Available online at www.sciencedirect.com
SCIENCE DIRECT®
Energy 30 (2005) 1533–1549
ENERGY
www.elsevier.com/locate/energy

Quality control of solar radiation data: Present status and proposed new approaches

S. Younes, R. Claywell, T. Muneer *

School of Engineering, Napier University, 10 Colinton Road, Edinburgh, EH10 5DT, UK

Abstract

During the past few decades, there has been a continual rise in interest in passive and active solar energy uses, not only in the governmental and commercial sectors, but also within the private sector. There is thus a need for taking measurements of solar irradiation and creating local and regional databases of irradiation and synoptic (meteorological) information. However, there is no guarantee of the quality of the data collected, as often due care is not exercised with respect to quality control of the measured dataset.

This article reviews the presently available procedures for quality assessment of the solar irradiation data. Furthermore, we propose a set of stringent physical and statistical measures to create a semi-automated procedure that is based on the creation of an envelope in the clearness index–diffuse to global irradiance ratio domain. The procedure is very general in nature and may be used with equal effectiveness for any terrestrial dataset.

1. Notation

I_G	global horizontal irradiation
I_D	diffuse horizontal irradiation
I_B	beam horizontal irradiation
I_E	extraterrestrial horizontal irradiation
I_{SC}	solar irradiation constant
$I_{G,C}$	global horizontal irradiation by Page clear-sky model
$I_{D,C}$	diffuse horizontal irradiation by Page clear-sky model
$I_{B,C}$	beam horizontal irradiation by Page clear-sky model
$I_{G,OC}$	global horizontal irradiation by Page overcast sky model

* Corresponding author. Tel.: +44-131-455-2541; fax: +44-131-455-2264.
E-mail address: t.muneer@napier.ac.uk (T. Muneer).

0360-5442/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.energy.2004.04.031

$I_{D,OC}$	diffuse horizontal irradiation by Page overcast sky model
DEC	earth's declination
SHA	solar hour angle
AST	apparent solar time
SOLALT	solar altitude
SOLAZM	solar azimuth
K_d	earth–sun correction factor
T_L	Linke turbidity factor
DN	Julian day number
δ_r	Rayleigh optical depth
T_{rd}	diffuse transmittance

2. Introduction

Architects, engineers and scientists involved in the field of solar energy systems require solar data measured at the vicinity of their application. It is important for active and passive users of solar energy to know the global solar energy that is available and its diffuse component. The quality of measurements of solar irradiance dataset needs to be assessed by staff with minimal training and with precision.

The patterns of the availability of solar resource, in time, are important, as this dictates the design of energy storage systems. Long-term solar irradiance measurements are available from a number of meteorological measuring stations around the globe. Close examination of the data regularly reveals problems with the data often for extended periods of time. This lack of complete datasets, free from any erroneous measurements, can be due to a combination of factors that are underlined in the proceeding text.

We had at our disposal datasets covering the northern hemisphere from Europe and Asia for the last two decades. In this article, we have analysed the quality control procedures that already exist and then we shall propose our own presently developed method using statistical and physical based tests.

3. Location of synoptic information

The dataset used for this work consists of 11 locations from the northern hemisphere and covers two continents. Out of the 11 sites, four were from Europe, two British sites: Bracknell (51.26 N; 0.46 W) and Eskdalemuir (55.32 N; 3.20 W) and two Spanish sites: Madrid (40.45 N; 3.73 W) and Gerona (41.97 N; 2.88 E). These are the most westerly locations used. We then cover Bahrain in the Middle-East (26.22 N; 50.65 E) before looking at the Indian sub-continent where we utilised data from four sites that are geographically and topographically different. The Indian sites are as follows: Chennai (13.0 N; 80.18 E), Pune (18.53 N; 73.85 E), Mumbai (19.12 N; 72.85 E) and New Delhi (28.60 N; 77.20 E). Finally, the most easterly sites we covered are from Japan: Fukuoka (33.52 N; 130.48 E) and Sapporo (43.05 N; 141.33 E).

These sites not only cover different longitudes and latitudes from the northern hemisphere but also different climates and topographies; some have semi-arid climate, others are temperate. The locations also differ by specific climatologies; the Bahrain site is affected by seasonal sand storms, while the Indian sites are affected by seasonal monsoons.

4. Equipment error, uncertainties and operation related problems

Any likely sources of errors or problems related to solar radiation measurement may be placed under the following two major categories: equipment error and uncertainty and operation related problems and errors.

With any measurement there exist errors, some of which are systematic and others inherent of the equipment employed. The most common sources of error arise from the sensors and their construction. These are broken down into the most general types of errors and are described below:

- Cosine response
- Azimuth response
- Temperature response
- Spectral selectivity
- Stability
- Non-linearity
- Shade-ring misalignment
- Dark offset (nocturnal) long-wave radiation error.

Of all the above listed errors, the cosine effect is the most apparent and widely recognised one.

This is the sensor's response to the angle at which radiation strikes the sensing area. The more acute the angle of the sun, i.e. at sunrise and sunset, the greater will be the error (at altitudes below 6°). Cosine error is typically dealt with through the exclusion of the recorded data at sunrise and sunset times.

The azimuth error is the result of imperfections of the glass domes and in the case of solarimeters, the angular reflection properties of the black paint. This is an inherent manufacturing error, which yields a similar percentage error as the cosine effect.

Like the azimuth error, the temperature response of the sensor is an individual fault for each cell. The photometers are thermostatically controlled; hence, the percentage error due to fluctuations in the sensor's temperature is reduced. The pyranometers rely on their construct, i.e. a double-glass envelope to prevent large temperature swings.

The spectral selectivity of the pyranometers is dependent on the spectral absorptance of the black paint and the spectral transmission of the glass. The overall effect contributes only a small error percentage to the measurements. Each sensor possesses a high level of stability with the deterioration of cells resulting in approximately $\pm 1\%$ change in the full-scale measurement per year. Finally, the non-linearity of the sensors is a concern especially with photometers. It is a function of illuminance or irradiance levels. It, however, tends to contribute only a small error percentage towards the measured values.

The work undertaken by the National Renewable Energy Laboratory (NREL) [1] under the US continental climate and a desert site in Saudi Arabia has shown that zero offsets of −5 to −20 W/m^2 occur in diffuse pyranometer measurements due to thermal imbalances. This error was reported for all instruments using black sensors. The offset for a black and white detector, however, was found to be insensitive to such offset errors.

In addition to the above sources of equipment-related errors, care must be taken to avoid operational errors highlighted below:

- Operation related problems and errors
- Complete or partial shade-ring misalignment
- Dust, snow, dew, water-droplets, bird droppings, etc.
- Incorrect sensor levelling
- Shading caused by building structures
- Electric fields in the vicinity of cables
- Mechanical loading of cables
- Orientation and/or improper screening of the vertical sensors from ground-reflected radiation
- Station shut-down.

The sources of operation relation errors itemised above are self-explanatory. It is a good practice to protect cables from strong electric fields such as elevator shafts. Another source of error that may arise is from cables under mechanical load (piezoelectric effects). The piezoelectric effect is the production of electrical polarisation in a material by the application of mechanical stress. Failure to protect cables from the above sources may produce 'spikes' in the data and these are shown as unusually high values of irradiance. Such errors are best highlighted via cross-plotting diffuse ratio ($k = I_D/I_G$) against clearness index ($k_t = I_G/I_E$), and a sample plot is shown in Fig. 1. Note that any consistent errors arising due to an operational

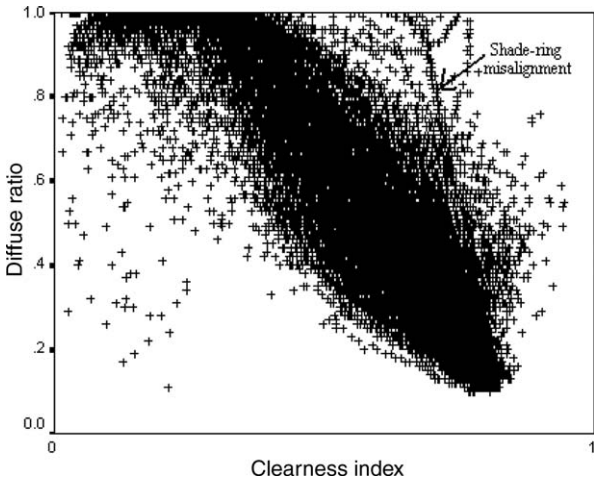

Fig. 1. Diffuse ratio–clearness index plot for Bahrain, 5-min averaged data for 28 March–30 September 2000.

problem, such as misaligned shade ring, are easily picked up by this type of plot. Any serious departure of data from the normally expected envelope is thus identified.

Stoffel et al. [1] give us a good representation of the scale of errors for carefully managed irradiance sensors. In their study, they [1] found that the range of error for a pyranometer compared with a reference pyranometer was from +2.5% to −10%, while for a pyrheliometer the range was ±2.5%.

5. Quality control procedures: present state

5.1. Page model

The Page model is based on the work undertaken for the production of the European Solar Radiation Atlas (ESRA) [2] and the CIBSE Guide on weather and solar data [3]. Page sets upper and lower boundaries for diffuse irradiation and also sets an upper boundary for global irradiation. For the former component, the overcast and clear-sky irradiance set the upper and lower limits, respectively. For the latter component, the upper limit is set by the global clear-sky model.

5.1.1. Page clear-sky model

This model computes hourly beam and diffuse irradiances under clear-sky conditions thus,

$$I_{B,C} = 1367 K_d \exp(-0.8662 m\ T_L \delta_r) \sin \mathrm{SOLALT} \tag{1}$$

$$I_{D,C} = K_d T_{rd} F(\mathrm{SOLALT}) \tag{2}$$

where $I_{B,C}$ and $I_{D,C}$ are the beam and diffuse irradiances under clear-sky conditions, respectively, and K_d is the mean earth–sun distance correction factor. The relative air mass m takes account of the presence of gases, liquids and solid particulate matter in the atmosphere.

The global irradiance $I_{G,C}$ under a clear-sky is simply the sum of the beam and diffuse components.

$$I_{G,C} = I_{B,C} + I_{D,C} \tag{3}$$

The Linke turbidity factor T_L is applied throughout the electromagnetic spectrum. Values of T_L at air mass of 2 are typically used in Page's model. T_L data are readily available on a monthly basis for many European locations [2]. The Rayleigh optical depth δ_r is an attenuation coefficient due to the Rayleigh scattering [2].

The diffuse transmittance T_{rd} is the theoretical diffuse irradiance on a horizontal surface when the sun is at the zenith. Thus,

$$T_{rd} = -21.657 + 41.752 T_L + 0.51905 T_L^2 \tag{4}$$

The solar elevation function F (SOLALT) is a polynomial function of the sine of the solar elevation,

$$F(\mathrm{SOLALT}) = x_0 + x_1\ \mathrm{sinSOLALT} + x_2\ \sin^2 \mathrm{SOLALT} \tag{5}$$

The x coefficients are given in Table 1.

Table 1
Coefficients for use in Eq. (5)

Coefficients	Sky type	
	Clear	Overcast
x_0	3.8175×10^{-2}	-6.7133×10^{-3}
x_1	1.5458	0.78600
x_2	−0.59980	0.22401

5.1.2. Page overcast sky model

Under overcast skies, global ($I_{G,OC}$) and diffuse ($I_{D,OC}$) irradiances are equal due to the absence of the beam component. Thus,

$$I_{G,OC} = I_{D,OC} = 572 \text{ SOLALT} \tag{6}$$

Ref. [4] provides further details and software for the Page model.

5.2. Helioclim quality control algorithm

Geiger et al. [5] have described the availability of a web-based service for quality control of solar radiation data. The service is available through the web site www.helioclim.net. The quality control procedure is a part of an on-going effort of the Group 'Teledetection and Modelisation' of the Ecole des Mines de Paris/Armines to provide tools and information to the solar radiation community through the world wide web. The object of that service is not to perform a precise and fine control but to perform a likelihood control of the data and to check their plausibility. This is achieved by comparing observations with expectations based upon the extraterrestrial irradiation and a simulation of the irradiation for clear skies. It offers a means to check time series of irradiation data. Inputs are provided via an HTML page by a copy and paste procedure and the return is also via similar means. Suspicious data are flagged upon return.

The user is requested to provide information to compute the quality control procedure: Geographical coordinates, elevation and dates. HTML pages are available to better understand and fill the forms for each quality control procedure. Documents explaining the algorithm used in the calculation and references to articles, web sites of interest and publications on solar radiation topics are also provided.

The quality control procedure has been divided into four HTML documents:

(a) Daily irradiation for a single day: single-value examination.
(b) Daily irradiation for several days: several daily values spread over many months and years to be analysed.
(c) Hourly irradiation for a single hour: single-value examination.
(d) Hourly irradiation for several hours: several hourly values spread over many days, months and years.

The quality algorithm used in the Helioclim web site is part of the SoDa project [6].

The Helioclim algorithm provides an interval of acceptance for hourly global irradiation data. The algorithm has been designed for locations with noon SOLALT greater than 2°.

$$\text{Upper limit} = \min(1.1 \times I_{G,C}, I_E) \tag{7}$$

$$\text{Lower limit} = 0.03 I_E \tag{8}$$

5.3. Molineaux and Ineichen's web based procedure and tools

Molineaux and Ineichen [7] describe the availability of yet another web-based facility for quality control of solar radiation data. Their computer programmes allow validation limits to be set on the tests so as to enable the user to increase the precision of the tests. The programme reads an input file based on the ASCII format and in turn creates an output file in the same format with AQC flags. Visualisation of the comparisons between measured and predicted values (based on the well established solar radiation models) is used to trace the errors. The programme carries out a series of coherence tests, which is then followed by the creation of a number of plots based on comparisons between modelled/calculated and measured values.

5.4. NREL SERI QC programme

The US-based NREL has developed alternate procedures and software for performing post-measurement quality assessment of solar radiation data. The assessments are also performed on the uncertainty of measured solar radiation data. In this respect, a quality assessment software package SERI QC [8] was developed by the NREL to address the above needs. SERI QC is based on the establishment of boundaries or limits within which acceptable data are expected to lie. This is similar to previous quality assessment procedures that use extraterrestrial values for the upper limit and zero for the lower limit within which solar radiation data were expected to lie. SERI QC increases the sophistication of the latter approach by establishing much more restrictive boundaries specific to each station month. SERI QC operates in a dimensionless space, i.e. solar radiation normalised to extraterrestrial values. An example of the expected limits and boundaries established by SERI QC is given in Fig. 1. The variables that form the abscissa and ordinate in this figure: k_B the atmospheric transmission of the direct beam radiation is defined as, $k_B = I_B/I_E$.

The hourly values plotted in Fig. 2 were the data collected by the NREL for Nashville, TN, for the period April 1978–April 1980. Established empirical limits and boundaries of acceptable data for this station are also shown within the latter figure. The heavy dashed lines represent the expected maximum global horizontal, and direct normal values and the curved boundaries around the scatter plot of the data were empirically determined by these data. This was implemented by positioning a limited set of boundary shapes around the data. The position of the boundaries was then adjusted in k_t increments of 0.025 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.

The three parts of Fig. 2 show data, maximum–minimum limits and boundaries for three different air masses. SERI QC assigns limits and boundaries for three air mass ranges

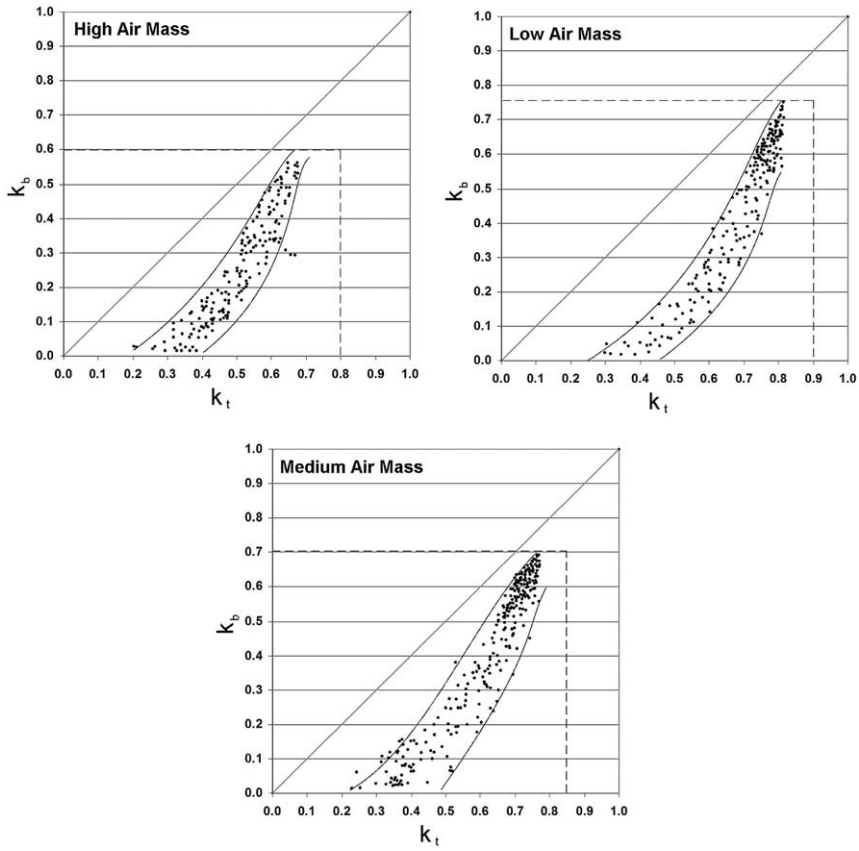

Fig. 2. Hourly beam-to-extraterrestrial irradiance plotted against clearness index (NREL's quality control procedure).

(low: 1.0–1.25, medium: 1.25–2.5 and high: 2.5–5.58). Changes in limits and boundary positions with smaller changes in air mass are not significant.

When all three solar radiation elements are available (global horizontal, direct normal and diffuse horizontal), redundancies may be used to further reduce the uncertainty of the data. This is accomplished by calculating the global from the direct normal and diffuse, and by comparing the calculated global with the measured global radiation. This comparison provides a direct indication of the accuracy of all three measurements.

Nevertheless, when hourly values of global horizontal, direct normal and diffuse horizontal radiation agree within a specified error limit, the lowest possible uncertainty for the data can be assigned. In addition to determining whether the solar radiation data fall within expected boundaries, SERI QC calculates the distance in *k*-space by which the data fall outside the boundaries.

The flagging system used by SERI QC records these distances and indicates whether one-element, two-element or three-element data were involved and whether the data point was below or above the expected boundaries. The SERI QC flags, therefore, permit the assignment of uncertainties that are dependent on the nature of the test performed (one, two or three components) and the distance by which the data point exceeds expected limits.

A point worth mentioning is that once the above filtering process has been completed and the erroneous data removed, there is a need to fill-in the 'holes' within the dataset. Unless this procedure is undertaken, the time series would be incomplete. Building energy simulation programmes in particular are prone to hick-ups with such problems. Gaps identified within the dataset may either be filled by the generation of irradiation data from other synoptic data such as sunshine or cloud cover, or data averaging techniques. In this respect, the reader is referred to the work of Muneer and Fairooz [9].

Furthermore, Rymes and Myers [10] have presented a method for smoothly interpolating averaged (coarsely resolved) data into data with a finer resolution, while preserving the deterministic mean of the data. Their technique claims to preserve the component relationship between direct, diffuse and global solar radiation when values for at least two of the components are available.

5.5. CIE automatic quality control

The Commission Internationale de l'éclairage (CIE) proposes the following quality control tests [11]. They note that automatic testing should not be performed when the solar elevation is less than 4° and when the global irradiance is less than 20 W/m^2.

They propose five levels of tests; the first two are related to global, beam and diffuse radiation and corresponding illuminance. The third test is related to the north, east, south and west global irradiance and illuminance. The fourth level test involves inter-comparisons between irradiance and illuminance and finally the fifth level test compares the zenith luminance with either diffuse irradiance or illuminance.

Herein we are interested in the first two levels of tests from the CIE. These are described in more detail below.

The first level tests are rough absolute checks that insure that no major problems exist.

$$0 < I_G < 1.2 \times I_E \tag{9}$$

$$0 < I_D < 0.8 \times I_E \tag{10}$$

$$0 < I_B < I_E \tag{11}$$

The second level tests are consistency tests that utilise the redundancy existing between direct, diffuse and global components:

$$I_G = (I_B + I_D) \pm 15\% \tag{12}$$

For stations that do not measure the direct component,

$$I_D < I_G + 10\% \tag{13}$$

Note that the 10% margin is an allowance for shade ring correction.

5.6. Muneer and Fairooz quality control procedure

The Muneer and Fairooz [9] quality control procedure is, in addition to other filters, a combination of tests based on the CIE quality control procedure (Section 5.5) and Page irradiance model (Section 5.1).

The Muneer and Fairooz model has four test levels. The first test is the CIE quality control method given by Eqs. (9)–(11).

The level two tests include consistency tests between diffuse and global irradiation, and between global and horizontal extraterrestrial irradiation. The third level tests are based on an expected diffuse ratio (I_D/I_G)–clearness index (I_G/I_E) envelope. This test involves a check that the diffuse irradiation data conform to the limits set out by an envelope of acceptance. A further fourth level check on the quality of diffuse irradiance is undertaken by comparing its value with the diffuse irradiance under the two extreme conditions, as defined by Page's clear and overcast sky model [2].

As a final (fifth-level) measure of check on global and diffuse irradiance data, turbidity is calculated for the given time series and checked for its limits. A Linke turbidity value that is less than 2.5 (obtained under exceptionally clear skies) or greater than 12 (under dust storm conditions) would demand close inspection of data.

6. Proposed procedure

The code for the presently proposed procedure was written in FORTRAN to process the databases available. Geographical information for the site is required, such as site elevation above sea level, latitude, longitude and local time meridian. Also, logging related information is required, either solar time or local civil time is accepted.

Operation related information is also required such as desired standard deviation, interval number and envelope cut-off point. Details regarding this type of information are provided later.

Before starting to test the data for its validity, solar position calculations are performed for each data entry. These are described by Muneer [4] and are listed below as per order of occurrence. Calculation of solar hour angle (SHA), apparent solar time (AST), declination angle (DEC), solar altitude (SOLALT) and finally the calculation of solar azimuth (SOLAZM) is performed by the code.

6.1. First test

The calculation then proceeds by eliminating entries that show a SOLALT less than 7°. For entries that have passed the first test, the day number (DN), the horizontal extraterrestrial irradiation and finally the clearness index (k_t) and the diffuse ratio (k) are calculated.

6.2. Second test

This test is a logical test, as the clearness index and the diffuse ratio are always positive and have values between zero and one.

$$0 < k_t < 1 \text{ and } 0 < k < 1 \tag{14}$$

6.3. Third test

At this stage, global and diffuse irradiation are compared with their corresponding Page-model upper and lower boundaries (Section 5.1). The global horizontal irradiation ought to be

less than or equal to the clear day global horizontal irradiation. Thus,

$$I_G \leq I_{G,C} \tag{15}$$

From Muneer and Fairooz (Section 5.6), it is proposed to test whether the diffuse horizontal irradiation is sandwiched between the clear day diffuse and the overcast day horizontal irradiation as defined by Page. Thus,

$$I_{D,OC} \leq I_D \leq I_{D,C} \tag{16}$$

Fig. 3 presents a flow chart for the presently proposed procedure.

6.4. Fourth test

The presently proposed construction of $k - k$t quality control envelope is a statistical procedure that requires estimation of k_t banded mean, weighted mean, ($\bar{k}$) and standard deviations of k values (σ_k). Typically, the k_t range of data may be divided into, say, 10 bands of equal width. For each band, the above-mentioned statistics is obtained. From this information, an envelope may be drawn that connects those points that represent the top ($\bar{k} + 2\sigma_k$) and bottom ($\bar{k} - 2\sigma_k$) curves, respectively. Fig. 4 shows such a plot, which is in effect a loose quality control envelope of the raw data presented in Fig. 1.

Hand drawn envelopes developed from visual inspection of the datasets such as those presented in Fig. 1 were used by Muneer and Fairooz [9] to clean datasets for clearness index–diffuse ratio scatter plots. The visual type of boundary that is hand drawn is herein referred to as envelope of expectancy, or the quality envelope.

The coordinates defining the envelope are then noted thus producing constraints for data filtering, i.e. the upper and lower bounds of acceptability of k for any given k_t are used.

Another technique that is employed by statisticians to identify erroneous data is the outlier analysis. Note that 'outlier' is a term that indicates an abnormality, and suggests that the datum is not typical of the rest of the data. As a rule, an outlier should be subjected to careful examination to identify logical explanations for its unusual behaviour. Outliers may, however, be rejected if the associated errors can be traced to erroneous observations, due to any one or a number of factors described above. Statistically, a 'near-outlier' is an observation that lies outside 1.5 times the inter-quartile range. The inter-quartile is the interval from the 1st quartile to the 3rd quartile. The near-outlier limits are mathematically defined by:

$$\text{Lower outlier limit} = \text{1st quartile} - 1.5 \times (\text{3rd quartile} - \text{1st quartile}) \tag{17}$$

$$\text{Upper outlier limit} = \text{3rd quartile} + 1.5 \times (\text{3rd quartile} - \text{1st quartile}) \tag{18}$$

A high number of outliers in the dataset signify that the observations have a high degree of variability or a large set of suspect data indicating poor station operation.

Note that 'far outliers' are those for which the factor of 1.5 used within Eqs. 17 and 18 is replaced with a value of 3.

Outlier analysis is, however, very computation intensive as it involves ordering of large datasets in an ascending order. The presently proposed technique based on standard deviation is much more economical in terms of the CPU time. The software that was developed under the

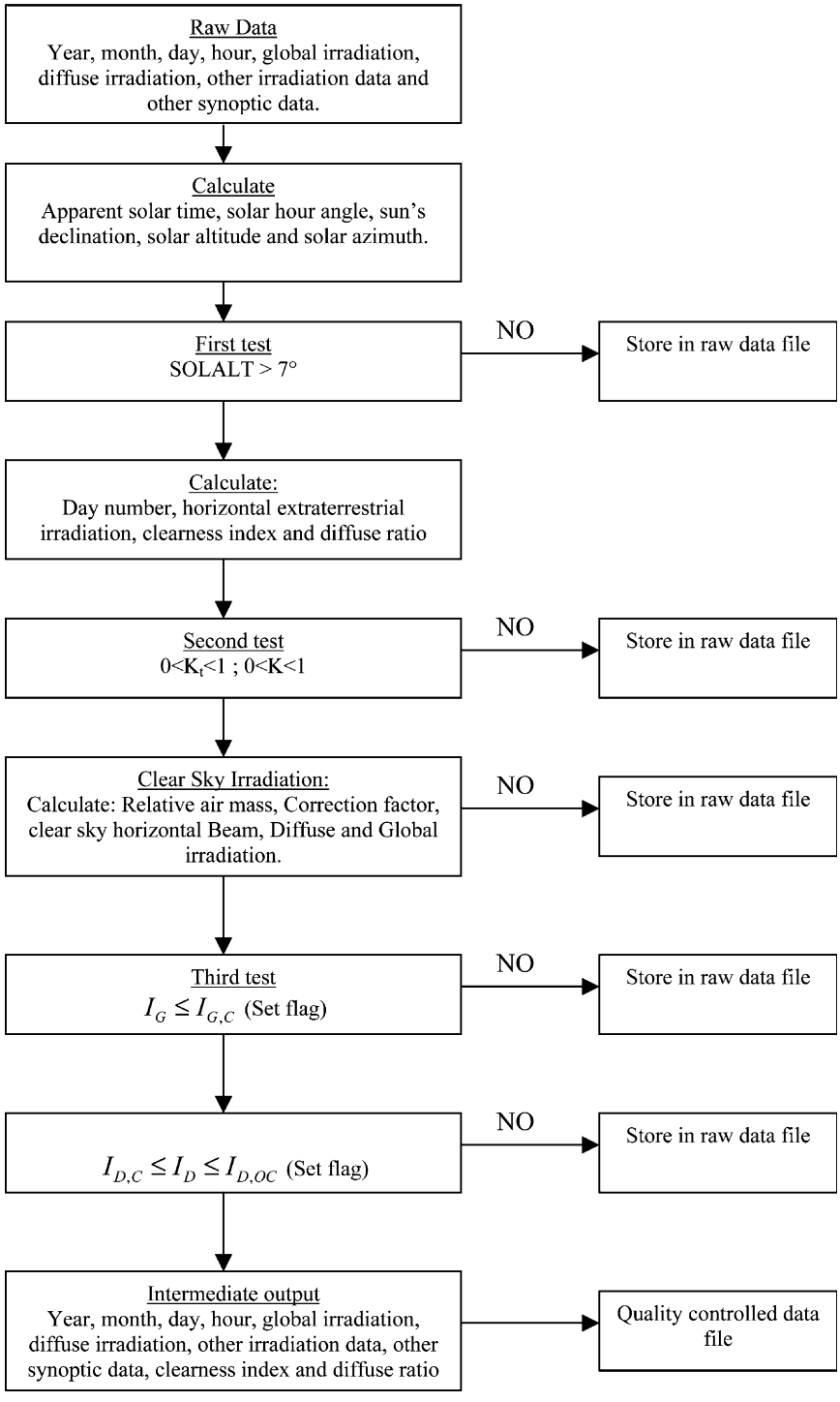

Fig. 3. Flow chart for processing raw solar irradiation data via presently proposed procedure.

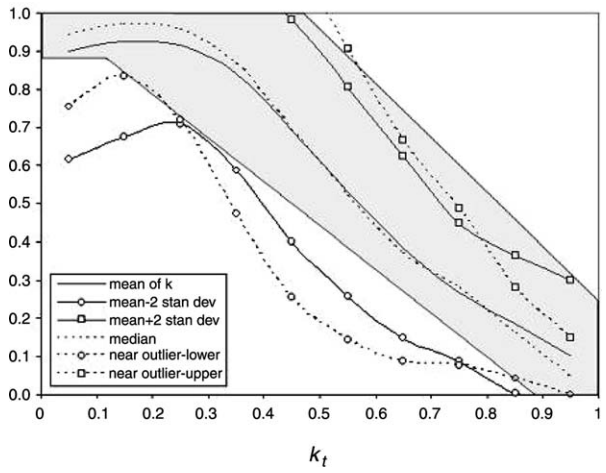

Fig. 4. Comparison between standard deviation and quartile analysis.

present research programme, however, produces both the near-outlier and standard deviation based envelopes. These are included in Fig. 4.

Note that the 'crude' hand-drawn envelope based purely on a visual observation is shown as a grey area.

The above described $k - k_t$ scatter plot procedure was undertaken for data from 11 sites (see Section 3). It was found that the $\bar{k} \pm 2\sigma_k$ provides the most optimum envelope. Note that $\bar{k}$ is the weighted mean of k values within a given k_t band. It also outperforms the quartile analysis method in 90% of cases, i.e. in terms of the restrictiveness of data inclusion.

Once the envelope constituted by the upper and lower boundaries is identified, it is possible to fit a polynomial for a mathematical description of the envelope of acceptance. A second degree

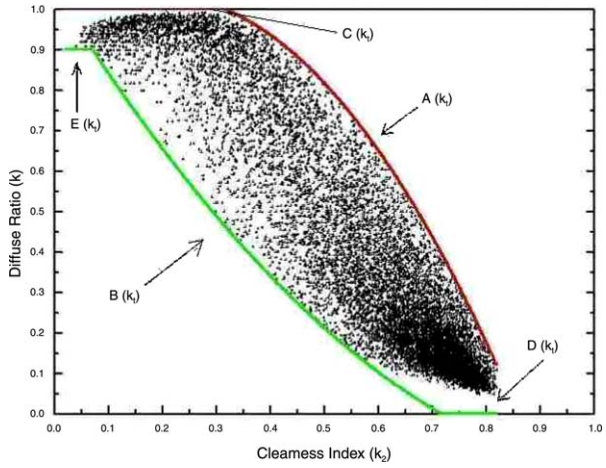

Fig. 5. Picture shows the boundary equations and functions for a typical database analysis, in this case Madrid. Madrid-Quality Controlled ($\bar{k} + 2\sigma_k$).

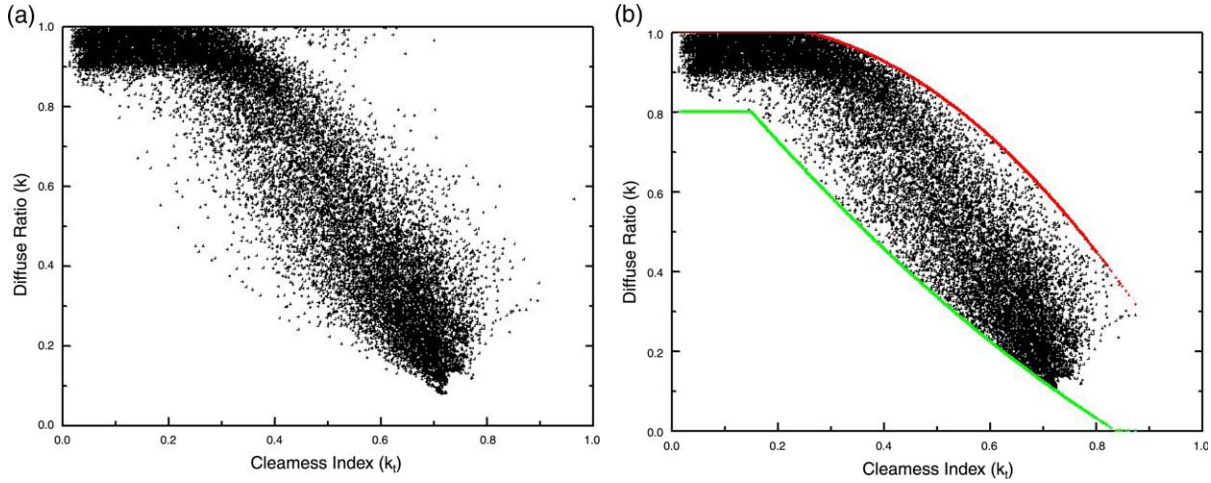

Fig. 6. Scatter plot for Fukuoka. Note that $\bar{k} \pm 2.0\sigma_k$ envelope contains 97.5% of the data that passed the third test (see Section 6.4), a) Raw data; b) Quality controller ($\bar{k} \pm 2.0\sigma_k$).

polynomial was found to be adequate. Thus, the upper and lower boundaries are represented as, respectively

$$\mathrm{A}\ (k_t) = \max\left(1, a_1 k_t^2 + b_1 k_t + c_1\right) \tag{19}$$

$$\mathrm{B}\ (k_t) = \min\left(0, a_2 k_t^2 + b_2 k_t + c_2\right) \tag{20}$$

Note that any given polynomial may generate data that can go beyond the physical limits of k, which lie between 0 and 1. The formulation given in Eqs. (19) and (20) satisfies the above

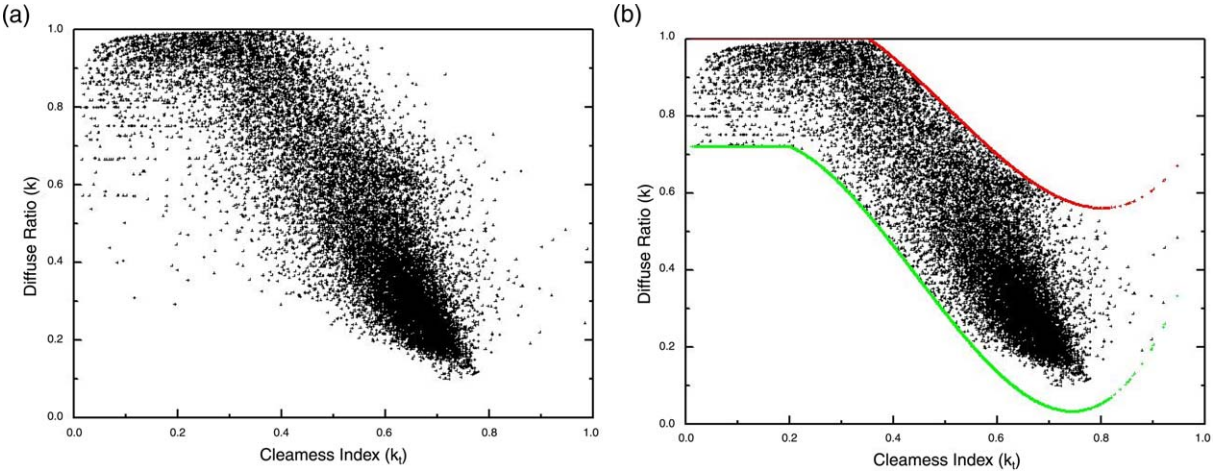

Fig. 7. Scatter plot for Mumbai. Note that $\bar{k} \pm 2.0\sigma_k$ envelope contains 94.62% of the data that passed the third test (see Section 6.4), a) Raw data; b) Quality controller ($\bar{k} \pm 2.0\sigma_k$).

constraints. Furthermore, due care has to be taken to incorporate the 'shoulder' effect caused by the respective intersection of the upper and lower polynomials with the $k = 1$ (upper) and $k = 0$ (lower) limits for the plot. By visual inspection of the plot (Fig. 6), it is possible to ascertain the latter points of intersection. The procedure of quality control can now be completed with the envelope of acceptance fully defined.

The presently proposed quality control procedure is therefore semi-automatic, as the user has to select the cut-off points by a visual inspection of the trend of the upper- and lower-bound polynomials. This is the only visual part of the process. For the 11 locations, we have worked with the cut-off point that has been below $k_t = 0.4$ (upper bound), and between $k_t = 0.75$ and $k_t = 1$ (lower bound).

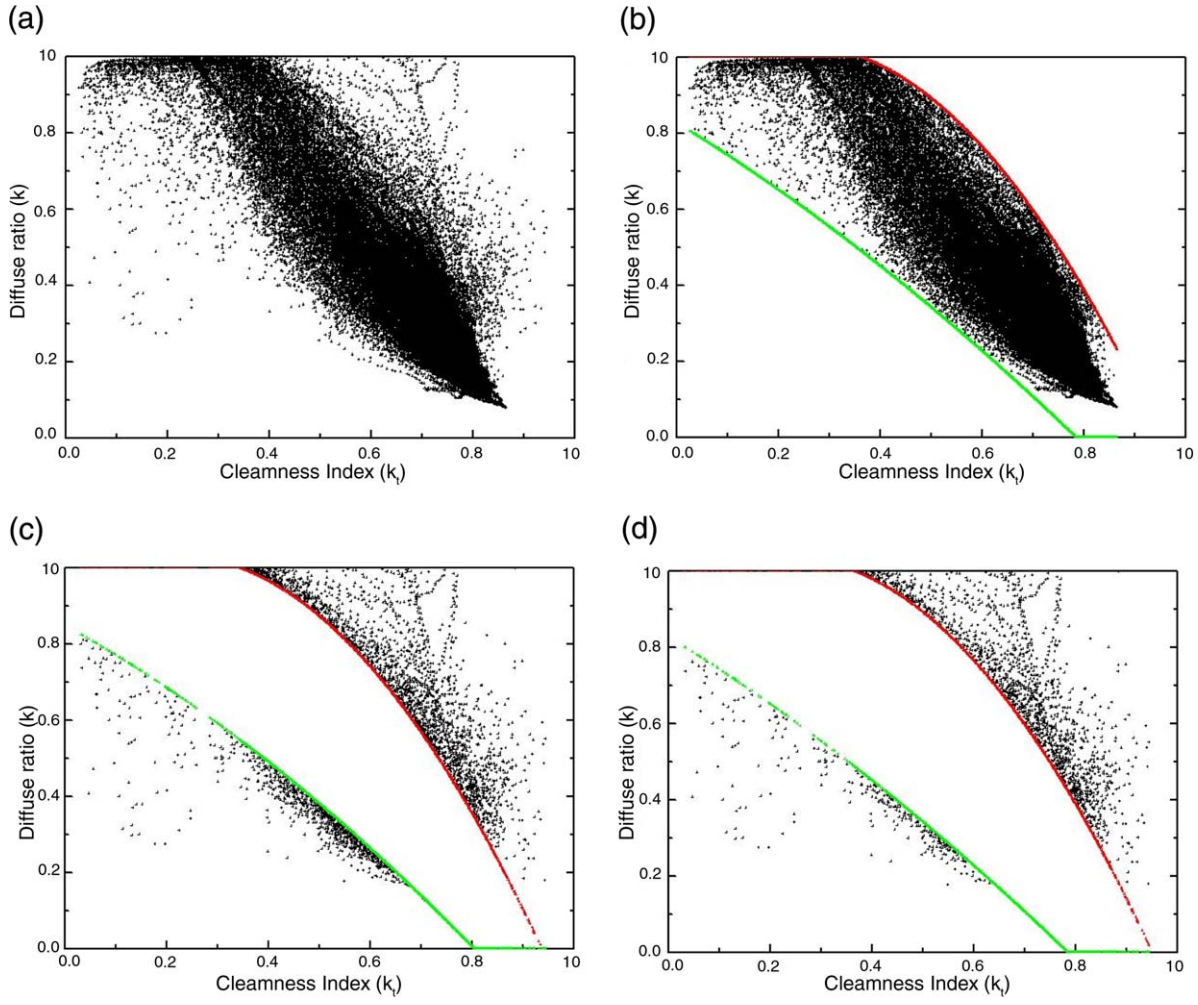

Fig. 8. Scatter plots for Bahrain. a) Raw data; b) Quality controlled ($\bar{k} \pm 2.3\sigma_k$); note that $\bar{k} \pm 2.0\sigma_k$ envelope contains 97.54% of the data that passed the third test (see Section 6.4), c) Rejected data based on ($\bar{k} \pm 2.0\sigma_k$) envelope; d) Rejected data based on ($\bar{k} \pm 2.3\sigma_k$) envelope.

Fig. 5 presents the latterly mentioned upper (C (k_t)) and lower (D (k_t)) lines of intersection. The last item to be mentioned in this context is that in certain cases there may be a need for the control of the lower-bound polynomial with respect to its upper limit. Note that within Fig. 5, an unconstrained 'flow' of the B (k_t) curve would exclude a small proportion of data belonging to heavy overcast regime. A cut-off shown as E (k_t) line is thus required, once again by visual inspection.

Fig. 3 presents the presently proposed sequential quality control procedure. This procedure was applied to the present dataset from 11 locations. Sample plots of raw, filtered and rejected data are presented in Figs. 5–8.

While Fig. 6 is self-explanatory, attention is drawn towards Figs. 5 and 7. The latter plot reveals that the shade ring correction factor has not been applied to diffuse irradiance measurements as is evident via examination of the top left-hand corner of the plot. Note that the data in the corner would be expected to attain the limiting value of $k = 1$ as $k_t \to 0$. This was later confirmed by the provider of the dataset. Fig. 8 also reveals that problems exist with respect to shade ring correction. Note that reference [12] provides a literature survey and assessment of diffuse shade ring irradiance correction procedure. Fig. 8 also demonstrates that the optimum envelope of acceptance for this particular location is $\bar{k} \pm 2.3\sigma_k$. This is evident via a comparison of Fig. 8(c) and (d). Note that the former plot rejects an excessive amount of data using a tighter ($\bar{k} \pm 2\sigma_k$) envelope. Within the context of the present study, it was found that the envelope of acceptance optima lay between the ($\bar{k} \pm 2\sigma_k$) and ($\bar{k} \pm 2.3\sigma_k$) limits.

7. Conclusion

A new procedure has been developed for efficiently assessing the quality of large solar irradiation datasets. The procedure in nature requires minimal input from the user. The procedure is based on physical and statistical tests. With minimum training provided to the staff engaged in the quality control of such data, important information such as the diffuse quality of shade ring corrections may be identified. Although presently semi-automated in nature, the procedure has the potential to become fully automated.

References

[1] Stoffel T, Redo I, Myers D, Renne D, Wilcox S, Treadwell J. Current issues in terrestrial solar radiation instrumentation for energy, climate and space applications. Metrologia 2000;37(5):399–402.
[2] Greif J, Scharmer K. European solar radiation atlas, 4th ed.. France: Presses de l'Ecole, Ecole des mines de Paris; 2000.
[3] Page J. Proposed quality control procedures for the Meteorological Office data tapes relating to global solar radiation, diffuse solar radiation, sunshine and cloud in the UK. Report submitted to CIBSE Guide Solar Data Task Group. 222 Balham High Road, London, UK: Chartered Institute of Building Services Engineers; 1997.
[4] Muneer T. Solar radiation and daylight models. Oxford: Elsevier; 2004.
[5] Geiger M, Diabate L, Menard L, Wald L. A web service for controlling the quality of measurements of global solar irradiation. Solar Energy 2002;73(6):475–80.
[6] Wald L. The Project SoDa for Solar Energy and Radiation, 2004. Sophia Antipolis, France: Ecole des Mines de Paris, Centre d'Energetique, Groupe Télédétection et Modélisation; 2004. Available from: http://www.soda-is.org.

[7] Molineaux B, Ineichen P. Automatic quality control of daylight measurement: software for IDMP stations. Vaulx en Velin, France: International Daylight Measurement Programme, Ecole National des Travaux Publics de l'Etat; 2003. Available from: http://idmp.ente.fr.
[8] Maxwell E, Wilcox S, Rymes M. 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; 1993.
[9] Muneer T, Fairooz F. Quality control of solar radiation and sunshine measurements—lessons learnt from processing worldwide databases. Building Services Engineering Research and Technology 2002;23(3):151–66.
[10] Rymes M, Myers D. Mean preserving algorithm for smoothly interpolating averaged data. Solar Energy 2001;71(4):225–31.
[11] Kendrick D, et al. Guide to recommended practice of daylight measurement. Report no. CIE-108. Wein, Austria: International Commission on Illumination (CIE); 1994.
[12] Muneer T, Zhang X. A new method for correcting shadow band diffuse irradiance data. Journal of Solar Energy Engineering 2002;124:34–43.