switzerland < https://www.wmo.int/pages/prog/www/OSY/Meetings/ET-AWS3/Doc4(1).pdf > .