Quality Control for Solar Irradiance Data

Dazhi Yang*, Gokhan Mert Yagli† and Hao Quan‡
*Singapore Institute of Manufacturing Technology, Agency for Science, Technology and Research (A*STAR), Singapore
Email: yangdazhi.nus@gmail.com
†Department of Electrical and Computer Engineering, National University of Singapore, Singapore
Email: mert.yagli@u.nus.edu
‡Experimental Power Grid Centre, Agency for Science, Technology and Research (A*STAR), Singapore
Email: quanh@epgc.a-star.edu.sg

***Abstract*—With the advent of sensing technology, high-resolution data have become ubiquitous among various smart grid applications. Whereas the data provide valuable empirical evidence, they are often, if not always, subject to significant measurement uncertainty. However, nor is there any standardized procedure to quality control (QC) these data. In this regard, this paper investigates the well-known dataset from the Baseline Solar Radiation Network (BSRN). BSRN archives (mostly) 1-min solar irradiance data from 60 stations around the world for over two decades. A six-step QC procedure is demonstrated using the BSRN data.**

***Keywords*-BSRN; quality control; solar irradiance**

I. Introduction

With the increasing penetration of solar energy into the electricity grid, the adoption of real-time monitoring system at the photovoltaic (PV) system installation location has become universal. Data collected by these monitoring systems can be used for forecasting, to mitigate the reliability issues during grid integration. The monitoring systems also provide valuable feedbacks on the systems' health and degradation. Among the various parameters being monitored, such as voltage, current, power, temperature or humidity, solar irradiance is *the* most important parameter, because it provides the "reference" for many secondary (calculated) parameters, such as the performance ratio or system yield. Therefore, the quality of the solar irradiance measurements need to be thoroughly checked before commencing any data analysis.

In the literature, quality control (QC) procedure for solar irradiance and PV data has always been diversified. The most noticeable reason is perhaps the lack of "true" reference—how do we know that an erroneous data point is in fact erroneous? In this regard, the obvious and reasonable approach is to contrast the measurements made by co-located sensors. In the case of PV system data, cross-system QC is gaining popularity [1], [2]. However, for the case of irradiance monitoring, the co-located measurements are not always available. Whereas research institutes may have the luxury of having multiple co-located irradiance sensors, commercially, it is common to have one irradiance sensor per PV system. Therefore, the QC procedure for irradiance data often has to rely on a set of physical limits (see below). This paper consolidates some of these physical limits of solar irradiance into six steps, and presents them to the smart grid community at this interdisciplinary conference.

II. Overview of irradiance measurement and the BSRN archive

A. Preliminaries of irradiance measurement

Solar irradiance measurement revolves around three components, namely, the global horizontal irradiance (GHI), diffuse horizontal irradiance (DIF) and direct normal irradiance (DNI). GHI is the total solar irradiance received on a horizontal (to ground) surface; DIF is the solar irradiance received on a horizontal surface that is due to scattering from molecules, aerosols, and clouds in the atmosphere; and DNI is the solar irradiance received on a plane that is normal to the sun ray. These three components are related by the *closure equation*:

$$\mathrm{GHI} = \mathrm{DIF} + \mathrm{DNI}\cos Z, \tag{1}$$

where Z denotes the zenith angle—the angle between the sun ray and the ground normal—which can be calculated with very small errors (the solar positioning algorithm can be found in Fortran, C, R, Python, and many other languages). The closure equation suggests that if the measurement of two of these components were perfect, the third component can be calculated deterministically. Note that one should not mix up DNI with beam irradiance (BI)—also known as the direct irradiance—which is the direct irradiance component received on a horizontal surface, i.e., $\mathrm{BI} = \mathrm{DNI}\cos Z$. Apart from these, other frequently-encountered irradiances include the extraterrestrial DNI (E_{0n}), extraterrestrial irradiance ($E_0 = E_{0n}\cos Z$), clear-sky irradiance (CSI), and various irradiance components on titled surfaces, see [2], [3]. All of the above-mentioned irradiance components have the unit of W/m^2.

978-1-5386-4291-7/18/$31.00 ©2018 IEEE

Solar irradiance measurements come from two complementary sources, namely, remote sensing and ground-based instruments [4]. Remote-sensing-based measurements utilize satellite-to-irradiance models, such as the intricate—and probably the best—Perez model [5], which usually have an hourly resolution. Due to the joint coverage of the earth's surface by various geostationary satellites, irradiance measurements can be obtained for any place in the world, except for the high-latitude regions where satellite images do not resolve. The ground-based instruments, on the other hand, mostly refers to radiometers, with the exception of the more recent camera-based measurement devices, e.g., see [6]. They are location-specific, but are able to measure solar irradiance at high resolution (with a μsec time response).

GHI is measured using a device called pyranometer. There are two types of pyranometers: (1) thermopile pyranometers, which make measurements based on the *thermoelectric effect*; and (2) photodiode-based pyranometers, which make measurements using silicon photodiodes similar to solar cells. DNI is measured using pryheliometers, which are mostly thermopile-type. Pryheliometers are mounted on a tracker, so that the device is always pointing directly to the sun. Lastly, DIF is measured using a pyranometer together with a shadowband or a shading ball. The purpose of the shadowband or shading ball is to shade the pyranometer from the direct sun ray, and thus only allows the device to measure the diffuse component. These shading devices are often attached to the same sun tracker that supports the pryheliometers. Today, one can purchase off-the-shelf radiometers at a variety of prices, from a few hundred dollars to more than ten thousand dollars. The measurement uncertainty of the radiometers largely correlates with their price. In the recent paper by Gueymard and Myers [7], different settings of measuring GHI, DNI and DIF were studied using various co-located radiometers. It was found that the *optimal* way to obtain GHI is to use the closure equation, with DIF and DNI measured with a Kipp & Zonen CM22 pyranometer (with shading ball) and a Kipp & Zonen CH1 pyrheliometer, respectively. For the detail of radiometry measurement devices, the reader is referred to the book by Vignola *et al.* [4] and the technical report by Sengupta *et al.* [8].

B. BSRN data

The Baseline Solar Radiation Network (BSRN) [9] is the central archive under the World Radiation Monitoring Center (WRMC). It contains (mostly) 1-min irradiance and other meteorological measurements from 60 stations around the world, covering a variety of climate types, such as arid, tropical, temperate or high albedo. The data for each station—in the form of monthly GNU-ziped station-to-archive files—are submitted by the respective station scientists after some preliminary quality control. These data are freely accessible via the ftp-server: ftp.bsrn.awi.de, or via PANGAEA, a data publisher for earth and environmental science. The data availability—in terms of the number of available monthly files—at each station varies. The measurements at several stations started since 1992, whereas some new station only contains three years of data. The status of station-to-archive files can be checked at https://dataportals.pangaea.de/bsrn/.

The measurements of all three irradiance components, namely, GHI, DHI and DIF, are archived at BSRN. At each station, DIF is measured with a shading ball attached to the sun tracker also supporting the pyrheliometer measuring DNI. GHI is measured with a separate thermopile pyranometer.

III. Quality control

The quality-control procedure for solar irradiance data proposed in this paper has two types: (1) basic QC for high-resolution data; and (2) QC for aggregated data. Basic QC aims at checking the quality of individual high-resolution data points. These QC tests are relatively loose by default, but can be tightened, subject to the user's needs. Whereas the quality-checked high-resolution data can still be erroneous—false cloud-enhancement events in particular—they need to be aggregated and undergo further checks. Throughout this section, 1-year data (year 2016) from one of the BSRN stations, namely, Boulder station (BOS), United States, are used to exemplify the QC procedure.

A. Basic QC for high-resolution data

Although the station-to-archive files can ensure basic data integrity, WRMC recommends all users to perform their own quality checks of the data. In this regard, WRMC offers a tool called BSRN-Toolbox that helps users to perform quality control. Users can download station-to-archive files using the BSRN-Toolbox. After the data quality is checked, the spurious data are flagged. Subsequently, the raw data are combined with these flags, and saved as tab-separated txt files. This QC step follows the BSRN Global Network recommended QC tests, V2.0 by Long and Dutton [10]. The details are listed below:

1) Physically-possible limits
   - $-4 <$ GHI $< 1.5 E_{0n} \cos^{1.2} Z + 100$
   - $-4 <$ DIF $< 0.95 E_{0n} \cos^{1.2} Z + 50$
   - $-4 <$ DNI $< E_{0n}$
2) Extremely-rare limits
   - $-2 <$ GHI $< 1.2 E_{0n} \cos^{1.2} Z + 50$
   - $-2 <$ DIF $< 0.75 E_{0n} \cos^{1.2} Z + 30$
   - $-2 <$ DNI $< 0.95 E_{0n} \cos^{0.2} Z + 10$

3) Comparisons
   - abs(Closr) $< 8\%$ for $Z < 75^\circ$ and $GHI > 50$
   - abs(Closr) $< 15\%$ for $93^\circ > Z > 75^\circ$ and $GHI > 50$
   - DIF/GHI < 1.05 for $Z < 75^\circ$ and $GHI > 50$
   - DIF/GHI < 1.10 for $Z > 75^\circ$ and $GHI > 50$

In the above list, E_{0n} is the extraterrestrial irradiance on a normal surface, and Closr denotes the closure error in percent, i.e., $100(\text{DIF}+\text{DNI}\cos Z - \text{GHI})/\text{GHI}$, c.f., Eq. (1).

In QC step 1)—denoted as QC1 hereafter—the individual data points are checked against the corresponding physical limits. The limits are modeled as a function of zenith angle and extraterrestrial irradiance. Since the physical limits vary geographically, the parameters are established for the entire globe. Data points violating these rules are flagged and rejected. Similarly to QC1, QC2 configures a smaller bound by tightening the parameters. Data points which pass QC1 but fail this test are physically possible yet extremely rare. They are flagged and kept, so that the user can decide whether to use them. Lastly, the tests in QC3 aim at identifying erroneous data points by comparing two or more irradiance-component measurements. These tests are non-definitive, i.e., the particular component that causes the test to fail is unknown after the test.

Using the data from the BOS station, QC1 and QC2 are visualized in Fig. 1. The physically-possible limits are drawn in red, where as the extremely-rare limits are drawn in blue. The width variations of these limit lines are caused by the changes in earth–sun distance through the year. Based on a visual inspection, these limits seem loose, however, they can be fine-tuned based on the long-term data from the station, see QC4 below. After counting, it was found that a total of 2164 out of 526,448 1-min data points were rejected after QC2. The results for QC3 are shown in Fig. 2. In this case, a total of 3420 data points failed the closure equation test, and 1095 data points failed the diffuse ratio test. It is observed that most of the closure-equation-test failures occur under low-irradiance-high-zenith-angle conditions.

Apart from QC1, QC2, and QC3, Long and Shi [11] extended the basic QC to also consider climatological limits and comparisons:

4) Climatological limits
   - $0 <$ GHI $< C_1 E_{0n} \cos^{1.2} Z + 50$
   - $0 <$ DIF $< C_2 E_{0n} \cos^{1.2} Z + 30$
   - $0 <$ DNI $< C_3 E_{0n} \cos^{0.2} Z + 10$

5) Climatological comparisons
   - DIF/GHI < 0.85 for GHI/Clr > 0.85 and DIF > 50
   - DIF $> R_L - 1.0$ for DIF/GHI < 0.8 and GHI > 50, where $R_L = 209.3\mu_0 - 708.3\mu_0^2 + 1128.7\mu_0^3 - 911.2\mu_0^4 + 287.85\mu_0^5 + 0.046725\mu_0\text{Prs}$

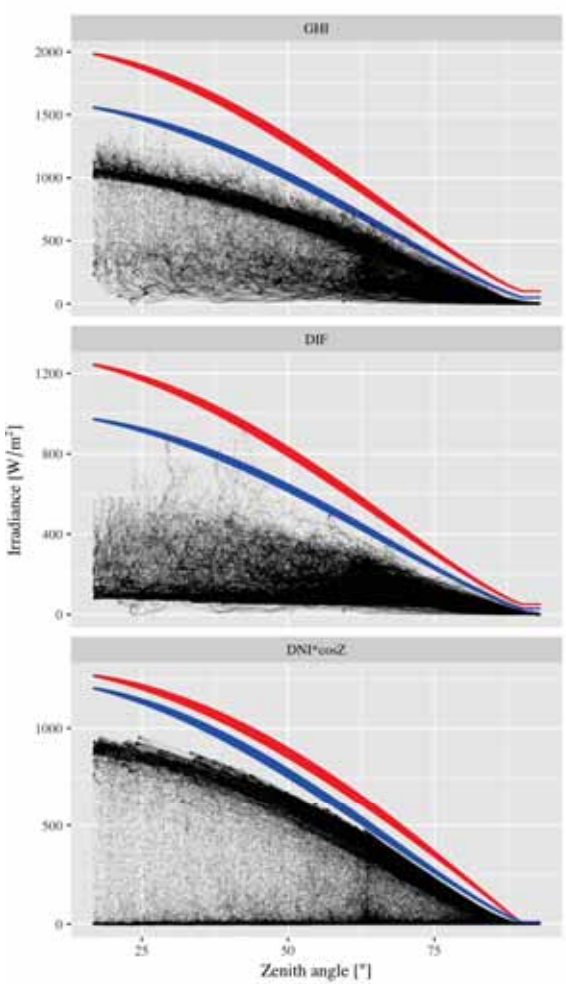

Figure 1. QC steps 1) and 2) for 1-min (top) GHI, (middle) DIF and (bottom) DNI measurements. The measurements are from the Boulder station over the year 2016. The red lines indicate the physically-possible limits, where as the blue lines show the extremely rare limits.

In the above list, C_1, C_2 and C_3 are station-specific parameters similar to—but usually tighter than—those in QC2; R_L is the Rayleigh limit; Prs is the surface station pressure in millibars; and Clr denotes the clear-sky GHI values. The clear-sky GHI is the irradiance received on a horizontal surface in a cloud-free situation. It can be calculated using various clear-sky models, such as the Ineichen–Perez model [12], or the REST2 model [13].

In QC4, the extremely rare limits in QC2 are re-configured for a particular location. This is usually done by analyzing the long-term measurements from that location. A simple approach to establish these station-specific parameters is to gradually reduce the values of C_i—lowering the limit curves—until a sudden increase in rejection rate is observed (indicating that the limit curves are stating to

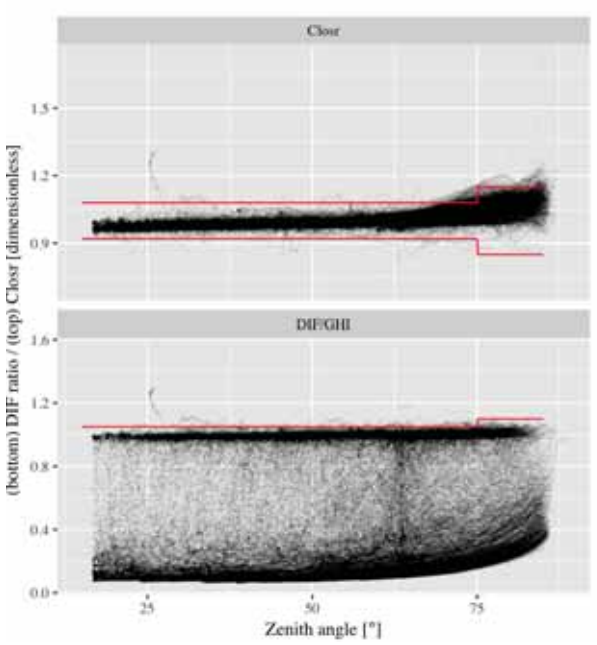

Figure 2. QC step 3) for 1-min comparison tests—(top) comparison based on the closure equation, and (bottom) comparison based on the diffuse ratio. The red lines indicate the rejection bounds. Only data points that have GHI > 50 are plotted.

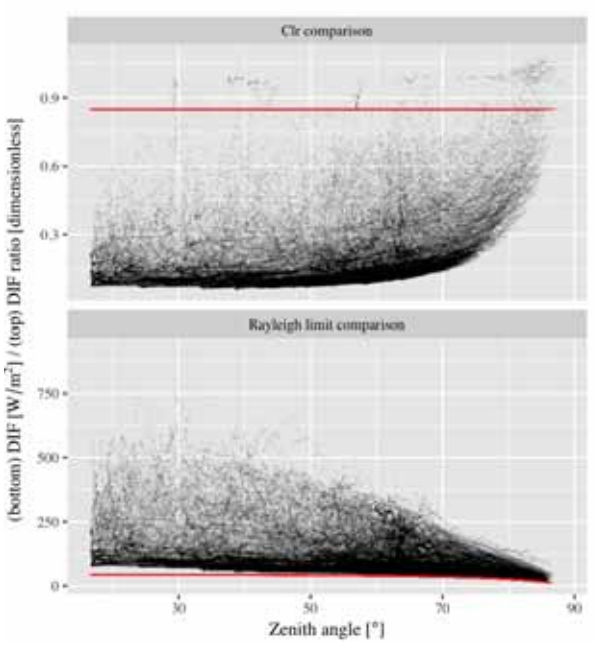

Figure 3. QC step 5) for 1-min comparison tests—(top) comparison based on the clear-sky conditions (for data points that have GHI/Clr > 0.85 and DIF > 50) and (bottom) comparison based on the Rayleigh limit (for data points that have DIF/GHI < 0.8 and GHI > 50).

"overlap" with the good data).

QC5 has two parts. The first part leverages on the contradictions of DIF and GHI during clear-sky situations. More specifically, if GHI/Clr > 0.85, it can be assumed that the sky is clear. When the sky is clear, DIF cannot be high, i.e., DIF/GHI < 0.85. The second part of QC5 describes the clear-sky DIF produced by the Rayleigh scattering only, i.e., no DIF measurement—usually subjective to additional scattering—should fall below the Rayleigh limit.

The results of QC5 using the BOS data are plotted in Fig. 3. The Ineichen–Perez model was used to generate Clr. A total of 1060 1-min data points were rejected by the clear-sky test, and 19 were rejected by the Rayleigh-limit test. It is noted that the rejection rate for BOS is quite high. When the same QC procedure is applied to the 2016 data from the Carpentras station (CAR), France, the numbers of rejected data points are 0, 94, and 580, for QC2, QC3, and QC5, respectively.

B. QC for aggregated data

The temporal resolution of solar irradiance data has a great influence when deciding what data can be considered reasonable [11]. For high-resolution data, GHI can occasionally exceed Clr by up to 50% owing to *cloud enhancement* [1]. The high GHI under cloud enhancement is caused by the simultaneous exposure of high DIF and high DNI. When the pyranometer is directly exposed to the sun (high DNI), the scattering effects of the clouds outside the solar disk may contribute a significant amount of DIF, resulting in high GHI. However, the cloud-enhancement events usually last around 20–140 s [14], and thus not seen in low-resolution data. Based on these facts, by applying QC to moderately-aggregated data, such as 5-min averages, cloud-enhancement events can be distinguished from the erroneous data.

To detect the cloud-enhancement events, a QC test similar to [1] is used here.

6) Cloud enhancement detection
   - sdev($\text{GHI}_{t-6\text{min}}$, GHI_t, $\text{GHI}_{t+6\text{min}}$) > 0.05 for GHI/Clr > 1.10 and $Z < 75°$
   - At least three valid 1-min data points are required for every aggregated 6-min data point

In the above QC step, sdev is the standard deviation operator. The 6-min aggregation interval—instead of the 5-min interval used in [1]—is used here because some BSRN stations only record 3-min-averaged irradiance prior to the year 2009. Furthermore, it is important to note that during data aggregation, lacking valid data points may cause inaccurate low-resolution data. Therefore, for every aggregated 6-min data point, at least three 1-min data points (or one 3-min data point) are required.

Following QC6, the 6-min-aggregated GHI data (red

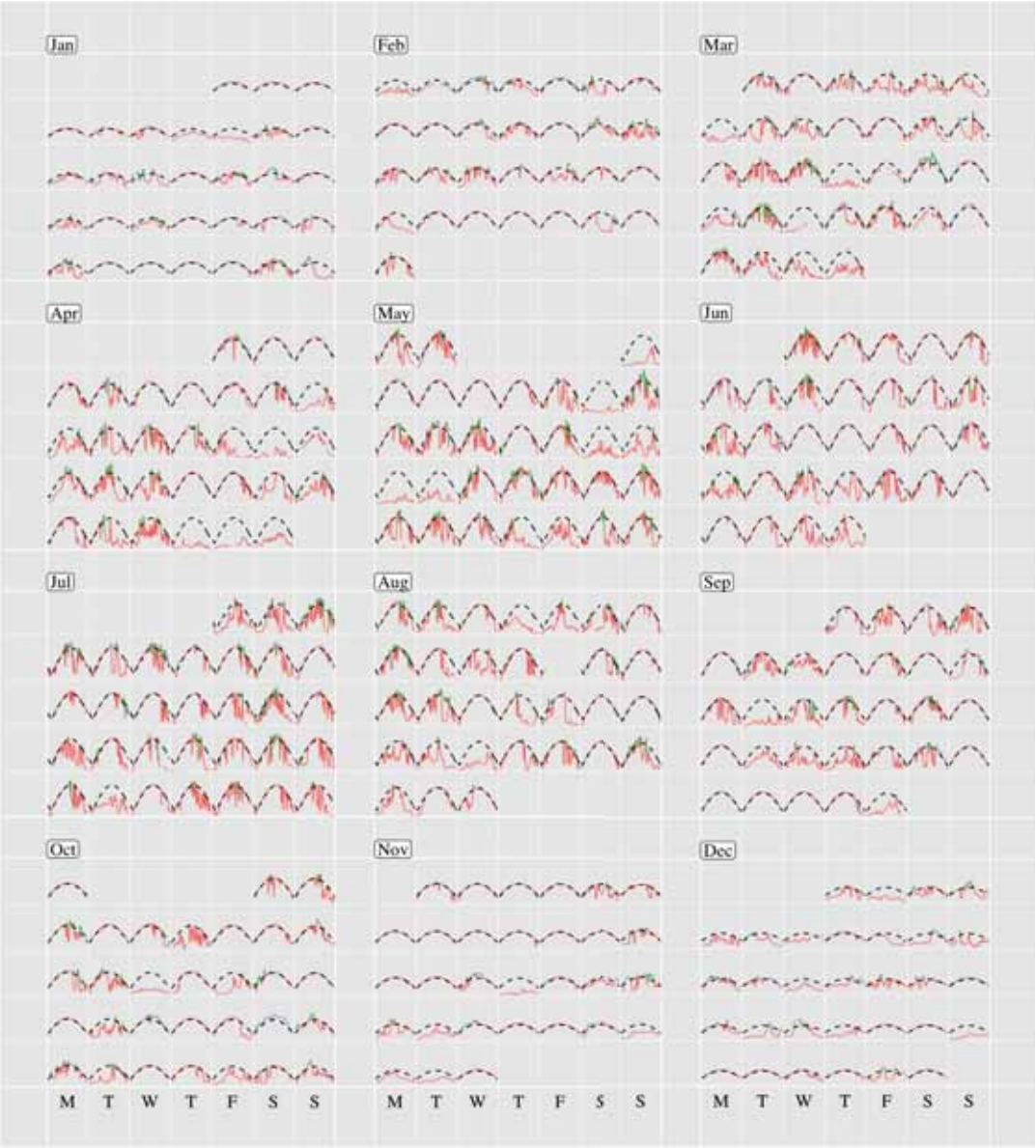

Figure 4. QC step 6) is displayed using calendar plots with the 6-min-averaged data. GHI is plotted using red solid lines. Cloud-enhancement events are marked with green, where as erroneous data are marked with blue. The Ineichen–Perez clear-sky irradiance is plotted in black dashed lines. Only data with $Z < 75^{\circ}$ are shown.

solid lines) for the entire year 2016 are plotted in Fig. 4. In addition, the Ineichen–Perez clear-sky model (black dashed lines) is overlaid. The cloud-enhancement events and the erroneous data are marked with green and blue, respectively. It is evident that not all data exceeding Clr are due to cloud enhancement. For example, such spurious data can be seen on October 22. In total, 2509 out of 87,840 6-min data points from BOS were identified as cloud-enhancement events, where as the number of erroneous data was 2784.

It is noted that the number of consecutive data points used in QC6 is subjected to the aggregation interval. For example, if 3-min averages are used, six points should be used when computing the standard deviation. Last but not the least, data gaps—due to sensor malfunction or logging issues—are found in most, if not all, long-term measurements. In some datasets, the data owners identify these gaps and perform imputation before data submission. In other cases, such as the BSRN data, these missing data points need to be manually identified by users. It is advised that data imputation should be performed at the

last QC stage, so that the imputation would not affect data aggregation. However, the missing data points can be marked with NA values at the very beginning of the QC, to note the discontinuity in the data. This is particularly important for time series analysis.

In practice, one typically develops his own method for quality controlling the aggregated data. For example, when irradiance data are used to validate irradiance separation models, only instances with $Z < 85^\circ$ and irradiance component > 0 should be considered [15]. For seasonal time series forecasting, nighttime data are manually set to zero [16]. It is in general difficult to know whether the various QC rules are sufficient for an application, or too stringent. On this point, visual inspection, such as using the calendar plot, or the daily irradiance *biplot* with time series features [17], should not be circumvented.

IV. Conclusion

Quality control of irradiance data is discussed in this paper. To perform a reasonably good QC, measurements of all three irradiance components, namely, GHI, DIF and DNI are required. In addition, solar positioning algorithm and thus clear-sky irradiance are also required, especially during the detection of cloud-enhancement events. Due to the high equipment and maintenance cost, such as the sun tracker or the high-precision pryheliometer, it is common to only monitor GHI at PV system sites. If this is the case, satellite-derived irradiance should be used to contrast the on-site irradiance measurements [18].

Acknowledgment

The authors would like to acknowledge the financial support by the Experimental Power Grid Centre at Agency for Science, Technology and Research (A*STAR).

References

[1] S. Killinger, N. Engerer, and B. Müller, "QCPV: A quality control algorithm for distributed photovoltaic array power output," *Solar Energy*, vol. 143, pp. 120 – 131, 2017.

[2] D. Yang, J. Kleissl, C. A. Gueymard, H. T. Pedro, and C. F. Coimbra, "History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining," *Solar Energy*, 2018, in press. [Online]. Available: https://doi.org/10.1016/j.solener.2017.11.023

[3] D. Yang, "Solar radiation on inclined surfaces: Corrections and benchmarks," *Solar Energy*, vol. 136, pp. 288 – 302, 2016.

[4] F. Vignola, J. Michalsky, and T. Stoffel, *Solar and Infrared Radiation Measurements*. CRC Press, Jun. 2012.

[5] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola, "A new operational model for satellite-derived irradiances: Description and validation," *Solar Energy*, vol. 73, no. 5, pp. 307 – 317, 2002.

[6] B. Kurtz and J. Kleissl, "Measuring diffuse, direct, and global irradiance using a sky imager," *Solar Energy*, vol. 141, pp. 311 – 322, 2017.

[7] C. A. Gueymard and D. R. Myers, "Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling," *Solar Energy*, vol. 83, no. 2, pp. 171 – 185, 2009.

[8] M. Sengupta, A. Habte, C. Gueymard, S. Wilbert, and D. Renne, "Best practices handbook for the collection and use of solar resource data for solar energy applications," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep. NREL/TP-5D00-68886, 2017.

[9] A. Ohmura, H. Gilgen, H. Hegner, G. Müller, M. Wild, E. G. Dutton, B. Forgan, C. Frölich, R. Philipona, A. Heimo, G. König-Langlo, B. McArthur, R. Pinker, C. H. Whitlock, and K. Dehne, "Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research," *Bulletin of the American Meteorological Society*, vol. 79, no. 10, pp. 2115 – 2136, 1998.

[10] C. N. Long and E. G. Dutton, "Bsrn global network recommended qc tests, v2." BSRN, Tech. Rep., 2002.

[11] C. N. Long and Y. Shi, "An automated quality assessment and control algorithm for surface radiation measurements," *The Open Atmospheric Science Journal*, vol. 2, no. 1, pp. 23 – 37, 2008.

[12] P. Ineichen and R. Perez, "A new airmass independent formulation for the Linke turbidity coefficient," *Solar Energy*, vol. 73, no. 3, pp. 151 – 157, 2002.

[13] C. A. Gueymard, "REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – validation with a benchmark dataset," *Solar Energy*, vol. 82, no. 3, pp. 272 – 285, 2008.

[14] N. H. Schade, A. Macke, H. Sandmann, and C. Stick, "Enhanced solar global irradiance during cloudy sky conditions," *Meteorologische Zeitschrift*, vol. 16, no. 3, pp. 295 – 303, 06 2007.

[15] C. A. Gueymard and J. A. Ruiz-Arias, "Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance," *Solar Energy*, vol. 128, pp. 1 – 30, 2016, special issue: Progress in Solar Energy.

[16] D. Yang, V. Sharma, Z. Ye, L. I. Lim, L. Zhao, and A. W. Aryaputera, "Forecasting of global horizontal irradiance by exponential smoothing, using decompositions," *Energy*, vol. 81, pp. 111 – 119, 2015.

[17] D. Yang, Z. Dong, L. H. I. Lim, and L. Liu, "Analyzing big time series data in solar engineering using features and PCA," *Solar Energy*, vol. 153, pp. 317 – 328, 2017.

[18] R. Urraca, A. M. Gracia-Amillo, T. Huld, F. J. M. de Pison, J. Trentmann, A. V. Lindfors, A. Riihelä, and A. Sanz-Garcia, "Quality control of global solar radiation data with satellite-based products," *Solar Energy*, vol. 158, pp. 49 – 62, 2017.