Available online at www.sciencedirect.com
ScienceDirect
Solar Energy 85 (2011) 72–86
SOLAR ENERGY
www.elsevier.com/locate/solener

Quality control of solar radiation data within the RMIB solar measurements network

Michel Journée *, Cédric Bertrand

Royal Meteorological Institute of Belgium, Avenue Circulaire 3, B-1180 Brussels, Belgium

Received 20 April 2010; received in revised form 21 September 2010; accepted 30 October 2010
Available online 30 November 2010

Communicated by: Associate Editor Frank Vignola

Abstract

Assessment of the solar resource is based upon measured data, where available. However, with any measurement there exist errors. Consequently, solar radiation data do not exhibit necessarily the same reliability and it often happens that users face time series of measurements containing questionable values though preliminary technical control has been done before the data release. To overcome such a situation, a major effort has been undertaken at the Royal Meteorological Institute of Belgium (RMIB) to develop procedures and software for performing post-measurement quality control of solar data from the radiometric stations of our in situ solar monitoring network. Moreover, because solar energy applications usually need continuous time series of solar radiation data, additional procedures have also been established to fill missing values (data initially lacking or removed via quality checks).

Keywords: Global, direct and diffused solar radiation; Sunshine duration; Quality control; Empirical models

1. Introduction

With the development of solar-based renewable energy technologies national meteorological services are faced with increasing demands for reliable solar resource data. Up-to-date series of these data are also being sought by architects and building services professionals for better design of buildings. Compared to measurements of other meteorological variables, the measurement of solar radiation is more prone to errors (Moradi, 2009). Younes et al. (2005) identified two major categories of possible sources of problems or errors related to in situ measurement of solar radiation: (1) equipment error and uncertainty and, (2) operation related problems and errors; with the most common sources of error arising from the sensors and their constructions (e.g., Muneer and Fairooz, 2002; Younes et al., 2005). Because of the difficulties frequently encountered when measuring solar radiation and the resultant unknown quality of some solar radiation data, it is crucial to perform a quality assessment of these data prior to their further processing (e.g., Hay, 1993). As an example Colle et al. (2001) have shown that uncertainty in life cycle savings for solar thermal and photovoltaic systems are linearly correlated with uncertainty in solar resource data.

Because quality control may be a lengthy and tedious task, most of the customers and scientists are ready to use data from meteorological offices in confidence without performing an additional precise and fine control. A major effort has therefore been recently undertaken at the Royal Meteorological Institute of Belgium (RMIB) to develop quasi-automated procedures and software for performing post-measurement quality control of solar data in addition to the human data monitoring.

The usual solar radiation parameters measured on ground are the global solar irradiance, the direct solar irradiance and the sunshine duration. We are currently measuring various combinations of these parameters in 14 automatic weather stations (AWS) in addition to the measurements performed in our main/reference station in

* Corresponding author. Tel.: +32 2 3730598.
E-mail address: Michel.Journee@oma.be (M. Journée).

0038-092X/$ - see front matter © 2010 Published by Elsevier Ltd.
doi:10.1016/j.solener.2010.10.021

Uccle (see Table 1 and Fig. 1). Since a few years, solar radiation data are integrated to bring them to a 10 min time step (local mean time or clock time) in addition to the historical 30 min (solar time) time step.

Based on previously proposed procedures for quality assessment of solar irradiation data (e.g., Maxwell et al., 1993; Molineaux and Ineichen, 1994; Terzenbach, 1995; Geiger et al., 2002; Muneer and Fairooz, 2002; Younes et al., 2005; Shi et al., 2008; Moradi, 2009; Tang et al., 2010), we present in this paper a new quality control scheme developed for the solar data measured within the RMIB solar radiation measurements network. In contrast to previously published quality control algorithms usually devoted to hourly or daily solar radiation data, the new scheme is intended for sub-hourly data (i.e., 10 min and 30 min averaged data). Since for solar energy applications there is a need for continuous time series of solar radiation data to correctly assess the usefulness of the particular application and in its implementation, additional procedures have also been developed to provide users with an estimation of missing solar values (i.e. data initially lacking or failing the quality assessment criteria).

The paper is organized as follows: the new quality control scheme is presented in the next section. Section 3 deals with procedures to fill missing values in the data time series and their validation. Conclusions are given in Section 4. The main notations used in this paper are explained in Table 2.

2. Quality control of solar radiation and sunshine measurements

The quality control of radiometric observations involves various tests to establish boundaries or limits within which acceptable data are expected to lie. These tests are guided either by physical reasoning (to detect physically impossible events) or by the statistical variability of the data (to detect very rare and thus questionable events). Furthermore, they either treat the various solar radiation parameters separately or compare them to each other. Temporal dependence in the data can eventually be taken into account. Finally, data from multiple sites can be used to investigate spatial dependence. The proposed scheme for quality control of radiometric data (see Fig. 2) involves several categories of tests that are detailed in the forthcoming sections and summed up in Table 3. While most of these tests can be used in a fully automated manner, some aiming at detecting specific error types (e.g., misleading calibrations of the instruments as well as the impacts of shadow and snow) require a decision taken by a human operator. Depending on the results of these tests, flags with attribute either "valid", "suspicious" or "erroneous" are set to inform the user of any departure of the data from expected values. All details on the rules followed to attribute the quality flags are provided in Table 4. Data that fail the tests are not deleted or modified but stored because they may provide valuable insight into the origin of failure.

Because the developed procedures can be applied to our 10 and 30 min averaged solar radiation data, solar irradiance values (in $W\,m^{-2}$) are first converted in terms of solar irradiation data (in $W\,h\,m^{-2}$). Sunshine duration values are expressed in minutes. Finally, it is important to mention that all threshold values used in these tests, except those derived from physical reasonings, were established from the RMIB solar radiation data with the objective to discriminate valid observations from erroneous ones. These values are thus representative of a low-altitude European region and might be adjusted for other regions.

2.1. Physical threshold tests

Physical limits on the solar radiation parameters are derived from the solar radiation at the top of the atmosphere (TOA) and at the Earth's surface for both very clear skies and overcast skies.

Table 1
Location of the ground stations involved in the RMIB solar radiation monitoring network, with the associated availability for the global horizontal radiation, the ground-reflected radiation, the diffuse horizontal radiation, the direct normal radiation and the sunshine duration.

Station code & name		Lat. (°N)	Long. (°E)	Alt. (m)	Global incident & reflected	Diffuse	Direct & duration
6407	Middelkerke	51.198	2.869	3	×		×
6414	Beitem	50.905	3.123	25	×		×
6418	Zeebrugge	51.349	3.196	8			×
6434	Melle	50.976	3.825	15	×		×
6439	St.-Katelijne-Waver	51.076	4.526	10	×		×
6447	Uccle	50.798	4.359	101	×	×	×
6455	Dourbes	50.096	4.596	233	×		×
6459	Ernage	50.583	4.691	157	×		×
6464	Retie	51.222	5.028	21	×		×
6472	Humain	50.194	5.257	296	×		×
6476	Saint-Hubert	50.040	5.405	557	×		×
6477	Diepenbeek	50.916	5.451	39	×		×
6478	Bierset	50.646	5.452	176			×
6484	Buzenol	49.621	5.589	324	×		×
6494	Mont Rigi	50.512	6.075	673	×		×

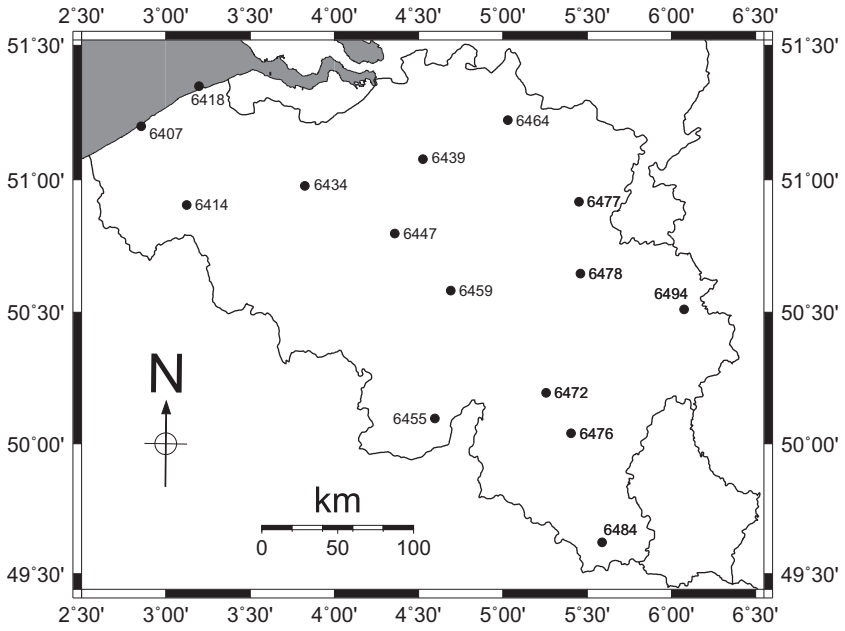

Fig. 1. RMIB solar radiation measurements network. See Table 1 for the stations names and geographical coordinates.

Table 2
Nomenclature.

E	Extraterrestrial solar radiation on a horizontal surface
G	Global solar radiation on a horizontal surface
D	Diffuse solar radiation on a horizontal surface
$B^{\perp}$	Direct beam solar radiation
$B = B^{\perp}\sin(h)$	Direct solar radiation on a horizontal surface
h	Solar elevation angle
SD	Sunshine duration (in minutes)
$SD_{\max}$	Maximal sunshine duration, i.e., either 10 or 30 min
R	Ground-reflected solar radiation
$K_t = G/E$	Clearness index
$K_n = B/E$	Beam transmittance
$K = D/G$	Diffuse ratio

First, the global solar radiation on a horizontal surface at the Earth's surface, G, has to be less than the corresponding extraterrestrial value incident on a horizontal surface, E. It is possible that solar irradiation at the surface is larger than at TOA for short time periods because of the diffusive effects of clouds that are not in the way of the solar beam (Shi et al., 2008). This situation, which can be quite frequent in high altitude regions in case of convective clouds (Yang et al., 2010), is however not expected for lowland areas except when the elevation of the sun above the horizon is small. We hence imposed the condition $G < E$ to all data with a solar elevation angle greater than 2°. Because the global radiation accounts for contributions of both the beam and the diffuse radiation, beam, B, and diffuse, D, radiation incoming on a horizontal surface cannot be larger than E.

Second, because solar radiation at the surface is affected by atmospheric absorbing gases, clouds, and aerosols the measured values cannot exceed the maximum income of solar radiation with clear sky and high atmospheric transparency, e.g., Gueymard (2004). As in Geiger et al. (2002), clear sky solar irradiations are assessed by means of the modified European Solar Radiation Atlas (ESRA) model. This model is based upon a parameterization formula by Kasten (ESRA, 1984; Kasten, 1996) and has been

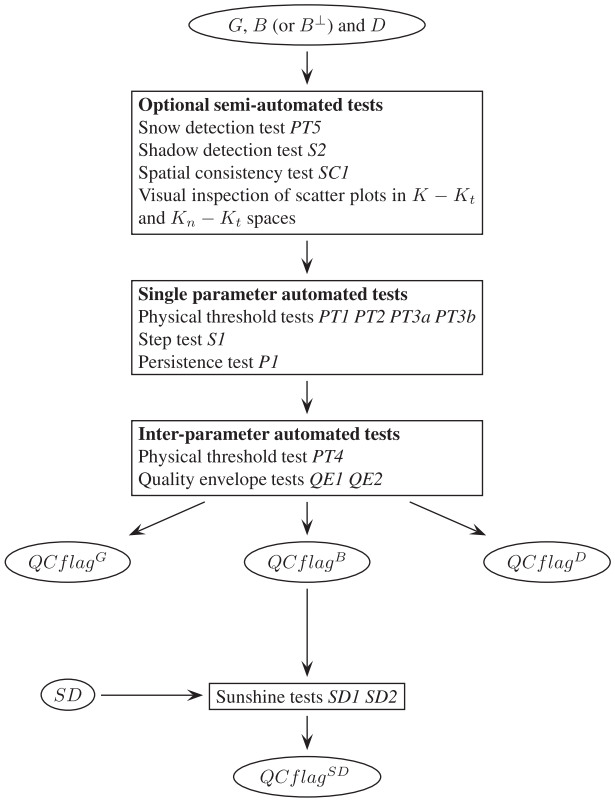

Fig. 2. Flow diagram for the quality control of radiometric data. See Table 3 for the description of the labels used to denote the various tests and Table 4 for the rules followed to assign quality flag attributes (*QC flag*).

Table 3
Quality criteria of radiometric data used in physical threshold tests (PT), step tests (S), persistence tests (P), quality envelope tests (QE), spatial consistency tests (SC) and sunshine tests (SD).

Test label	Test description	Quality criteria
PT1	Comparison of surface solar radiation data against the extraterrestrial solar radiation	$G/E < 1$ if $h > 2°$ $B/E < 1$ $D/E < 1$ if $h > 2°$
PT2	Comparison of surface solar radiation values against the outputs of the modified ERSA model for clear sky	$G/G_{cs} \leqslant 1.1$ if $h > 2°$ $G/G_{cs} \leqslant 2$ if $h \leqslant 2°$ $B/B_{cs} \leqslant 1$ $D/D_{cs} \geqslant 1$
PT3a	Lower bounds on G and B derived from heavily overcast conditions with low atmospheric transparency	$G/E \geqslant 10^{-4}(h-10)$ if $h > 10°$ $G \geqslant 0$ if $h \leqslant 10°$ $B \geqslant 0$
PT3b	Lower bounds on G derived from heavily overcast conditions with low atmospheric transparency. The daily mean μ is derived from all data acquired from sunrise to sunset	$\mu(G/E) \geqslant 0.03$
PT4	Tests derived from the relation $G = B + D$	$B/G \leqslant 0.95$ if $h > 2°$ $B/G < 1$ if $h \leqslant 2°$ $D/G \leqslant 1$
PT5	Test to detect snow-corrupted measurements of the global solar radiation. Special care has to be taken when these two conditions hold simultaneously	$B/G \geqslant 1$ $R/G \geqslant 0.7$
S1	Bounds on the variations of the solar radiation parameters between two successive timestamps	$\left\lvert\frac{G(t)}{E(t)} - \frac{G(t-1)}{E(t-1)}\right\rvert < 0.75$ if $h(t) > 2°$ $\left\lvert\frac{B(t)}{E(t)} - \frac{B(t-1)}{E(t-1)}\right\rvert < 0.65$ if $h(t) > 2°$ $\left\lvert\frac{D(t)}{E(t)} - \frac{D(t-1)}{E(t-1)}\right\rvert < 0.35$ if $h(t) > 2°$
S2	Test to detect shadow obscuring the global sensor. Special care has to be taken when these four conditions hold simultaneously	$h(t) > 2°$ $B(t) \geqslant G(t)$ and $B(t-1) < G(t-1)$ $\left\lvert\frac{G(t)}{E(t)} - \frac{G(t-1)}{E(t-1)}\right\rvert > 0.1$ $\frac{G(t)-G(t-1)}{B(t)-B(t-1)} > 3$ or < -1
P1	Persistence tests. The daily mean μ and standard deviation σ are derived from all data acquired from sunrise to sunset	$\frac{1}{8}\mu\left(\frac{G}{E}\right) \leqslant \sigma\left(\frac{G}{E}\right) \leqslant 0.35$ $\frac{1}{6}\mu\left(\frac{D}{E}\right) \leqslant \sigma\left(\frac{D}{E}\right) \leqslant 0.2$ $\lambda \leqslant \sigma\left(\frac{B}{E}\right) \leqslant 0.3$ where $\lambda = \begin{cases} 0 & \text{if } \mu\left(\frac{B}{E}\right) \leqslant 0.01, \\ \frac{1}{2}\mu\left(\frac{B}{E}\right) & \text{if } 0.01 < \mu\left(\frac{B}{E}\right) \leqslant 0.2, \\ 0.1 + \frac{1}{20}\left(\mu\left(\frac{B}{E}\right) - 0.2\right) & \text{otherwise.} \end{cases}$
QE1	Quality envelope test in the $K_t - K$ space	See Section 2.4 for details
QE2	Quality envelope test in the $K_t - K_n$ space	See Section 2.4 for details
SC1	Comparison of daily totals of global solar radiation against values obtained by spatial interpolation of nearby stations data	See Section 2.5 for details
SD1	Comparison of sun duration values against extreme values	$0 \leqslant SD \leqslant SD_{\max}$
SD2	Compatibility of sunshine duration values with direct solar radiation data	$B^{\perp} < 120$ W m^{-2} if $SD = 0$

detailed in Rigolier et al. (2000). Corrections were brought to this model as proposed by Remund and Page (2002). This clear-sky model makes use of the Linke turbidity factor (a parameter quantifying the atmospheric visibility – aerosol plus water vapor – under clear skies) and has the time, the date, the latitude, the longitude and the altitude of the site as inputs. To represent the clear-conditions, the Linke factor with an atmospheric mass of two is set at a value of one which stands for very optically clean atmosphere (Geiger et al., 2002). Measured values of the three components of surface solar irradiation are compared against the outputs of the model by means of the following thresholds:

- The upper limit for G is set as 1.1 times the calculated clear sky global solar irradiation value, G_{cs}.
- The upper limit for the direct beam irradiation, $B^{\perp}$, is set as the calculated clear sky beam irradiance, $B^{\perp}_{cs}$.
- Because diffuse irradiation usually increases with the atmospheric turbidity and cloud amount, the lower limit for D is set to its calculated value under clear-sky and high atmospheric transparency condition, D_{cs}.

Following Geiger et al. (2002), the upper bound on G needs to be relaxed (i.e., $G < 2G_{cs}$) when the elevation of the sun above the horizon is small (i.e., smaller than 2°)

Table 4
Decision rules for quality flag attributes of solar radiation and sunshine duration data. These rules follow the flow diagram in Fig. 2. See Table 3 for the description of the labels used to denote the various tests.

Failure of semi-automated tests	
Detection of a snow-corrupted value of G by test PT5	The misleading value is flagged as erroneous.
Detection of shadow contamination in the G values by test S2	Misleading values are replaced by linearly interpoled values when possible or otherwise flagged as erroneous.
Calibration error detected by test SC1 or by visual inspection of scatter plots in $K - K_t$ and $K_n - K_t$ spaces	A compensation factor is applied to all misleading data when possible. These values are otherwise flagged as erroneous.
Failure of single parameter tests	
Failure of the physical threshold tests PT1, PT2, PT3a or the step test S1	The misleading value is flagged as erroneous.
Failure of the physical threshold test PT3b	All G data from sunrise to sunset are flagged as erroneous.
Failure of the persistence test P1	All data from sunrise to sunset of the misleading solar radiation component are flagged as erroneous.
Failure of intra-parameter tests (only for $\{G,B\}$ data that succeeded all previous tests)	
Failure of physical threshold tests PT4 or the quality code resulting from the quality envelope tests QE1 and QE2 equals 2	The B data is flagged as erroneous, except if quality envelope tests indicate that the direct sensor is reliable and the temperature is negative. In the latter situation, the G data is flagged as erroneous.
The quality code resulting from the quality envelope tests QE1 and QE2 equals 3	The B data is flagged as suspicious.
The quality code resulting from the quality envelope tests QE1 and QE2 equals 4	The B data is flagged as erroneous.
Failure of sunshine tests	
Failure of sunshine tests SD1 or SD2	The SD data is flagged as erroneous.
In all other cases	
Solar radiation data that succeeded these tests are considered as valid.	
Sunshine duration data that succeeded these tests are assigned with the same quality flag as the concurrent direct solar radiation data.	

because of the instrumental precision and atmospheric phenomena like refraction that are present.

Regarding lower limits on G and B, the opposite sky conditions need to be considered, i.e., heavily overcast conditions with low atmospheric transparency. If under these conditions the direct irradiance can be null, the diffuse irradiance never reaches zero during day time. Geiger et al. (2002) proposed to set a lower limit on G as 3% of the extraterrestrial value incident on horizontal surface in the case of hourly and daily cumulated data. Because this bound leads to excessively large rates of rejection in the case of sub-hourly data, we chose to replaced it by a less restrictive condition of the form: $G/E \geqslant 10^{-4}\cdot(h - 10)$ for all solar elevation angles, h, greater than 10° with the additional condition that the lower bound at 3% of the extraterrestrial radiation holds on daily average for all global solar radiation data acquired from sunrise to sunset.

Finally, because of the geometrical relation between global, direct and diffuse irradiance components, $B + D = G$, the basic inequalities $D \leqslant G$ and $B \leqslant G$ must be verified. Because the diffuse irradiance never reaches zero during day time and is minimal under clear-sky and high atmospheric transparency condition, the upper limit on the direct horizontal irradiation B can be further reduced to $G - D_{cs}$, which is well approximated by the condition $B \leqslant 0.95G$ when the solar elevation angle is not too small (i.e., greater than 2°). Because global irradiance is measured from a fixed instrument while direct irradiance requires to track the sun, the latter quantity is the most delicate to measure precisely (Molineaux and Ineichen, 1994). Hence, the values of direct solar radiation are the first to be considered doubtful when a lack of compatibility is identified between B and G values, except when there is sufficient insight to incriminate G rather than B. For instance, the condition $B < G$ is usually not satisfied when the global sensor is covered by snow. Evidence for snow-corrupted G values can rely on measurements of the ground-reflected solar radiation, R, that are available at most stations of the RMIB solar measurements network (see Table 1). An underestimated value of G leads to an overestimation of the surface albedo ($=R/G$) that cannot exceeds certain limits. This snow test is optional and let to the operator's decision.

The threshold values used for the three components of surface solar irradiation are summed up in Table 3.

2.2. Step tests

Step tests check for a plausible rate of change from a preceding acceptable level (e.g., detection of unrealistic spikes or jumps in values, or dead band caused by blocked sensors).

First, the variations of the solar radiation measurements between two successive timestamps cannot exceed certain limits. Maximal variations of up to 800 W m^{-2} are usually recommended for instantaneous (i.e., 1 min averaged) solar irradiances (Shafer et al., 2000; WMO, 2007). In the case of 10 min data, we preferred to impose upper bounds on the variations of the clearness index $K_t = G/E$, the beam transmittance $K_n = B/E$ and the diffuse transmittance $K_d = D/E$,

$$\begin{aligned}\left|\frac{G(t)}{E(t)}-\frac{G(t-1)}{E(t-1)}\right| &< 0.75,\\ \left|\frac{B(t)}{E(t)}-\frac{B(t-1)}{E(t-1)}\right| &< 0.65,\\ \left|\frac{D(t)}{E(t)}-\frac{D(t-1)}{E(t-1)}\right| &< 0.35,\end{aligned} \tag{1}$$

where $t-1$ and t are two successive timestamps. Because of the instrumental precision, the conditions (1) are considered only when the solar elevation angle is greater than 2°.

Second, the variations of the various solar radiation parameters need to be, in some sense, compatible with each other, e.g., large fluctuations of the beam irradiance need to be reflected in the time series of both B and G. Although it is difficult to set strict conditions for compatible rates of change in the overall case, a comparison of the G and B diurnal evolutions is very useful to detect shadow contaminations in the G data. In such a case, G exhibits large variations (i.e., a rapid drop of the values followed by a sudden increase when the shadow disappears) that are either not reported or reported with a slight time delay in the B time series (see Fig. 3). The ratio $\frac{G(t)-G(t-1)}{B(t)-B(t-1)}$ is large either in the positive direction (e.g., greater than 3) or in the negative direction (e.g., smaller than -1). In addition, a significant variation of the clearness index is observed, $\left|\frac{G(t)}{E(t)}-\frac{G(t-1)}{E(t-1)}\right|>0.1$. Although this situation is most often already identified by the non satisfaction of the compatibility condition $B(t)<0.95G(t)$, comparing the rates of change enables to further identify the origin of the failure. The final attribution has however to rely on a visual inspection of the G and B time series. Because only a few number of successive data are usually misleading, the erroneous data can often be replaced by linearly interpolated values.

All threshold values used in these step tests are provided in Table 3.

2.3. Persistence tests

Persistence tests check the variability of the measurements. When a sensor fails, it often reports a constant value, leading to a small standard deviation. The standard deviation can even be null if the sensor is out for the entire reporting period. On the other hand, when an instrument works intermittently and produces reasonable values interspersed with misleading values, the variability can sometimes be excessively high. Thus, data should be flagged for further check when the standard deviation exceeds lower and upper limits. Statistical quantities are computed from all solar radiation data acquired over the day (from sunrise to sunset). For all three components of the solar radiation, the lower bound on the standard deviation, σ, of the daily series of data is expressed as an increasing function of the mean μ:

$$\begin{aligned}\frac{1}{8}\mu\left(\frac{G}{E}\right) \leqslant \sigma\left(\frac{G}{E}\right) &\leqslant 0.35,\\ \frac{1}{6}\mu\left(\frac{D}{E}\right) \leqslant \sigma\left(\frac{D}{E}\right) &\leqslant 0.2,\\ \lambda \leqslant \sigma\left(\frac{B}{E}\right) &\leqslant 0.3,\end{aligned} \tag{2}$$

with the lower limit $\lambda=\begin{cases}0 & \text{if } \mu\left(\frac{B}{E}\right)\leqslant 0.01,\\ \frac{1}{2}\mu\left(\frac{B}{E}\right) & \text{if } 0.01<\mu\left(\frac{B}{E}\right)\leqslant 0.2,\\ 0.1+\frac{1}{20}\left(\mu\left(\frac{B}{E}\right)-0.2\right) & \text{otherwise.}\end{cases}$

Variations in the incoming solar radiation due to the cloud cover are not expected to exceed these upper limits, at least over Belgium.

2.4. Quality envelope tests

In contrast to previous tests that dealt with data acquired on short temporal horizons (from a single timestamp to a full day of data), quality envelope tests first require to accumulate data over a longer time period (e.g., two months). Based on Maxwell et al. (1993) and Younes et al. (2005), this type of tests operate in a dimensionless space within which expectancy envelopes or quality envelopes are defined. The variables that form the abscissa and ordinate of this

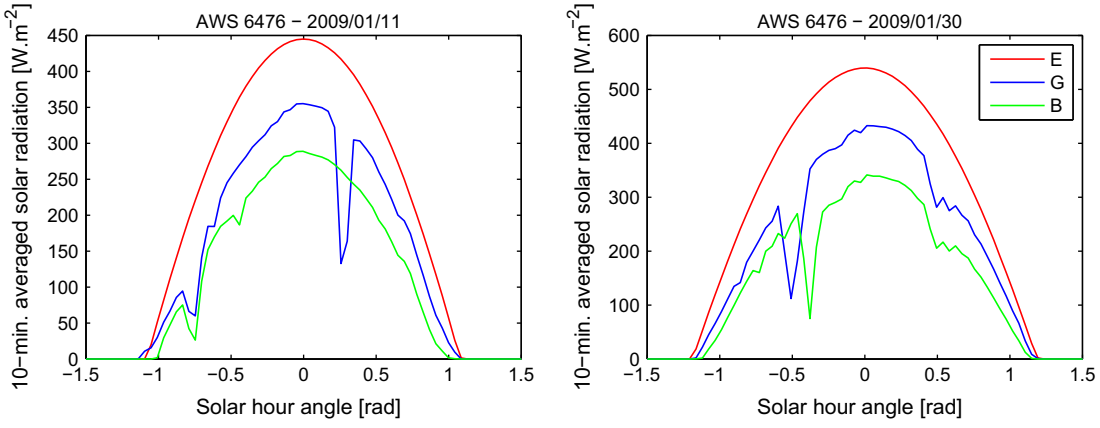

Fig. 3. Time series of solar radiation data contaminated by shadow (contamination of the G data only (left panel) or both the G and B data (right-panel)). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

dimensionless space are the clearness index $K_t = G/E$ for the abscissa and either the atmospheric transmission of the beam radiation as in Maxwell et al. (1993), $K_n = B/E$, or the diffuse ratio $K = D/G$ as in Younes et al. (2005). When no measurement of the diffuse solar radiation is available, the diffuse ratio is estimated from the difference between the global and the beam horizontal radiation, $K = (G - B)/G$. In the case of sub-hourly data, these tests are restrained to solar radiation data acquired for a solar elevation angle greater than 2°. Typical distributions of the data in the $K_t - K$ and $K_t - K_n$ spaces are provided in the top plots of Fig. 5. Although the philosophy is similar, the proposed approach differs from Younes et al. (2005) and Maxwell et al. (1993) in the establishment of the boundary shapes around the data.

The SERI QC quality control algorithm proposed by Maxwell et al. (1993) relies on the use of climatological boundaries established from scatter plots of historical good quality solar radiation data in the $K_t - K_n$ space. When performing data quality control, the boundaries are first automatically selected from a set of predefined empirical boundary shapes using a criterion whereby the selection of tighter fitting boundaries results in a greater percentage increase in errors than the percentage decrease in the acceptable data between the boundaries. The position of the selected boundaries are then manually adjusted such that up to 5% of the data lay outside the boundaries. This criterion was based both on the assumption that some of the data were in error and a desire to limit the acceptance of erroneous data to small percentages, while similarly limiting the rejection of good data.

Younes et al. (2005) proposed a statistical approach to construct the $K - K_t$ quality control envelopes on the analyzed data. Basically, the K_t range of data is divided into n bands of equal width, within which the mean and standard deviation of the K values, μ_i and σ_i, are calculated. The top and bottom boundary shapes are identified by fitting a polynomial through the points $\{\mu_i + a\ \sigma_i\}_{i=1,\ldots,n}$ and $\{\mu_i - a\sigma_i\}_{i=1,\ldots,n}$, respectively. Polynomial values are limited between 0 and 1 to respect the physical limits of K. Finally, the cut-off points (the respective intersection of the upper and lower polynomials with the $K = 1$ and $K = 0$ limits) are selected by visual inspection. The rejection rate lies usually between 2% and 6%.

The proposed method follows the statistical approach of Younes et al. (2005) in the establishment of the boundary shapes around the data, but does not require any manual operation and can be used in both $K_t - K$ and $K_t - K_n$ spaces. Moreover, our analysis uses recorded and simulated data, and considers both climatological envelopes as in Maxwell et al. (1993) and envelopes directly computed from the analyzed data.

Direct irradiance is the most delicate quantity to measure precisely (Molineaux and Ineichen, 1994) and deficiencies in the direct sensor may remain unseen in the absence of the diffuse component. Therefore, the procedure first starts by estimating the direct solar horizontal radiation, $\widehat{B}$, from the measured global horizontal solar radiation G using the formulation of Skartveit and Olseth (1987) (see Section 3.1). Although direct solar radiation is difficult to estimate precisely, the accuracy of this G-to-B model is sufficient to detect gross deviations in the measurements. Both the measured and the estimated direct solar radiation components are then plotted in the $K_t - K$ and $K_t - K_n$ spaces (see Fig. 4). For each distribution of points, a mean curve as a function of K_t, $\mu(K_t)$ or $\hat{\mu}(K_t)$, is computed by dividing the K_t range in n bands and fitting a polynomial through the mean of the K (or K_n) values in each band (see the blue and green lines on the four panels in Fig. 4). The root mean squared distance between the two mean curves,

$$\Delta = \sqrt{\frac{1}{N}\sum_{k=1}^{N}\left(\mu(K_t^{(k)}) - \hat{\mu}(K_t^{(k)})\right)^2},$$

is then evaluated, where N is the number of points in the scatter plots and $K_t^{(k)}$ is the clearness index of the kth point. Depending on the value of Δ with respect to a given threshold (see Table 5), the direct sensor is considered either as reliable (see top plots of Fig. 4) or as deficient (see bottom plots of Fig. 4). In the former case, quality envelopes are directly constructed from the measured data by fitting polynomials through the points located at a few standard deviations above and below the mean in each band to define the top and the bottom boundary shapes, as in Younes et al. (2005). The quality envelope width was set to five times the standard deviation in each band, which is 25% more than suggested by Younes et al. (2005). This choice was made to avoid as much as possible the rejection of valid observations potentially related to rare events. In addition to the fitted top and bottom boundaries, an upper limit $K_t^{\max}$ is set on the K_t values in such a way that $K_t^{\max}$ is the smallest value for which the right outliers (i.e., the data points such that $K_t > K_t^{\max}$) are distant by more than 0.02 in K_t values from the accepted data points.

When the direct sensor is found inefficient, a climatological quality envelope is used to filter the data. The climatological quality envelope was determined from two years of data (2008 and 2009) that succeeded all previous quality control tests. Examples of quality envelopes are provided in Fig. 5.

Depending on the instruments' reliability and the position of the data with respect to the quality envelope, a quality code ranging from 1 to 4 is given to each data pair $\{G, D\}$ or $\{G, B\}$, as detailed in Table 6. When no measurement of the diffuse radiation is available, two quality codes (one for each of the $K_t - K$ and $K_t - K_n$ spaces) are assigned to each pair $\{G,B\}$. Because these quality codes can differ from each other, we retain the smallest value of them, i.e., a sensor is considered as being deficient only if it found as it in both spaces.

Finally, it should be mentioned that a visual inspection of the data distributions in the $K_t - K$ and $K_t - K_n$ spaces can sometimes highlight failures in the global sensor. This is particularly under concern as the default assumption is to attribute the wrong measurements to the direct sensor. For instance, a calibration error on one sensor is reflected

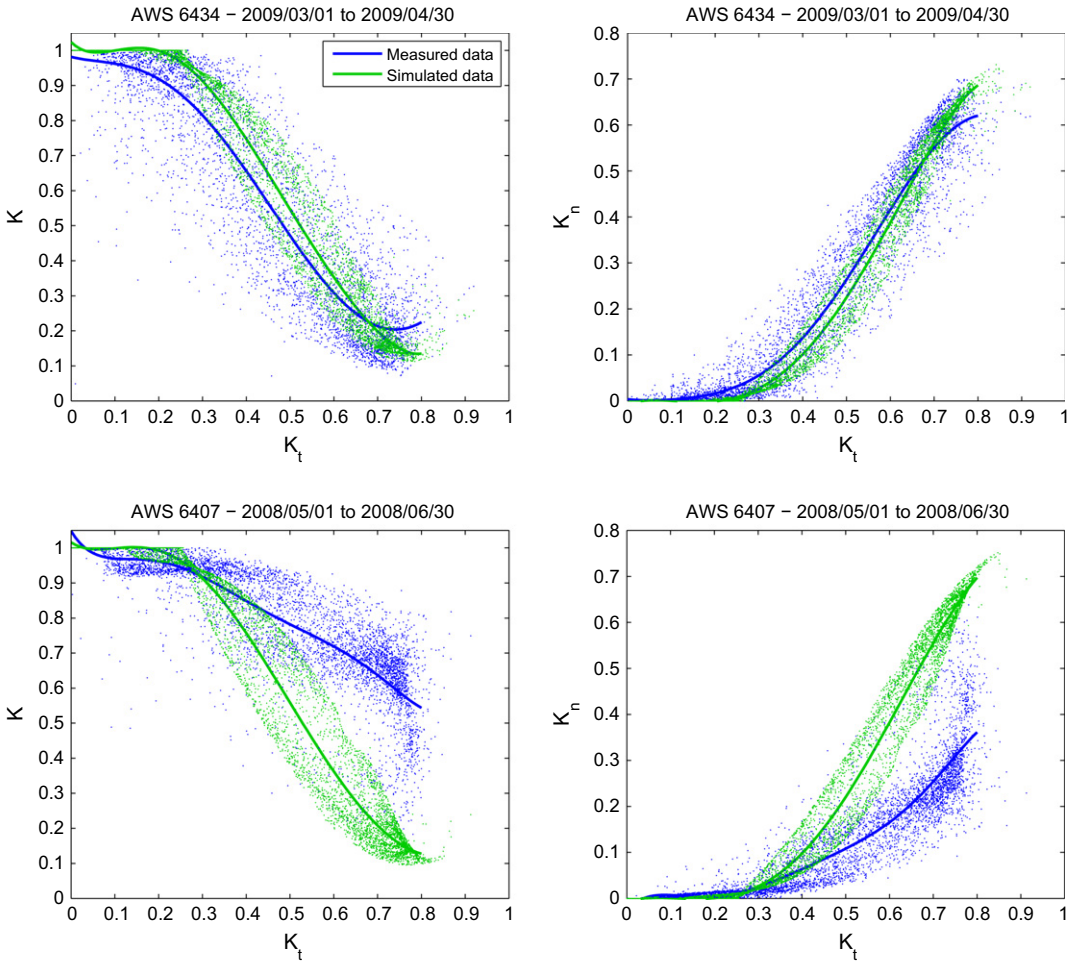

Fig. 4. Scatter plots and corresponding mean curves of solar radiation data in the $K_t - K$ and $K_t - K_n$ spaces for reliable instruments (top plots) and in case of a deficient instrument for the measure of the beam solar radiation B (bottom plots). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

Table 5
Upper limits on the root mean squared distance Δ used in quality envelope tests to evaluate the overall reliability of the direct sensor when the model of Skartveit and Olseth (1987) is used to estimate the direct solar radiation. The seasonal dependence reflects the larger dispersion of the data in summer because of the higher frequency of clear and partly clear skies.

Season	$K_t - K$ space	$K_t - K_n$ space
Winter	0.085	0.050
Spring	0.115	0.070
Summer	0.125	0.075
Autumn	0.115	0.070

in the $K_t - K$ or $K_t - K_n$ spaces by two distinct clusters of points (see Fig. 6).

2.5. Spatial consistency tests

Data can be analyzed by comparing neighboring stations with each other. Because the cloud cover can be highly variable even on small distances, spatial consistency tests have to rely on cumulated data (at least daily values). As an alternative, neighboring stations might be compared on periods with concurrent clear sky. Finally, these tests should be restained to sites with similar climate and geography.

Here, the approach proposed by Terzenbach (1995) is used to evaluate the spatial consistency between daily totals of global solar irradiation data. In this method, the value at a station located in x_0, $G(x_0)$, is compared against an estimation obtained through a spatial interpolation of the measurements performed at nearby locations x_i, i.e., $\widehat{G}(x_0) = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i G(x_i)$, where N is the number of stations located within a given maximum distance R of x_0 and the interpolation coefficients w_i are a decreasing function of the distance d_i between the locations x_0 and x_i, e.g., $w_i = \left(\frac{1-d_i/R}{d_i/R}\right)^2$. The bias $|\widehat{G}(x_0) - G(x_0)|$ is expected to stay below certain limits. In practice, biases that are significantly larger (e.g., larger by more than 50%) than the mean bias over all sites are checked manually to attest whether their origin is natural or instrumental.

2.6. Sunshine tests

Sunshine duration values are first assigned with the same quality index than the associated B data. Then, several additional checks are performed if the B value has

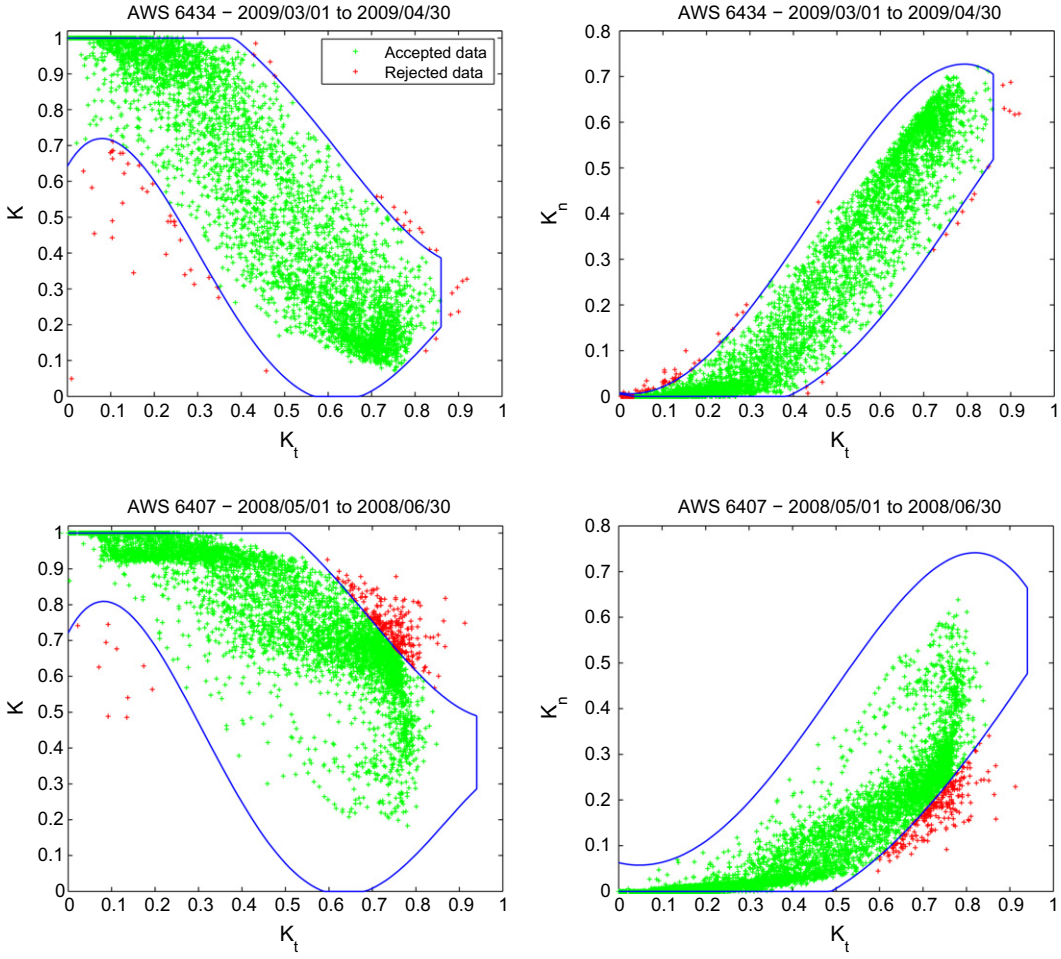

Fig. 5. Quality envelopes in the $K_t - K$ and $K_t - K_n$ spaces. In case of reliable instruments, these envelopes are directly computed from the analyzed data (top plots). Climatological envelopes are otherwise used (bottom plots, case of a deficient direct sensor). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

Table 6
Quality code information for quality envelope tests.

Quality code	Status of the sensor	Position of the data
1	Reliable	Inside the quality envelope
2	Reliable	Outlier
3	Deficient	Inside the quality envelope
4	Deficient	Outlier

succeeded the quality assessment tests. Sunshine duration values are to lie between zero and the sampling period $SD_{\max}$ (i.e., 10 min or 30 min for the RMIB data set). Furthermore, in accordance with the WMO definition of the sunshine duration, the direct solar irradiance averaged on the period $SD_{\max}$ cannot exceed 120 W m^{-2} if $SD = 0$ and has to be above the lower bound $\frac{SD}{SD_{\max}}$ 120 W m^{-2}.

3. Solar radiation modeling

Because the availability of solar radiation data in time are important to assess the performance of solar energy systems, additional procedures have been developed to provide users with an estimation of data initially lacking or failing the quality assessment criteria. When two out of the three components of the solar radiation succeed the quality assessment checks, the remaining component can be directly estimated from the relation $D + B = G$. When only one solar radiation parameter is available, one can resort to empirical models to estimate one missing solar parameter from the available one and then use the relation $D + B = G$ to derive the last parameter. Further models enable to estimate the sunshine duration from the knowledge of the solar radiation parameters.

In this section we evaluate some empirical models based on 2 years (2008 and 2009) of validated 10 min solar data with a solar elevation angle greater than 2°. Models' performance is quantified by means of the following statistical indices:

- the mean bias error $MBE = \mathbb{E}[\hat{x} - x]$,
- the mean absolute error $MAE = \mathbb{E}[|\hat{x} - x|]$,
- the root mean square error $RMSE = \sqrt{\mathbb{E}\left[(\hat{x} - x)^2\right]}$,
- the relative mean bias error $rMBE = \mathbb{E}\left[\frac{\hat{x} - x}{x}\right]$,
- the relative mean absolute error $rMAE = \mathbb{E}\left[\frac{|\hat{x} - x|}{x}\right]$ and,

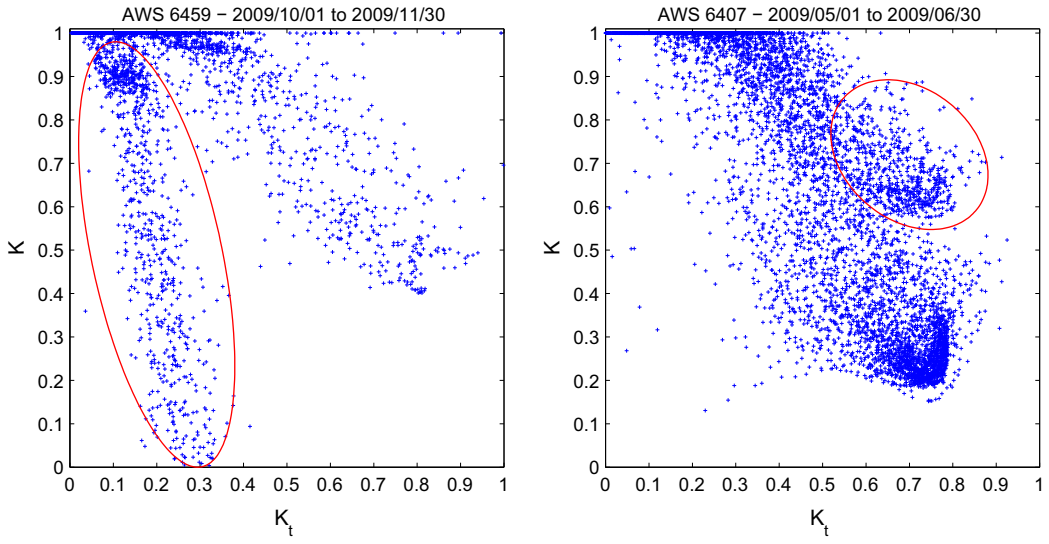

Fig. 6. Calibration errors on solar radiation measurements as highlighted on scatter plots of solar radiation data in the $K_t - K$ space. On the left-hand plot, measurements of the global solar radiation circled in red are underestimated by a factor of 2.35. On the right-hand plot, circled measurements of the direct solar radiation are underestimated by a factor of two. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

– the relative root mean square error $rRMSE = \sqrt{\mathbb{E}\left[\left(\frac{\hat{x}-x}{x}\right)^2\right]}$,

where x is the measurement, $\hat{x}$ is the model output and $\mathbb{E}$ is the expectation operator. Note that further indices can be found in the literature (e.g., Espinar et al., 2009). For a proper estimation of the relative error statistics, the measurement x is chosen greater than 1.67 W h m^{-2} (i.e., greater than 10 W m^{-2} in terms of average irradiance) when evaluating solar radiation models. These statistical indices are not derived for sunshine duration models to include zero values in the analysis. At the end, a data set with 140,000 (resp. 244,000) instances that account for various locations in Belgium, sky conditions, solar elevation angles and seasons was used to compare the models for solar radiation (resp. sunshine duration) described in the following sections.

3.1. Models for the direct solar radiation

Derivation of models to convert hourly global solar radiation to hourly direct (or diffuse) radiation has been an intensive research topic (see Perez et al. (1990b) for a review). Four models have been evaluated, namely the Erbs et al. (1982) model (hereafter referred as E82), the Skartveit and Olseth (1987) model (hereafter S87), the Maxwell (1987) model (hereafter M87) and the Perez et al. (1992) model (hereafter P92). In these models, the beam horizontal solar radiation is computed as a fraction of either the global horizontal solar radiation (E82 and S87) or the extraterrestrial horizontal solar radiation (M87 and P92) with the ratio $\frac{B}{G}$ or $\frac{B}{E}$ being an increasing function of the clearness index K_t. These models also differ by the way the dependence in K_t is parameterized and by the additional input parameters they might require. For instance, while E82 exclusively relies on the clearness index, the other models require either the solar elevation angle (S87) or the airmass (M87 and P92). It should be noted that P92 is an extended version of M87 that accounts for the dynamics of the global radiation time series and the atmospheric precipitable water, when available.

Although these four models are intended for hourly data, they were evaluated against 10 min data (see Table 7). Note that P92 was used without providing information on the atmospheric precipitable water. In overall, this model performs better in terms of the absolute error indices, while the rMAE and rRMSE errors are the smallest for S87. Compared to its static version (i.e., M87), it is clear that exploiting the dynamics of the G time series gives a noticeable improvement. Regarding the impact of sky conditions (Fig. 7), all models underestimate the direct radiation for overcast skies and slightly overestimate it for very clear skies. For all models, the rRMSE decreases for larger clearness indices. As far as the influence of the solar elevation angle is concerned (Fig. 7), the strongest dependence in this parameter is

Table 7
Comparison of the models E82 (Erbs et al., 1982), S87 (Skartveit and Olseth, 1987), M87 (Maxwell, 987) and P92 (Perez et al., 1992) that convert global radiation into direct radiation against good quality 10 min data. The absolute error statistics MBE, MAE and RMSE are expressed in W h m^{-2}.

	E82	S87	M87	P92
MBE	−0.165	−2.036	2.189	0.360
MAE	7.781	7.226	7.599	6.762
RMSE	10.924	10.225	11.246	9.624
rMBE	−0.088	−0.085	0.071	−0.016
rMAE	0.384	0.346	0.376	0.349
rRMSE	0.596	0.546	0.642	0.584

observed for E82. This is not really surprising as this model does not incorporate solar elevation as an input parameter. In overall, S87 and P92 appear as the best models and exhibit a rather equivalent performance. Missing or erroneous direct solar radiation within the RMIB data set are therefore estimated by means of the model S87 that is conceptually more simple.

3.2. Models for the global solar radiation

Sunshine duration was traditionally the main and often the only measurement of solar radiation. Several models have been proposed to convert sunshine duration values into global solar radiation values. The most widely used model of Ångström (1924) and Prescott (1940) propose a linear relationship between the clearness index G/E and the sunshine fraction $SD/SD_{\max}$,

$$\frac{G}{E} = a + b\frac{SD}{SD_{\max}}, \tag{3}$$

where the coefficients a and b are site-dependent. More complex nonlinear models have been proposed in the literature but without showing significant improvement with respect to the linear formulation (3) (Almorox and Hontoria, 2004). Although model (3) was initially proposed for daily data, it is often used in the case of hourly data (see, e.g., Yang and Koike (2005)).

As an alternative, the global solar radiation can be inferred from the direct solar radiation by inverting the models discussed in Section 3.1. In the case of E82 (see Appendix A for the detailed formulation of this model), the clearness index K_t can be expressed as a function of the beam transmittance $K_n = B/E$,

$$\begin{cases} \text{if } 0 < K_n \leqslant 0.004356 : K_t = \sqrt{K_n/0.09}, \\ \text{if } 0.004356 < K_n < 0.668 : \text{find the root of } f(K_t) - K_n = 0 \\ \quad \text{where, } f(K_t) = 0.0489K_t + 0.1604K_t^2 - 4.388K_t^3 \\ \qquad\qquad + 16.638K_t^4 - 12.336K_t^5, \\ \text{if } 0.668 \leqslant K_n < 0.835 : K_t = K_n/0.835. \end{cases} \tag{4}$$

Special care has to be taken for values of K_n that are close to zero. The global solar radiation cannot fall below a certain limit (e.g., $K_t \geqslant 0.15$) because the diffuse radiation never reaches zero during day time. The equation $f(K_t) - K_n = 0$ has to be solved in a numerical manner by means, for instance, of the bisection method. Numerical inversion of S87, M87 and P92 is also possible but more involved because of the additional input parameters. Although these models are not considered in the forthcom-

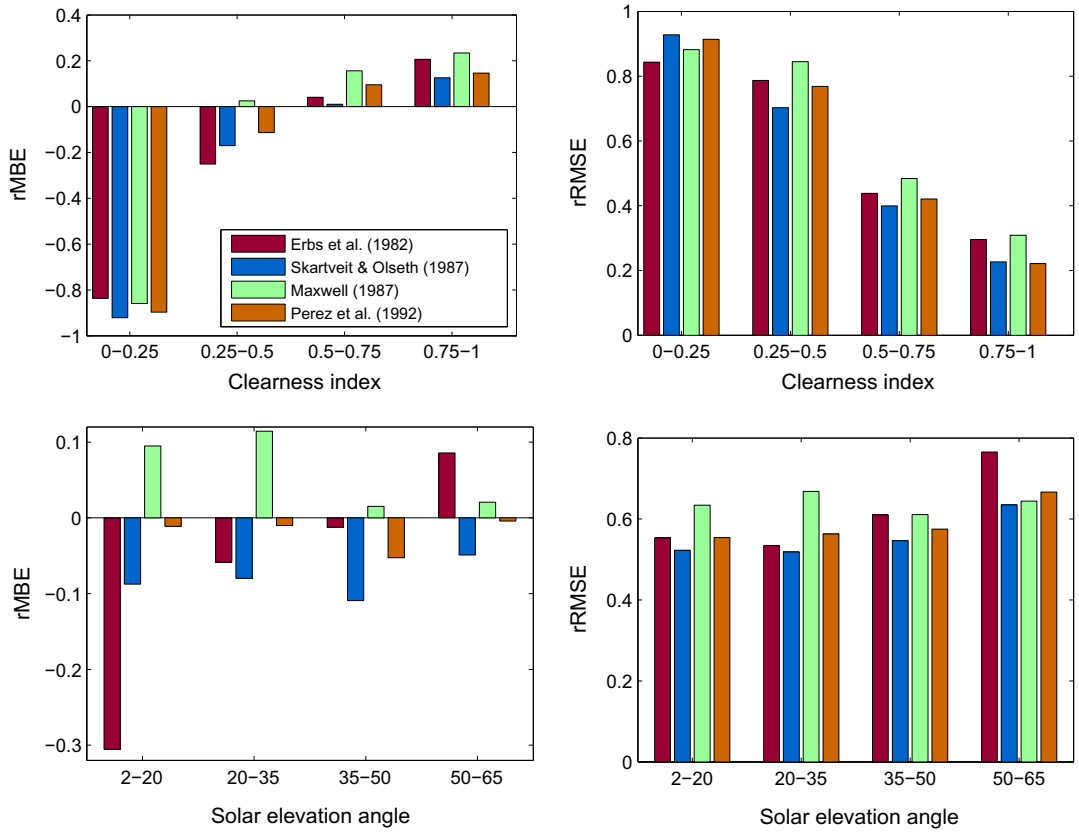

Fig. 7. Distribution of the relative mean bias error (rMBE) and relative root mean square error (rRMSE) between measured and estimated values of direct solar radiation as a function of sky conditions (top plots) and solar elevation angle (bottom plots). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

Table 8
Comparison of the models described by Eqs. (3) and (4) that estimate global radiation against good quality 10 min data. The absolute error statistics MBE, MAE and RMSE are expressed in W h m^{-2}.

	Model (3) with site-specific coefficients	Model (3) with averaged coefficients	Model (4)
MBE	−2.450	−2.446	1.314
MAE	9.090	9.376	6.433
RMSE	12.071	12.466	8.912
rMBE	0.051	0.057	0.089
rMAE	0.183	0.189	0.149
rRMSE	0.286	0.299	0.264

ing comparison, one can expect, in view of the results discussed in Section 3.1, that the inverted version of S87 may slightly outperform the model (4).

Because model (3) involves site-dependent coefficients, the RMIB data set was split randomly in two subsets of equal size for each site to estimate the coefficients in a first hand and evaluate the models on the other hand. In addition, coefficients averaged over all considered stations (i.e., $a = 0.295$ and $b = 0.335$) were used to evaluate the spatial sensitivity of the model. Although the site-specific coefficients may vary by up to 16% with respect to the averaged values, the performance degradation of the averaged model is rather small (e.g., the RMSE – see Table 8 – is increased by 3.54%). Following the study of Almorox and Hontoria (2004), more complex polynomial regression models that estimate global solar radiation values from sunshine duration measurements were considered (although not reported in Table 8), but we did not notice any improvement with respect to the linear model (3). Hence, the model (4) that is based on direct solar radiation measurements performs significantly better than sunshine duration-based models to estimate global solar radiation values, i.e., all error indices are smaller for this model, except the rMBE that is small for all models (Table 8). Sunshine duration values, which are truncated from direct radiation values with respect to the 120 W m^{-2} threshold, provide in fact a less accurate information on solar radiation.

As far as the impact of the sky conditions is concerned, both models (3) and (4) overestimate the global solar radiation for overcast sky and underestimate it for clear skies (Fig. 8). The model (4) exhibits the best agreement with the measurements in case of intermediate and clear sky conditions (i.e., $K_t > 0.25$), while both models encounter difficulties in overcast conditions.

In view of its overall better performance, model (4) is used to estimate missing and erroneous global solar radiation values.

3.3. Models for the sunshine duration

When measurements of the global solar radiation are available, sunshine duration values can be derived from the linear relation (3). More complex polynomial regression model that rely on global and/or direct solar radiation measurements are in addition considered in this study (see Table 9 for the formulation of these models). These models involve as input parameters either the clearness index $K_t = G/E$ or the modified (i.e., elevation angle-independent) clearness index K_t' proposed by Perez et al. (1990a) and/or the beam ratio B/G. In accordance with the physical limits of the sunshine duration, the outputs of these polynomial models are truncated below at zero and above at SD_{max}. As a consequence, the zero mean bias property of the regression models is lost.

As in Section 3.2, the RMIB data set was split in two subsets to estimate the regression coefficients and compare the models on distinct data. Average models over Belgium have been inferred in addition to the site-specific ones. Because the degradation in performance was small, the following comparative study is focused on the average models. When models that involve only one input parameter are compared (SD_1 to SD_6), the smallest MAE and RMSE are observed for those based on the beam ratio B/G (Table

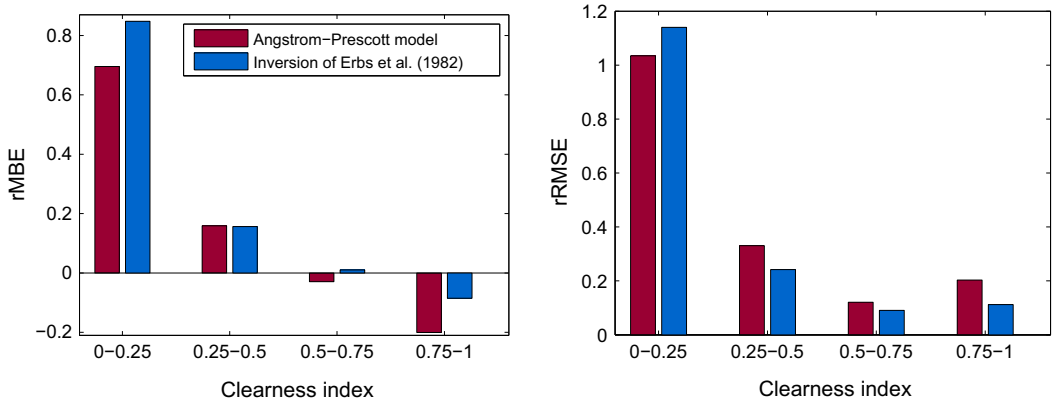

Fig. 8. Distribution of the relative mean bias error (rMBE) and relative root mean square error (rRMSE) between measured and estimated values of global solar radiation as a function of sky conditions. The Ångström–Prescott model refers to model (3) with the averaged coefficients, while the inverted version of Erbs et al. (1982) is model (4). (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

Table 9
Polynomial regression models to estimate sunshine duration values. The sunshine fraction $\frac{SD}{SD_{max}}$ is truncated below at zero and above at one. The regression coefficients are computed on the basis of good quality data from several sites in Belgium.

Model label	Model equation	Coefficient values
SD_1	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t$	$a_0 = -0.266$ $a_1 = 1.789$
SD_2	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t + a_2 K_t^2$	$a_0 = -0.241$ $a_1 = 1.612$ $a_2 = 0.214$
SD_3	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t'$	$a_0 = -0.309$ $a_1 = 1.608$
SD_4	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t' + a_2 K_t'^2$	$a_0 = -0.194$ $a_1 = 0.914$ $a_2 = 0.738$
SD_5	$\frac{SD}{SD_{max}} = a_0 + a_1 \left(\frac{B}{G}\right)$	$a_0 = -0.002$ $a_1 = 1.407$
SD_6	$\frac{SD}{SD_{max}} = a_0 + a_1 \frac{B}{G} + a_2 \left(\frac{B}{G}\right)^2$	$a_0 = -0.054$ $a_1 = 2.428$ $a_2 = -1.355$
SD_7	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t' + a_2 \frac{B}{G}$	$a_0 = -0.061$ $a_1 = 0.519$ $a_2 = 1.033$
SD_8	$\frac{SD}{SD_{max}} = a_0 + a_1 K_t' + a_2 K_t'^2 + a_3 \frac{B}{G} + a_4 \left(\frac{B}{G}\right)^2$	$a_0 = -0.044$ $a_1 = -0.175$ $a_2 = 0.562$ $a_3 = 2.104$ $a_4 = -1.277$

10). When only the global solar radiation is available, the estimation is slightly more accurate when the modified clearness index is used instead of the usual one. The best performance is however obtained when both global and direct solar radiation measurements are involved through the modified clearness index and the beam ratio (SD_7 and SD_8). Finally, quadratic models provide a significant improvement with respect to their linear counterparts (i.e., SD_2, SD_4, SD_6, SD_8 versus SD_1, SD_3, SD_5, SD_7, respectively).

As far as the impact of sky condition is concerned (Fig. 9), the models overestimate the sunshine duration for overcast to intermediate skies (i.e., for $0 \leqslant K_t < 0.5$) and underestimate it for clearer sky conditions (i.e., for $0.5 \leqslant K_t < 1$). All models perform almost equivalently for overcast conditions. Using measurements of the direct solar radiation is particularly beneficial for intermediate sky conditions. In the case of very clear skies, the best scores are however obtained by models that involve global solar radiation values (i.e., all models except SD_5 and SD_6).

Finally the performance of SD_8 was evaluated when one of its both inputs is not available (Table 10). Although the use of an estimation instead of the actual measurement degrades the performance of the model ($SD_{8,\widehat{B}}$ and $SD_{8,\widehat{G}}$ versus SD_8), the resulting MAE and RMSE are smaller than those observed for models based on a single measurements ($SD_{8,\widehat{B}}$ versus SD_4 and $SD_{8,\widehat{G}}$ versus SD_6). Hence, missing or erroneous sunshine duration values are estimated by means of SD_8 and we resort to solar radiation models when one of the both inputs is not available.

Table 10
Comparison of models to estimate sunshine duration against good quality 10 min data. The error statistics MBE, MAE and RMSE are expressed in minutes. The model labels refer to the descriptions in Table 9. The subscripts $\widehat{B}$ and $\widehat{G}$ denote that estimations are used for either B or G instead of the actual measurement. These estimations are computed by the model S87 (Skartveit and Olseth, 1987) for $\widehat{B}$ and by model (4) for $\widehat{G}$.

Model label	MBE	MAE	RMSE
SD_1	0.088	1.372	2.034
SD_2	0.053	1.365	2.032
SD_3	0.169	1.333	1.981
SD_4	0.032	1.273	1.942
SD_5	−0.220	0.818	1.447
SD_6	0.109	0.716	1.281
SD_7	0.360	0.936	1.376
SD_8	0.077	0.663	1.179
$SD_{8,\widehat{B}}$	−0.198	0.871	1.685
$SD_{8,\widehat{G}}$	0.133	0.671	1.230

3.4. Uncertainty of the filled data set

In addition to the limitations inherent to the empirical models, uncertainty in filled solar radiation time series is also function of the number of estimations they contain and the time at which they are applied. To assess the reliability of the filled data set, daily totals derived from time series of G, B and SD data containing a varying proportion of estimations were compared. Basically, we compiled a set of complete daily time series of good quality 10 min data (i.e., G, B and SD data were available over the entire day and succeeded the quality assessment tests). Based on these complete daily time series, new daily time series containing an increasing proportion of estimations were generated by replacing at randomly selected timestamps measurements by the corresponding estimations. Daily totals derived from the original and the synthetic time series were finally compared against each other. On average, the discrepancy between the filled and the measured data appears to increase linearly with the proportion of estimations included in the time series but stays within reasonable limits (Table 11). For instance, when daily totals are derived from as much as 50% of estimations and 50% of measurements, the RMSE amounts to 9%, 7.2% and 5.7% of the average of daily value for the direct solar radiation, global solar radiation and sunshine duration, respectively. These results might be representative for data acquired with identical instruments in regions with similar climate.

4. Conclusions

This paper describes the procedures developed at the Royal Meteorological Institute of Belgium for post-

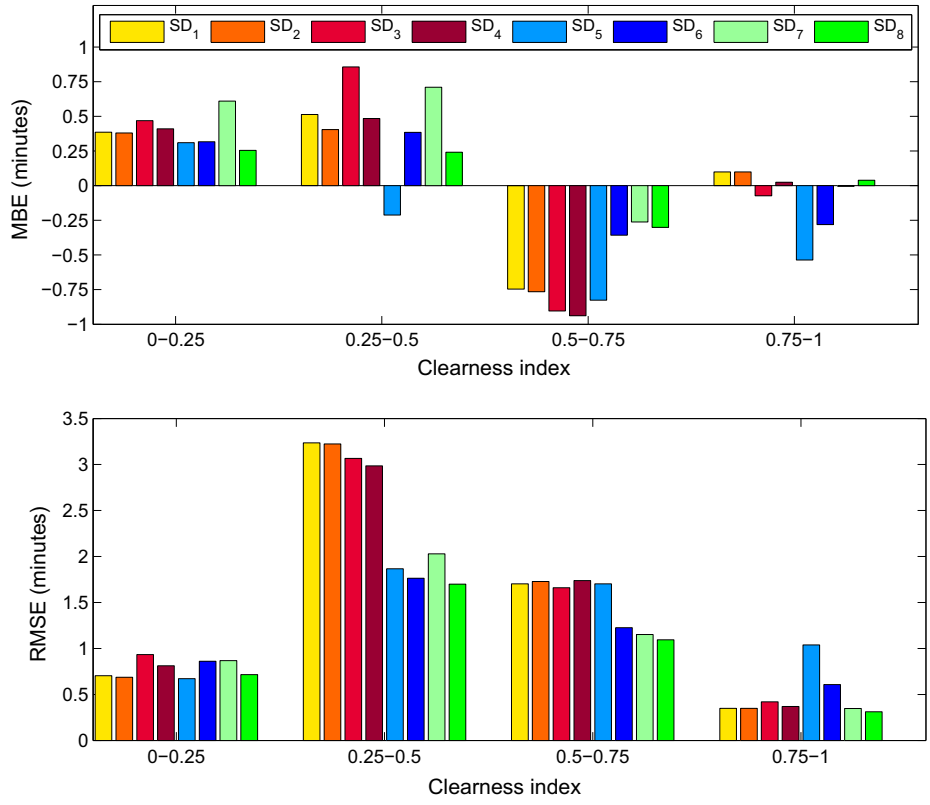

Fig. 9. Distribution of the mean bias error (MBE) and root mean square error (RMSE) between measured and estimated values of sunshine duration as a function of sky conditions. The model labels refer to the descriptions in Table 9. (For interpretation to colours in this figure, the reader is referred to the web version of this paper.)

Table 11
Comparison of measured daily totals against daily totals that contain a proportion p of estimations. The comparison indices are given in absolute value (i.e., in W h m^{-2} for the global and direct solar radiations and in minutes for the sunshine duration) as well as in relative value with respect to the average of the measured daily totals.

	$p = 10$ (%)	$p = 25$ (%)	$p = 50$ (%)	$p = 75$ (%)	$p = 100$ (%)
Estimation of direct solar radiation values by means of the model S87 (Skartveit and Olseth, 1987)					
MBE	−0.106 (−0.008)	−0.022 (−0.002)	−0.224 (−0.017)	−0.304 (−0.023)	−0.458 (−0.034)
MAE	17.62 (1.32)	38.72 (2.91)	73.01 (5.49)	107.03 (8.04)	141.24 (10.62)
RMSE	30.21 (2.27)	63.92 (4.80)	119.19 (8.96)	173.77 (13.06)	228.66 (17.18)
Estimation of global solar radiation values by means of model (4)					
MBE	−9.362 (−0.348)	−23.291 (−0.865)	−46.691 (−1.734)	−69.870 (−2.595)	−93.331 (−3.467)
MAE	31.08 (1.15)	72.98 (2.71)	142.70 (5.30)	212.38 (7.89)	282.34 (10.49)
RMSE	45.11 (1.68)	101.49 (3.77)	194.76 (7.23)	288.11 (10.70)	381.80 (14.18)
Estimation of sunshine duration values by means of the model SD_8 of Table 9					
MBE	0.303 (0.113)	0.704 (0.263)	1.441 (0.538)	2.168 (0.809)	2.890 (1.079)
MAE	2.57 (0.96)	5.52 (2.06)	10.33 (3.86)	15.15 (5.66)	19.88 (7.42)
RMSE	4.08 (1.52)	8.37 (3.13)	15.36 (5.73)	22.39 (8.36)	29.33 (10.95)

measurement quality control of solar radiation data from our radiometric stations. The data consists in measurements of the global, direct and diffuse solar radiations as well as sunshine duration with either a 10 min or a 30 min time step. The resulting quality control scheme is based on previously proposed as well as new quality criteria. It involves a set of optional semi-automated procedures to detect specific error types followed by fully-automated procedures. Because solar energy applications need continuous time series of radiation data to correctly assess the usefulness of the particular application and in its implementation, additional procedures have furthermore been established to fill missing values (data initially lacking or removed via quality checks) in the time series of solar radiation data. These procedures rely on empirical models that estimate missing solar parameters from available ones. Several of such models, either found in the literature or proposed in this paper, have been evaluated on

the basis of 10 min data that succeeded the quality control assessments. As far as the estimation of direct solar radiation is concerned, we selected the G-to-B model of Skartveit and Olseth (1987), which exhibited a root mean square error (RMSE) of 10.2 W h m^{-2}. The inverted version of the G-to-B model of Erbs et al. (1982), which lead to a RMSE of 8.9 W h m^{-2}, was chosen to estimate global solar radiation values from direct solar radiation measurements. A regression model was also proposed to estimate sunshine duration from measurements of global and direct solar radiation (RMSE around 1.2 min). The uncertainty of the filled data was finally observed to increase linearly with respect to the proportion of estimated values involved.

Acknowledgements

This study was supported by the Belgian Science Policy under the research project: Contribution de l'IRM au développement de l'énergie renouvelable en Belgique.

Appendix A.

In this appendix, we provide the formulation of the G-to-B model E82 proposed by Erbs et al. (1982) that underlies the B-to-G model (4). The model E82 estimates the beam horizontal solar radiation as a fraction of the global horizontal solar radiation, $B = \alpha G$ where α is an increasing function of the clearness index K_t,

$$\begin{cases} \alpha = 0.09K_t \quad \text{if } K_t \leqslant 0.22, \\ \alpha = 0.0489 + 0.1604K_t - 4.388K_t^2 + 16.638K_t^3 \\ \quad -12.336K_t^4 \quad \text{if } 0.22 < K_t < 0.8, \\ \alpha = 0.835 \quad \text{if } K_t \geqslant 0.8. \end{cases}$$

References

Ångström, A., 1924. Solar and terrestrial radiation. Quarterly Journal of the Royal Meteorological Society 50, 121–125.

Almorox, J., Hontoria, C., 2004. Global solar radiation estimation using sunshine duration in Spain. Energy Conversion and Management 45, 1529–1535.

Colle, S., De Abreu, S.L., Ruther, R., 2001. Uncertainty in economic analysis of solar water heating and photovoltaic systems. Solar Energy 70, 131–142.

Erbs, D.G., Klein, S.A., Duffie, J.A., 1982. Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. Solar Energy 28, 293–302.

Espinar, B., Ramírez, L., Drews, A., Beyer, H.G., Zarzalejo, L.F., Polo, J., Martín, L., 2009. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Solar Energy 83, 118–125.

European Solar Radiation Atlas (ESRA), 1984. In: Palz, W. (Ed.), Second Improved and Extended Edition, Vols. I and II, Commission of the European Communities, DG Science, Research and Development, Brussels, Report No. EUR 9344.

Geiger, M., Diabate, L., Menard, L., Wald, L., 2002. A web service for controlling the quality of measurements of global solar irradiation. Solar Energy 73, 474–480.

Gueymard, C., 2004. The sun's total and spectral irradiance for solar energy applications and solar radiation models. Solar Energy 76, 423–453.

Hay, J.E., 1993. Solar radiation data: validation and quality control. Renewable Energy 3, 349–355.

Kasten, F., 1996. The Linke turbidity factor based on improved values of the integral Rayleigh optical thickness. Solar Energy 56, 239–244.

Maxwell, E., 1987. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Report No. SERI/TR-215-3087. Solar Energy Research Institute, Golden, Colorado.

Maxwell, E., Wilcox, S., Rymes, M., 1993. Users manual for SERI QC software, assessing the quality of solar radiation data. Report No. NREL-TP-463-5608. 1617 Cole Boulevard, Golden, Colorado, National Renewable Energy Laboratory.

Molineaux, B., Ineichen, P., 1994. Automatic quality control of daylight measurements: software for IDMP stations. Technical report : guide to recommended practice of daylight measurement, Vienna, Commission internationale de l'éclairage, pp. 34–42.

Moradi, I., 2009. Quality control of global solar radiation using sunshine duration hours. Energy 34, 1–6.

Muneer, T., Fairooz, F., 2002. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. Building Services Engineering Research and Technology 23, 151–166.

Perez, R., Ineichen, P., Seals, R., Zelenka, A., 1990a. Making full use of the clearness index for parameterizing hourly insolation conditions. Solar Energy 45, 111–114.

Perez, R., Seals, R., Zelenka, A., Ineichen, P., 1990b. Climatic evaluation of models that predict hourly direct irradiance from hourly global irradiance: prospects for performance improvements. Solar Energy 44, 99–108.

Perez, R., Ineichen, P., Maxwell, E., Seals, R., Zelenka, A., 1992. Dynamic global-to-direct conversion models. ASHRAE Transactions 98, 354–369.

Prescott, J.A., 1940. Evaporation from water surface in relation to solar radiation. Transactions of the Royal Society of Australia 64, 114–125.

Remund, J., Page, J., 2002. Advanced parameters. Chain of algorithms. Part I: shortwave radiation. Report to the European Commission, SoDa project IST-1999-12245.

Rigolier, C., Bauer, O., Wald, L., 2000. On the clear sky model of the 4th European solar atlas with respect to Heliosat method. Solar Energy 68, 33–48.

Shafer, M.A., Fiebrich, C.A., Arndt, D.S., Fredrickson, S.E., Hughes, T.W., 2000. Quality assurance procedures in the Oklahoma Mesonetwork. Journal of Atmospheric and Oceanic Technology 17, 474–494.

Shi, G.Y., Hayasaka, T., Ohmura, A., Chen, Z.H., Wang, B., Zha, J.Q., Che, H.Z., Xu, L., 2008. Data quality assessment and the long-term trend of ground solar radiation in China. Journal of Applied Meteorology and Climatology 47, 1006–1016.

Skartveit, A., Olseth, J.A., 1987. A model for the diffuse fraction of hourly global radiation. Solar Energy 38, 271–274.

Tang, W., Yang, K., He, J., Qin, J., 2010. Quality control and estimation of global solar radiation in China. Solar Energy 84, 466–475.

Terzenbach, U., 1995. European Solar Radiation Atlas: Quality Control Algorithms on Solar Radiation Data. EC-Contract No. JOU2-CT94-0305, Task II-Algorithms. Deutscher Wetterdienst, Frahmredder 95, D-22393 Hamburg, Germany.

World Meteorological Organization, 2007. Guide to the global observing system, Appendix VI.2. WMO-No 448.

Yang, K., Koike, T., 2005. A general model to estimate hourly and daily solar radiation for hydrological studies. Water Resources Research 41, W10403. doi:10.1029/2005WR003976.

Yang, K., He, J., Tang, W., Qin, J., Cheng, C.C.K., 2010. On downward shortwave and longwave radiations over high altitude regions: observation and modeling in the Tibetan Plateau. Agricultural and Forest Meteorology 150, 38–46.

Younes, S., Claywell, R., Muneer, T., 2005. Quality control of solar radiation data: present status and proposed new approaches. Energy 30, 1533–1549.