solar

MDPI

Article

Automating Quality Control of Irradiance Data with a Comprehensive Analysis for Southern Africa

Francisca Muriel Daniel-Durandt † and **Arnold Johan Rix** *,†

Department of E&E Engineering, Stellenbosch University, Stellenbosch 7602, South Africa; 18296033@sun.ac.za
* Correspondence: rix@sun.ac.za
† These authors contributed equally to this work.

Abstract: A review of quality control for large irradiance datasets is applied as a case study for the Southern African Universities Radiometric Network (SAURAN) database. The quality control procedure is automated and applied to 24 stations from the database with a total of 848,189 hourly datapoints. From this, the individual station's data quality is also analysed. The assessment validates the automated methodology without the need for a user-based review of the data. The SAURAN database can play a significant role in advancing solar and wind energy; however, the number of offline stations hinders this process. Data scarcity remains an obstacle to these goals, and therefore, recommendations are provided to address this. Recommendations regarding each site's usability in time-series and discrete applications are made, which provides an overall indication of the SAURAN database's irradiance measurement quality. Of the 24 measuring stations assessed, eight are recommended, 11 are recommended with cautious use, and five are recommended with extremely cautious use. These recommendations are based on multiple factors, such as whether a dataset has more than one full year of data or is missing minimal datapoints. Further, a study of the irradiance correlation between the stations was conducted. The results indicated groupings of different stations that showed highly correlated irradiance measurements and similar weather patterns. This is useful if a proposed renewable energy power plant, such as PV, falls within a cluster where the data from the SAURAN database can be used as a substitute if no data is available. SAURAN presents an opportunity for Southern Africa to increase its research outputs in solar and wind energy and lessen its dependency on fossil fuel-based energy production.

Keywords: quality control; data processing; automated quality control; ground-based measurements; irradiance dataset

Citation: Daniel-Durandt, F.M.; Rix, A.J. Automating Quality Control of Irradiance Data with a Comprehensive Analysis for Southern Africa. *Solar* **2023**, *3*, 596–617. https://doi.org/10.3390/solar3040032

Academic Editor: Loreto Valenzuela and Jürgen Heinz Werner

Received: 25 July 2023
Revised: 13 October 2023
Accepted: 24 October 2023
Published: 30 October 2023

1. Introduction

There is a growing globalised movement to avoid fossil-fuel-based energy production and increase the adoption of renewable energy sources. Solar photovoltaic (PV) systems are ideal for harvesting solar energy into electrical energy. High-quality radiation measurements are required for accurate designs and performance monitoring of PV systems [1,2].

Data-driven techniques such as power plant maintenance and operations, forecasting, and designing require good-quality irradiance measurements. Ground-based data contains periods, even if brief, of erroneous data caused by various reasons [3]. Quality control (QC) of large irradiance datasets is therefore instrumental in ensuring the continuing growth of PV systems. Automating these quality control procedures can speed up the process of the life cycle phases of PV systems by removing the laborious manual QC process [2].

The Southern African Universities Radiometric Network (SAURAN) is an initiative of the Centre for Renewable and Sustainable Energy Studies (CRSES) at Stellenbosch University and the Group for Solar Energy Thermodynamics (GSET) at the University of KwaZulu-Natal [4]. SAURAN aims to 'make high-resolution, ground-based solar radiometric data available from stations across the Southern African region'. South African data can be freely obtained from SAURAN [4]. The research indicates that there has not been a large-scale quality control assessment of the database's irradiance measurements, and this paper aims to address this. To the author's knowledge, an extensive automated survey has yet to be conducted on SAURAN in this manner.

SAURAN provides a guide and resources to QC, as per the terms of use, that the 'data quality is not assessed as part of the SAURAN initiative'. A QC program is provided as a guideline for the user in Python. The program provided does not provide QC measures on all the database stations. The program uses criteria such as the comparison of theoretical calculations of direct normal irradiance (DNI) with the measured values, filtering lower irradiance levels (such as values less than 5 W/m^2) and assessing missing timestamps [4,5]. The suggested quality control measures provide a guideline for a semi-automatic process, where the user must review data or add additional manual QC checks.

The goal of this paper is two-fold: to validate the automated QC methodology proposed by [2] with an application to the SAURAN database and to assess the quality of the SAURAN database. This is performed to prove that the proposed automated quality control process is sufficient as a minimum measure for ensuring a high standard of data quality. Each station is assessed individually to ensure that the automated QC procedure is sufficient in removing erroneous data from a dataset. Each station is evaluated until 31 October 2022, and hourly-aggregated irradiance measurements from the database are used. Only the hourly irradiance data, meaning the diffuse horizontal irradiance (DHI), global horizontal irradiance (GHI) and DNI, are assessed for this study and not additional meteorological measurements such as temperature, wind, and UV measurements. The evaluation will indicate the advantages and shortcomings of large irradiance datasets and their quality control. The SAURAN database is highlighted to further the exposure for other researchers and users of its value to the worldwide goal of improving sustainability. A manual quality control process and review of each datapoint of the dataset is a highly time-consuming process and increases as the datasets get larger. Removing this step without compromising on the overall integrity of the database is a step forward for modelling and monitoring these systems, applicable specifically to PV and concentrating solar power.

2. Background

PV systems require solar radiation data for design and monitoring purposes. Irradiance measurements such as DHI, GHI and DNI are usually used for these purposes. Common errors in irradiance datasets include errors caused by equipment, operational problems, and data-processing errors [6]. The most general errors are the cosine response, azimuth response, temperature response, spectral selectivity, stability, non-linearity, shadering misalignment, and nocturnal long-wave radiation errors [6,7]. The interested reader is referred to [7].

Quality control of large irradiance datasets is a critical data-processing technique that ensures common errors are flagged and removed from the dataset. There have been significant publications on various QC methodologies [1,3,8–13]. The Baseline Surface Radiation Network (BSRN) recommendations by [10] are a popular QC methodology and have been used as a proposed QC methodology in various works [3,13–16]. Most of the research previously mentioned includes a form of user review, which can become time-consuming as dataset sizes increase. An automated QC procedure can significantly reduce the manual time it would have taken for the user to review the data while ensuring a high standard of data quality [2].

Figure 1 shows the weather stations of the SAURAN network over Southern Africa. The orange and blue dots indicate online and permanently offline stations, respectively. The SAURAN network is supported by partners such as The Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) and the United States Agency for International Development (USAid) [4].

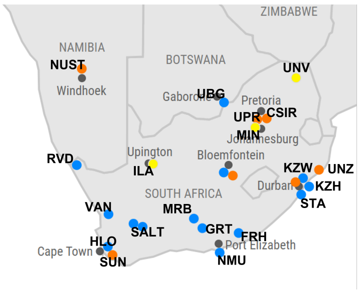

Figure 1. SAURAN stations across Southern Africa [4].

With the use of the SAURAN database, the scientific community has been gaining traction in recently published works [17–25]. Not all authors include documented QC measures when using the SAURAN database.

Documented QC methodologies for SAURAN data in publications include Jacovides et al. [26] in [27], McArthur [28]'s recommendations in [29] and Long and Shi [11]'s recommendations in [30]. In Mutombo et al. [21]'s study of the Mangosuthu University of Technology (STA) station of the SAURAN database, the authors replaced missing data with the previous years' averages. They combined the missing data from different years into one year [21]. This method is not ideal, as the integrity of the data is compromised. Sun et al. [17] exclude all abnormal GHI measurements and timestamps where the solar zenith angle θ_Z exceeded 85° [17]. Missing timestamps were replaced with satellite data in the report of [18]. Nwokolo et al. [31] applied QC techniques such as using the clearness index K_t, among other indices, that fall within the range of 0 to 1 [31].

A proposed QC methodology by [2] combines four QC methodologies [1,10,26,32]. An additional tracking error test was included. The results presented in the paper showed adequate QC capabilities for an automatic procedure, i.e., automatic elimination. The data is flagged based on whether the data conforms to certain limits. The paper only assessed two stations for a period from 2019 to 2021.

This paper presents an in-depth analysis of the SAURAN stations by applying the proposed QC methodology of [2] to assess the quality of all 24 available SAURAN stations at the time of this study, but also determine if the automated QC procedure is valid without a user review.

3. Methodology

This study will cover 24 stations at SAURAN [4]. A network summary is discussed, followed by the QC procedure, which will be applied to each station.

3.1. SAURAN Station Summary

SAURAN allows the download of up to one year's data at a time and not the entire dataset. Data averaged by minute, hour and each day is available for download and is usually downloaded in a comma-separated values (csv) format. The stations use Kipp & Zonen radiometers for irradiance measurements, and a detailed summary of available measurements and equipment can be seen in Table 1 [4,9].

Table 2 provides a summary of the SAURAN stations: the location, coordinates, as well as elevation. All data is hourly-averaged measurements and is available on the SAURAN website as of 31 October 2022.

Table 1. Measurements and equipment of SAURAN stations [4].

Parameter	Typical Instrument	Unit
GHI	Kipp&Zonen CMP11 pyranometer	W/m^2
DNI	Kipp&Zonen CHP1 pyrheliometer	W/m^2
DHI	Kipp&Zonen CMP11 pyranometer	W/m^2

Table 2. SAURAN station summary [4].

	Label	Name (Location)	Coordinates (Lat (°S), Long (°E))	Elevation (m)
1	CSIR	CSIR Energy Centre (Pretoria, South Africa)	25.747, 28.279	1400
2	CUT	Central University of Technology (Bloemfontein, South Africa)	29.121, 26.216	1397
3	FRH	University of Fort Hare (Alice, South Africa)	32.785, 26.845	540
4	GRT	Graaff-Reinet (Graaff-Reinet, South Africa)	32.485, 24.586	660
5	HLO	Mariendal (Mariendal, South Africa)	33.854, 18.824	178
6	ILA	Ilanga CSP Plant (Upington, South Africa)	28.490, 21.520	884
7	KZH	University of KwaZulu-Natal Howard College (Durban, South Africa)	29.871, 30.977	150
8	KZW	University of KwaZulu-Natal Westville (Durban, South Africa)	29.817, 30.945	200
9	MIN	CRSES Mintek (Johannesburg, South Africa)	26.089, 27.978	1521
10	MRB	Murraysburg (Murraysburg, South Africa)	31.890, 24.056	1548
11	NMU	Nelson Mandela University (Gqeberha, South Africa)	34.009, 25.665	35
12	NUST	Namibian University of Science and Technology (Windhoek, Namibia)	22.565, 17.075	1683
13	PMB	University of KwaZulu-Natal Pietermaritzburg (Pietermaritzburg, South Africa)	29.621, 30.397	680
14	RVD	Richtersveld (Alexander Bay, South Africa)	28.561, 16.761	141
15	SALT	Eskom Sutherland SALT (Sutherland, South Africa)	32.378, 20.812	1761
16	STA	Mangosuthu University of Technology (Umlazi, South Africa)	29.970, 30.915	95
17	SUN	Stellenbosch University (Stellenbosch, South Africa)	33.935, 18.867	119
18	SUT	Sutherland (Sutherland, South Africa)	32.222, 20.348	1450
19	UBG	Gaborone (Gaborone, Botswana)	24.661, 25.934	1014
20	UFS	University of Free State (Bloemfontein, South Africa)	29.111, 26.185	1491
21	UNV	Venda (Vuwani, South Africa)	23.131, 30.424	628
22	UNZ	University of Zululand (KwaDlangezwa, South Africa)	28.853, 31.852	90
23	UPR	University of Pretoria (Pretoria, South Africa)	25.753, 28.229	1410
24	VAN	Vanrhynsdorp (Vanrhynsdorp, South Africa)	31.617, 18.738	130

3.2. Automated Quality Control

The proposed QC methodology by [2] suggests a time-series visualisation and removes duplicate and missing timestamps. An automated QC methodology is applied where erroneous datapoints are automatically eliminated from the dataset without needing a user review.

The automatic elimination flagging procedure includes datapoints which are physically impossible (from [10]), removing timestamps where no irradiance is recorded, i.e., night-time values (using [1,26]) *K*-tests (from [32]) and tracking errors. These QC procedures are well-established and used within the available literature [2].

The automatic elimination process consists of three parts: removing night-time values, *K*-tests, and then the individual limits of GHI, DHI, and DNI. The *K*-tests involve the direct beam transmittance K_n, diffuse transmittance K_d and K_t. These K-values are given as:

$$K_t = \frac{GHI}{I_{0n} \cos \theta_Z}, \tag{1}$$

$$K_d = \frac{DHI}{GHI}, \tag{2}$$

$$K_n = \frac{DNI}{I_{0n}}. \tag{3}$$

I_{0n} is the extraterrestrial irradiance on a normal surface:

$$I_{0n} = I_{SC}\left(1 + 0.033 \cdot \cos\left(\frac{360 \cdot n}{365}\right)\right) \tag{4}$$

where n denotes the day of the year and I_{SC} denotes the solar constant (1367 W/m^2).

The horizontal extraterrestrial irradiance G_{0h} is defined by:

$$G_{0h} = I_{0n} \cdot \cos \theta_Z. \tag{5}$$

GHI, DHI and DNI units are in W/m^2, and all K-values (K_t, K_n and K_d) are unitless.

3.2.1. Night-Time Values

The night-time values are removed using:

$$GHI > 5, \tag{6}$$

$$\theta_Z < 85^\circ. \tag{7}$$

Most flagged data should remove the night-time constraints where irradiance levels are very low, and the sun is close to sunrise or sunset.

3.2.2. *K*-Tests

K-tests are applied by assessing the K_d, K_n and K_t-space. The BSRN tests, as well as the *K*-tests, are applied as follows:

$$K_d < 1.05 \text{ for } GHI > 50 \text{ and } \theta_Z < 75^\circ, \tag{8}$$

$$K_d < 1.10 \text{ for } GHI > 50 \text{ and } \theta_Z > 75^\circ, \tag{9}$$

$$K_n < 0.8, \tag{10}$$

$$K_d < 0.96 \text{ for } K_t > 0.6. \tag{11}$$

3.2.3. Individual Limits

The individual limits for GHI, DHI, and DNI are assessed using:

$$-4 < GHI < 1.5 I_{0n} (\cos\theta_Z)^{1.2} + 100, \tag{12}$$

$$-4 < DHI < 0.95 I_{0n} (\cos\theta_Z)^{1.2} + 50, \tag{13}$$

$$-4 < DNI < I_{0n}, \tag{14}$$

$$DHI < 0.8 \cdot G_{0h}, \tag{15}$$

$$GHI - DHI < G_{0h}, \tag{16}$$

$$DNI < 1100 + 0.03 Elev. \tag{17}$$

$Elev$ is the elevation in m.

3.2.4. Tracking Error

A tracking error is included when $GHI \approx DHI$ and $DNI \approx 0$:

$$0.8 < K_d < 1.2 \text{ and } K_n < 0.01. \tag{18}$$

The removal of night-time values should be conducted first. This optimises the computational processing, which will immediately exclude unusable datapoints. The tracking error test is a proposed way of automatically deducing whether a dataset has a tracking error.

3.3. Correlation Assessment

The stations' irradiance relationships are assessed by analysing the subsequent correlation matrix. An analysis of highly correlated stations identifies areas within Southern Africa with similar irradiance patterns. Potential PV plant locations within these areas can use the SAURAN database if no on-site irradiance data is available. This was performed by assessing the stations' overlapping timestamps, as shown in Figure 2. Therefore, stations with no overlapping timestamps will not have a calculated correlation.

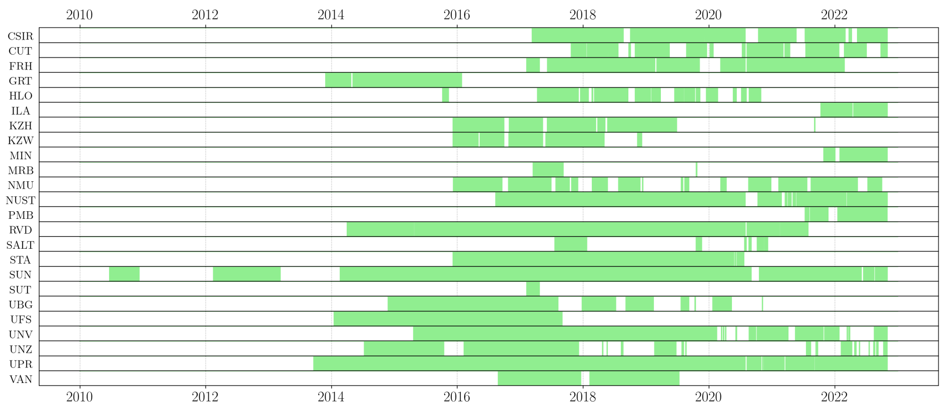

Figure 2. Overlapping periods of stations within SAURAN database.

The Pearson correlation coefficient ρ indicates the correlation between data and is defined as:

$$\rho = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\sum(x_i - \bar{x})^2\right)\left(\sum(y_i - \bar{y})^2\right)}} \tag{19}$$

where x_i and y_i is the i-th measurement of the x and y variables, and $\bar{x}$ and $\bar{y}$ are the mean of the values of the x and y variables.

Therefore, closer to -1 has a negative correlation, meaning if one variable increases, the other decreases. In contrast, closer to 1 has a positive correlation, meaning if one variable increases, the other would also [33].

4. SAURAN Database Review

Figures 3 and 4 show a practical example of applying the QC procedure to a dataset. Each figure shows a period in the dataset where data has been automatically flagged as erroneous. Underneath the time-series visualisation, the different flags according to Sections 3.2.1–3.2.4 are highlighted if the automatic QC process has flagged them.

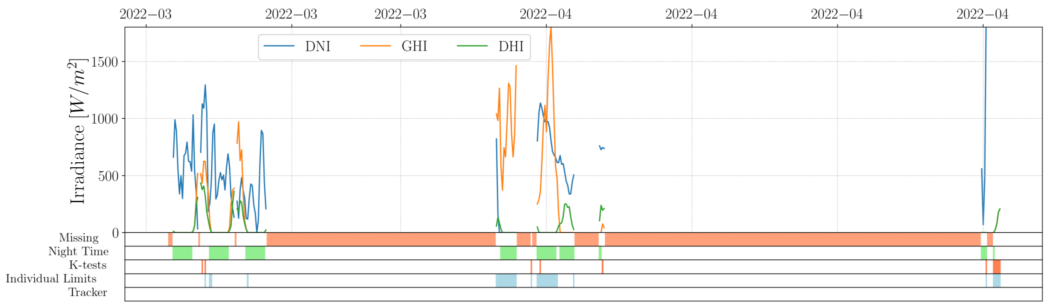

Figure 3. Application of quality control procedure: identifying faulty data.

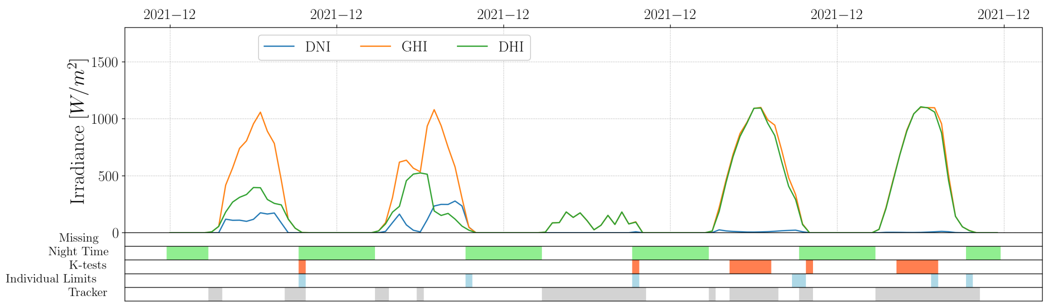

Figure 4. Application of quality control procedure: identifying tracking error.

Figure 3 shows substantial datapoints flagged as missing data. Night-time datapoints were flagged as expected, as well as the individual limits and K-tests. It is clear from the figure that there are faulty measurements in this dataset, which the QC procedure has automatically flagged.

Figure 4 has no missing data, and the night-time datapoints were also flagged as expected. It is visually evident in Figure 4 that a tracking error starts to occur on the third day, which was automatically flagged by the K-tests and tracking error flag. In comparison, the first day in Figure 4 shows how the very low irradiance levels at sunrise and sunset are triggered by the different flags that will be removed from the dataset accordingly.

The observed trend is that if there is faulty data, it is usually flagged by more than one limit. The QC process is, therefore, also useful as a real-time monitoring tool to identify any faulty measurement tools and correct them accordingly. The individual stations will show the missing, night-time datapoints and the other corresponding flags, as discussed in Sections 3.2.2–3.2.4.

Each station is assessed on an individual level. A time-series visualisation of the dataset is presented, which shows how the automatic QC procedure flags the erroneous data. Three groupings of flagged data are presented:

1. missing data;
2. night-time values;
3. and *K*-tests, individual limits and tracking error.

This is performed to present a visual tool for how the QC process works to identify erroneous data. An overall summary of the SAURAN network data is tabulated with recommendations for the usability and quality of each station. Table 3 summarises the stations' dataset sizes before and after QC, as well as the start and end dates of the datasets. This study assesses 848,189 hourly datapoints from the SAURAN database.

Table 3. SAURAN datasets [4].

Station	Dataset Size				Start Date	End Date
	Before QC	Night-Time & Duplicates Removed	Other Flags Removed	After QC		
CSIR	46,434	26,539	9560 (21%)	16,979 (37%)	11 March 2017	31 October 2022
CUT	28,077	14,619	2737 (10%)	11,882 (42%)	24 October 2017	31 October 2022
FRH	40,895	22,233	8148 (20%)	14,085 (34%)	7 February 2017	24 February 2022
GRT	18,541	9774	2438 (13%)	7336 (40%)	27 November 2013	24 January 2016
HLO	21,532	11,728	3503 (16%)	8225 (38%)	8 October 2015	27 October 2020
ILA	8832	4676	1057 (12%)	3619 (41%)	13 October 2021	31 October 2022
KZH	52,323	38,898	29,612 (57%)	9286 (18%)	7 December 2015	07 August 2022
KZW	20,291	10,756	4503 (22%)	6253 (31%)	7 December 2015	12 December 2018
MIN	8185	4423	1308 (16%)	3115 (38%)	28 October 2021	31 October 2022
MRB	4201	2462	850 (20%)	1612 (38%)	17 March 2017	22 October 2019
NMU	39,969	23,130	11,171 (28%)	11,959 (30%)	10 December 2015	30 September 2022
NUST	52,004	27,401	6096 (12%)	21,305 (41%)	26 July 2016	31 October 2022
PMB	9773	5415	2337 (24%)	3078 (31%)	13 July 2021	31 October 2022
RVD	63,716	34,457	8234 (13%)	26,223 (41%)	27 March 2014	28 July 2021
SALT	14,151	9908	7526 (53%)	2382 (17%)	21 July 2017	22 December 2020
STA	40,256	21,751	10,413 (26%)	11,338 (28%)	7 December 2015	19 April 2021
SUN	87,720	47,733	14,304 (16%)	33,429 (38%)	24 May 2010	31 October 2022
SUT	1715	902	115 (7%)	787 (46%)	8 February 2017	20 April 2017
UBG	38,917	20,646	6534 (17%)	14,112 (36%)	26 November 2014	6 November 2020
UFS	31,665	17,152	4060 (13%)	13,092 (41%)	16 January 2014	30 August 2017
UNV	59,100	33,144	15,226 (26%)	17,918 (30%)	23 April 2015	31 October 2022
UNZ	56,399	30,373	18,953 (34%)	11,420 (20%)	11 July 2014	31 October 2022
UPR	78,792	42,128	10,464 (13%)	31,664 (40%)	19 September 2013	31 October 2022
VAN	24,701	13,234	3414 (14%)	9820 (40%)	26 August 2016	10 July 2019

4.1. Quality Control of SAURAN Stations

An automated quality control procedure is applied to each station of the SAURAN database. Each station is discussed and summarised, and a final recommendation is given in Section 5.

4.1.1. CSIR

The CSIR Energy Centre (CSIR) station is in Pretoria, South Africa. The location's information is described in Table 2. The earliest available data is from March 2017, and the station is shown as currently online [4]. The pre-processed CSIR data are shown in Figure 5, with the flags triggered by the automated QC procedure. Unusually high DHI measurements and physically impossible DHI measurements are noticed. There are also two significant gaps in the dataset, which have been flagged accordingly. These datapoints will be removed from the dataset accordingly.

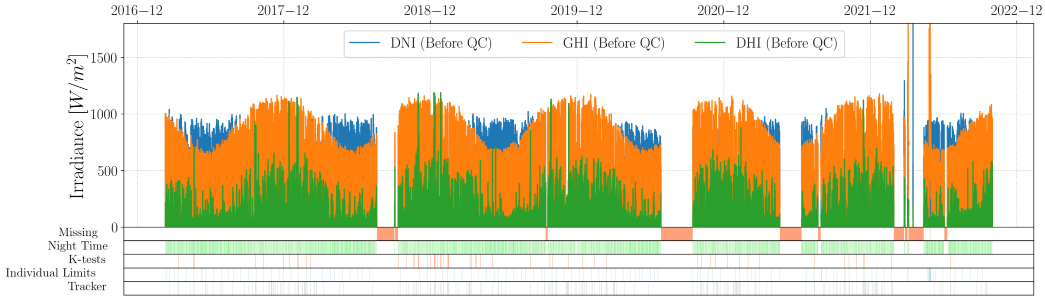

Figure 5. Pre-processed CSIR station with flagged datapoints.

4.1.2. CUT

The Central University of Technology (CUT) station is in Bloemfontein, South Africa (see Table 2 for information regarding the station's geographical information). Available data starts from October 2017, and the station is currently online [4]. Figure 6 visualises the pre-processed CUT dataset and how the QC procedure has flagged the dataset. There are significant missing data gaps in the set, and it would be ideal to have at least one year's worth of quality data available for research applications. The *K*-tests do not entirely encapsulate faulty DHI measurements around December 2012—a flaw of the automated QC procedure. However, the automated QC does remove the greater majority of faulty measurements.

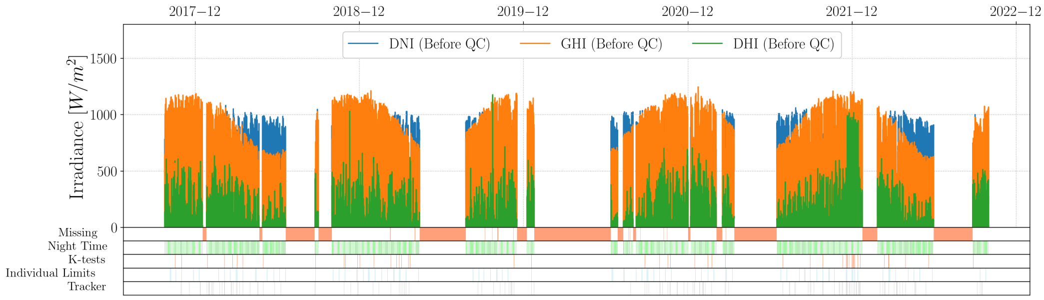

Figure 6. Pre-processed CUT station with flagged datapoints.

4.1.3. FRH

The USAid University of Fort Hare (FRH) station is in Alice, South Africa, and the station's geographical description is in Table 2. The data is available from February 2017, and the station is permanently offline from February 2022 [4]. Figure 7 visualises the pre-processed FRH dataset with the associated flagged data. A tracking error was observed during 2017, and there is a significant missing data period. The QC procedure successfully identifies the erroneous data by flagging them accordingly.

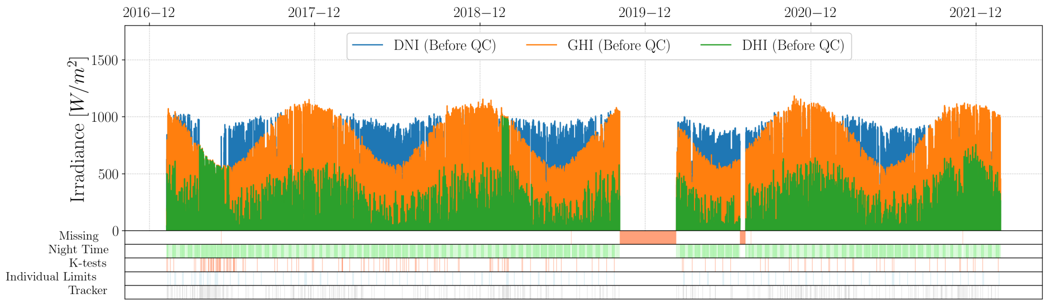

Figure 7. Pre-processed FRH station with flagged datapoints.

4.1.4. GRT

The GIZ Graaff-Reinet (GRT) station is situated in Graaff-Reinet, South Africa, and the stations' geographical properties are described in Table 2. Data is available from November

2013 to January 2016 and is permanently offline [4]. Figure 8 visualises the pre-processed GRT dataset. Other than the small missing data gap, the dataset has no obvious flaws.

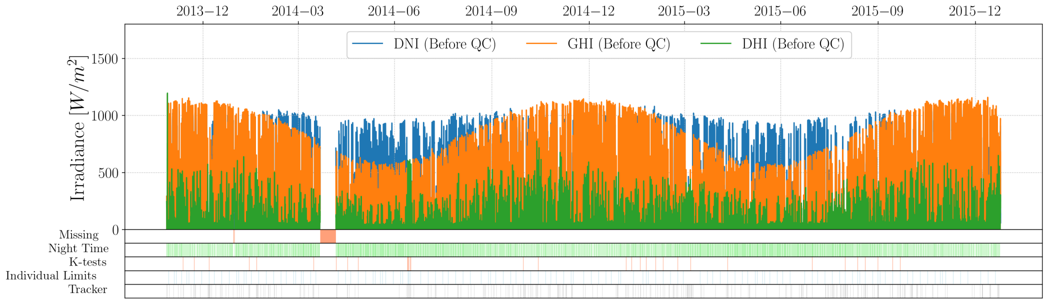

Figure 8. Pre-processed GRT station with flagged datapoints.

4.1.5. HLO

The Mariendal (HLO) station is located in Mariendal, South Africa, and the geographical description is summarised in Table 2. The dataset spans from October 2015 until October 2020 and is permanently offline [4]. Figure 9 visualises the pre-processed HLO dataset. There are significant gaps in the data and unusual measurements visible. There is not an entire year's worth of data available.

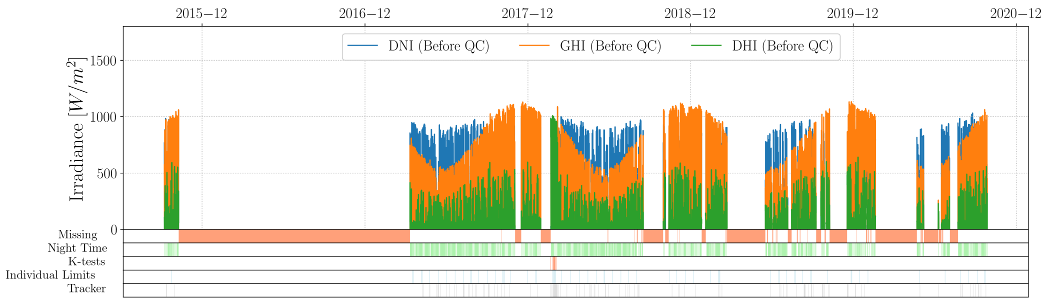

Figure 9. Pre-processed HLO station with flagged datapoints.

4.1.6. ILA

The Ilanga CSP plant (ILA) is situated in Upington, South Africa, and its geographical properties are listed in Table 2. The pre-processed dataset of ILA with the associated flags is presented in Figure 10. Data is available from October 2021 onwards, and the station is online [4]. It is noted that the station has very low DHI measurements in general. The station is relatively new, and the flagged data combined with the visualised time series shows that there is not one reliable year's worth of data available.

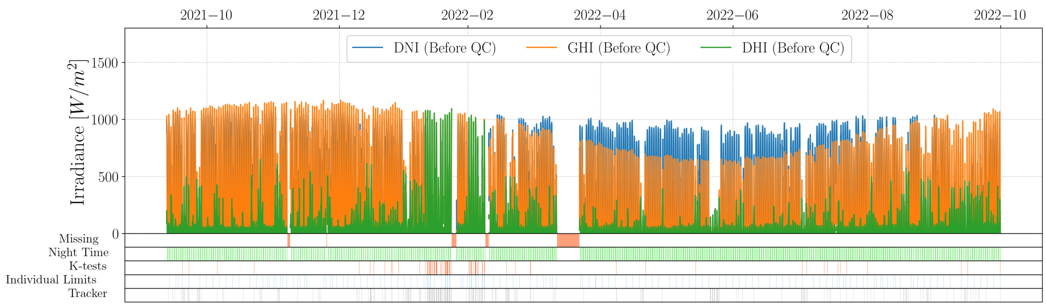

Figure 10. Pre-processed ILA station with flagged datapoints.

4.1.7. KZH

The University of KwaZulu-Natal Howard College (KZH) station is in Durban, South Africa, and the geographical location is described in Table 2. Figure 11 visualises the pre-processed KZH dataset. Data is available from December 2015 to August 2022, but there is a significant amount of missing data and the dataset is unreliable from June 2019 to August 2022. The station is permanently offline [4].

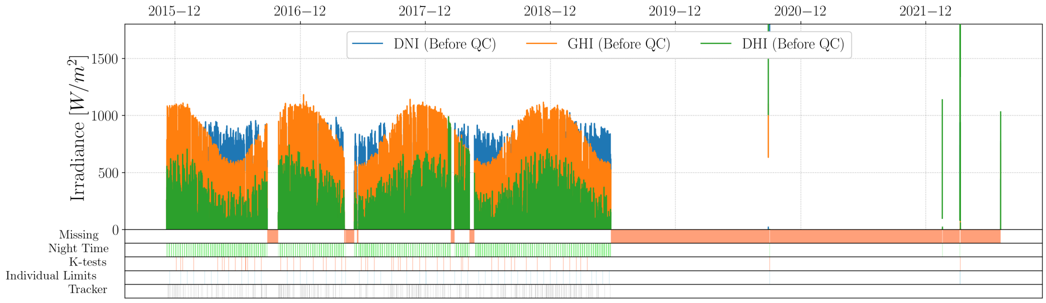

Figure 11. Pre-processed KZH station with flagged datapoints.

4.1.8. KZW

The University of KwaZulu-Natal Westville (KZW) station is located in Durban, South Africa, and its geographical properties are listed in Table 2. Data is available from December 2015 to December 2018, and the station is permanently offline [4]. Figure 12 visualises the pre-processed KZW dataset. There are physically impossible DNI measurements and a noticeable gap in the dataset.

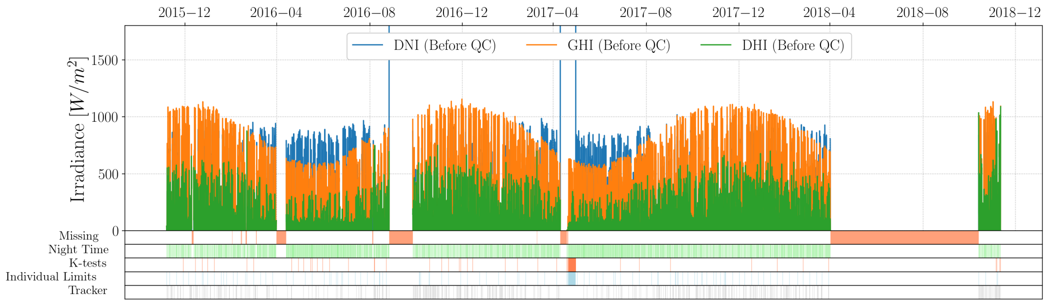

Figure 12. Pre-processed KZW station with flagged datapoints.

4.1.9. MIN

The Mintek (MIN) station is located in Johannesburg, South Africa, and its geographical description is in Table 2. The dataset has data available from October 2021 onwards and is currently online [4]. The pre-processed MIN dataset is shown in Figure 13. A tracking error was observed at the end of 2021. The MIN station is relatively new, and there is not one reliable year's worth of data available.

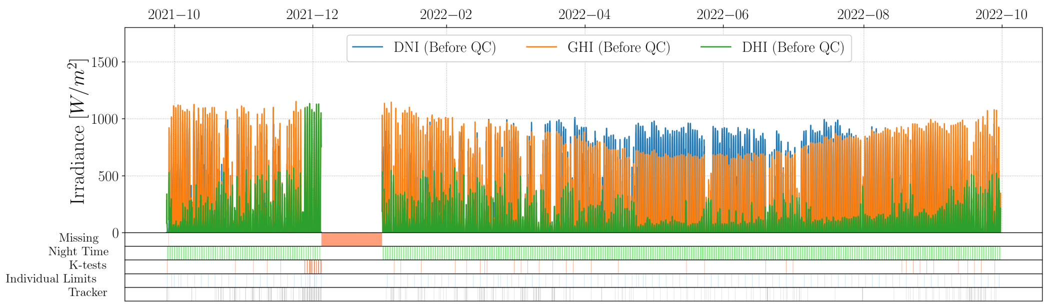

Figure 13. Pre-processed MIN station with flagged datapoints.

4.1.10. MRB

The GIZ Murraysburg (MRB) station is located in Murraysburg, South Africa, and its geographical properties are listed in Table 2. Data is available from March 2017 until

October 2019; however, it should be noted that there is considerable missing data from 5 September 2017 onwards. Figure 14 visualises the pre-processed dataset. The dataset is empty for most of 2018 and 2019, with a few singular datapoints in 2019. The dataset is not ideal as there is not a complete year available.

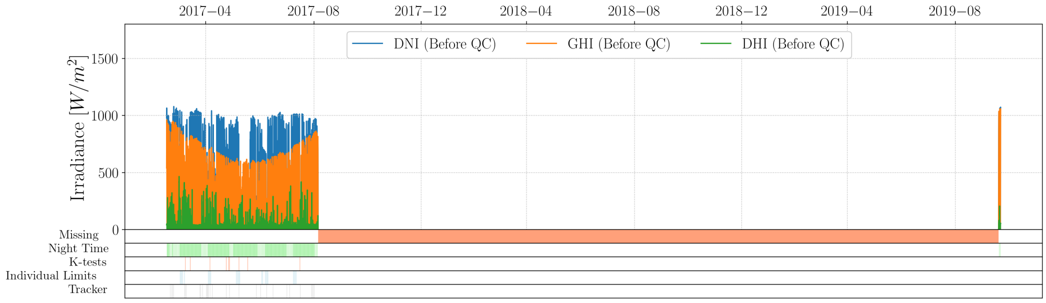

Figure 14. Pre-processed MRB station with flagged datapoints.

4.1.11. NMU

The Nelson Mandela University (NMU) is located in Gqeberha (previously known as Port Elizabeth), South Africa, and Table 2 describes the geographical location properties. Data is available from December 2015 to September 2022, and the station is permanently offline [4]. In Figure 15, the pre-processed NMU dataset indicates obvious missing data gaps, tracking errors, and unusually high DNI measurements. There is not a complete year available in the dataset.

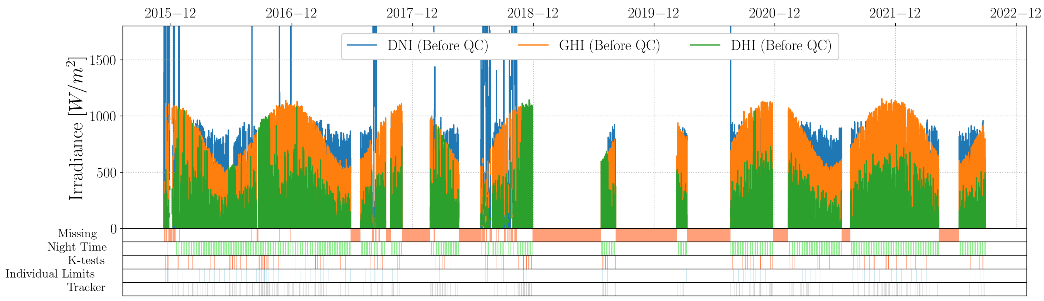

Figure 15. Pre-processed NMU station with flagged datapoints.

4.1.12. NUST

The USAid Namibian University of Science and Technology (NUST) is located in Windhoek, Namibia, and Table 2 summarises the geographical properties of the station. Data is available from July 2016, and the station is currently online [4]. Figure 16 visualises the pre-processed dataset. Low DHI measurements are observed, which could be attributed to the locations' climate, and a tracking error was observed in 2021.

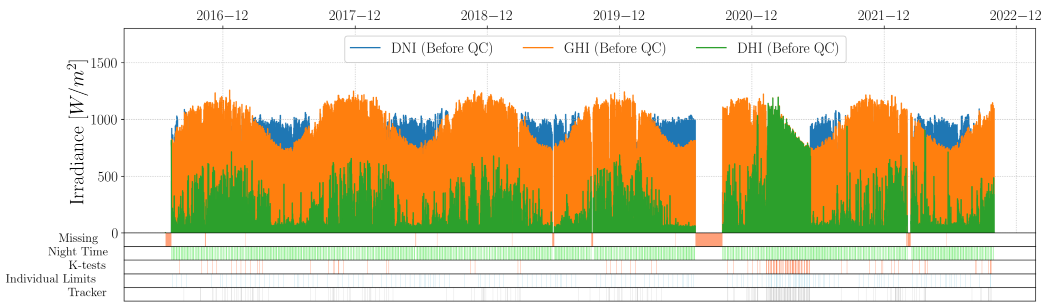

Figure 16. Pre-processed NUST station with flagged datapoints.

4.1.13. PMB

The University of KwaZulu-Natal Pietermaritzburg (PMB) station is located in Pietermaritzburg, South Africa, and its geographical properties are listed in Table 2. Data is available from July 2021, and the station is currently online [4]. Figure 17 shows the pre-processed dataset. The station is relatively new, and there are missing data periods. Thus, a full year's worth of data is unavailable in the figure.

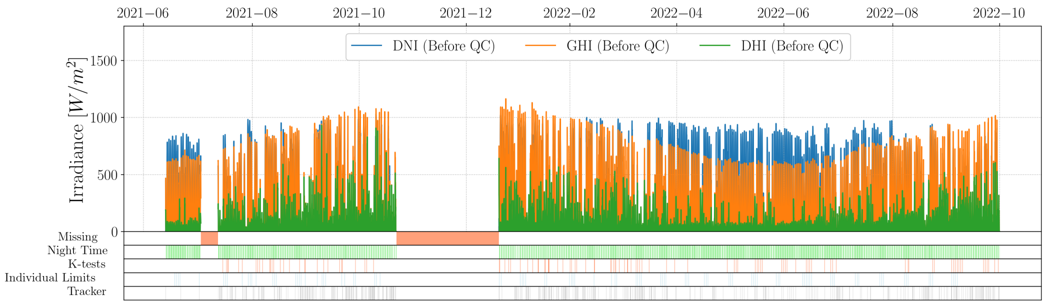

Figure 17. Pre-processed PMB station with flagged datapoints.

4.1.14. RVD

The Richtersveld (RVD) station is located in Alexander Bay, South Africa, and Table 2 describes the station's geographical properties. Data is available from March 2014 to July 2021, and the station is permanently offline [4]. Figure 18 visualises the pre-processed dataset. Other than the tracking error where $DHI \approx GHI$ is observed for an interval at the end of 2019, there seem to be minimal additional problems with the dataset. The QC procedure removes most of the faulty data. However, it was noted that there were suspicious DHI data for a short period, which the *K*-tests and Individual Limits tests did not flag.

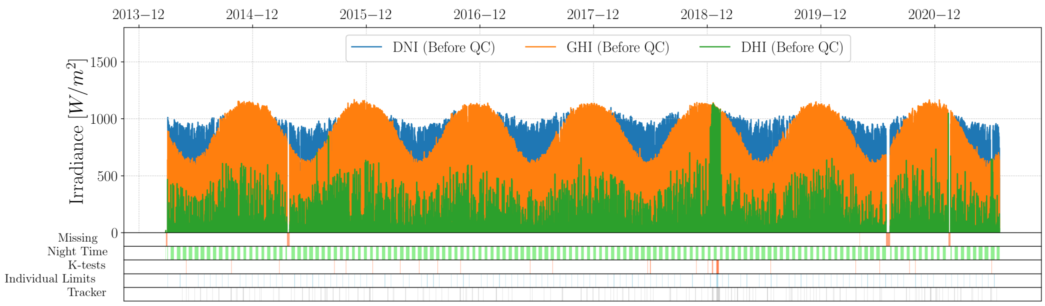

Figure 18. Pre-processed RVD station with flagged datapoints.

4.1.15. SALT

The Eskom Sutherland SALT (SALT) station is located in Sutherland, South Africa, and Table 2 describes the location. Data is available from July 2017 to December 2020, and the station is permanently offline [4]. The previous station, abbreviated as SUT, has data available from February 2017 to April 2017 [4] and will be discussed separately due to the location change. Figure 19 visualises the SALT dataset. The dataset is incomplete, especially after applying the QC procedure, when the tracking error ($DHI \approx GHI$) is removed.

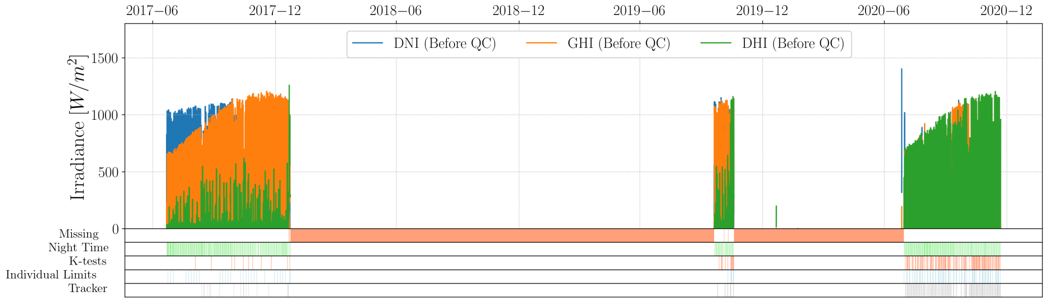

Figure 19. Pre-processed SALT station with flagged datapoints.

4.1.16. STA

The STA station is located in Umlazi, South Africa, and the geographical properties are described in Table 2. Data is available from December 2015 to April 2021, and the station is permanently offline [4]. Figure 20 visualises the pre-processed dataset. There are unusual DNI measurements that indicate that the data is faulty. A tracking error is also observed where $DHI \approx GHI$. After the QC steps are applied, as shown in Figure 20, the tracking error is not entirely removed. This could be because of faulty DNI measurements. The GHI and DHI data are still usable for applications.

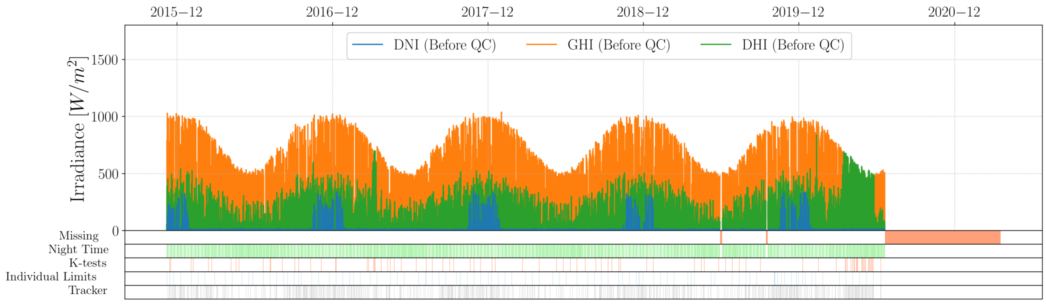

Figure 20. Pre-processed STA station with flagged datapoints.

4.1.17. SUN

The Stellenbosch University (SUN) station is located in Stellenbosch, South Africa, and its geographical description is in Table 2. Data is available from May 2010 and is currently online [4]. Figure 21 visualises the pre-processed dataset. There are unusually high DNI measurements and prominent gaps in the dataset, specifically from 2010 to 2013. The *K*-tests and Individual Limit tests do not entirely remove faulty DHI, as with the CUT and RVD datasets; however, most faulty data was removed.

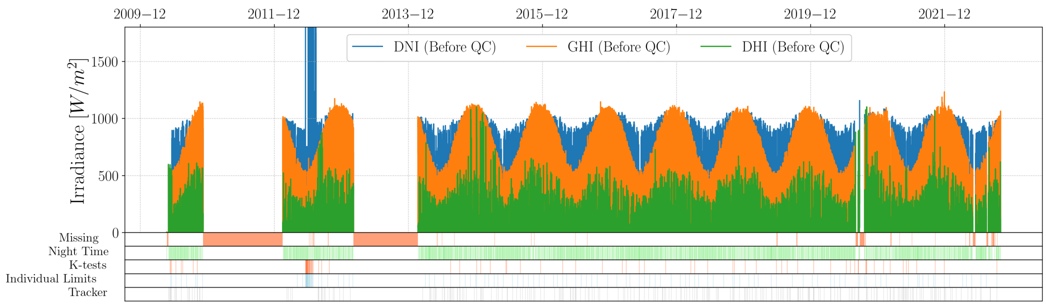

Figure 21. Pre-processed SUN station with flagged datapoints.

4.1.18. SUT

The Sutherland (SUT) station is located in Sutherland, South Africa, and its geographical properties are listed in Table 2. Data is available from February to April 2017, after which it was moved to the SALT station [4]. The pre-processed dataset is visualised in

Figure 22. The dataset does not include an entire year's worth of data; however, the dataset presented does have minimal missing data.

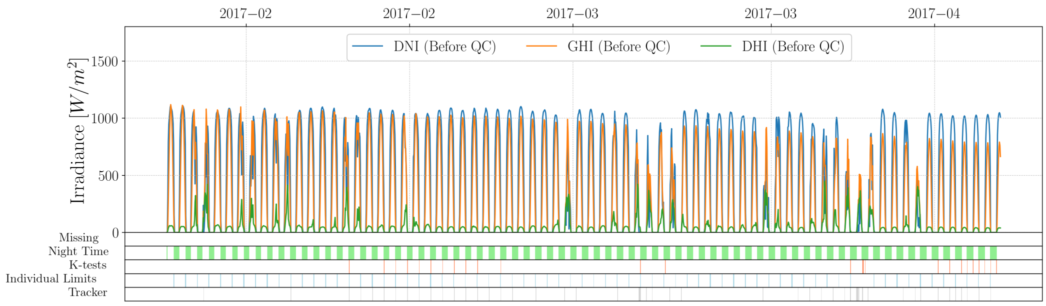

Figure 22. Pre-processed SUT station with flagged datapoints.

4.1.19. UBG

The USAid Gaborone (UBG) station is located in Gaborone, Botswana, and its geographical properties are described in Table 2. Data is available from November 2014 until November 2020, and the station is permanently offline [4]. Figure 23 visualises the pre-processed UBG dataset. There is missing data in the dataset and unusually high DHI measurements. A period of a tracking error is noticeable (where $DHI \approx GHI$). The QC steps remove these errors, except for the extremely low irradiance not flagged by the automated QC steps.

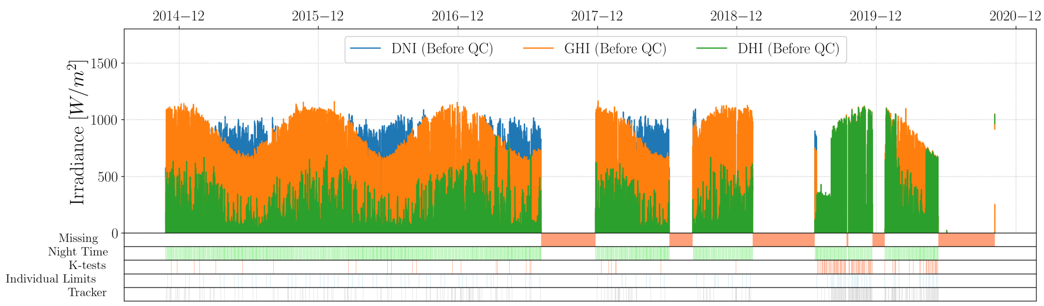

Figure 23. Pre-processed UBG station with flagged datapoints.

4.1.20. UFS

The GIZ University of Free State (UFS) is situated in Bloemfontein, South Africa, and the geographical information is shown in Table 2. Data is available from January 2014 until August 2017, and the station is permanently offline [4]. Figure 24 visualises the pre-processed UFS dataset. There are no apparent flaws in the dataset.

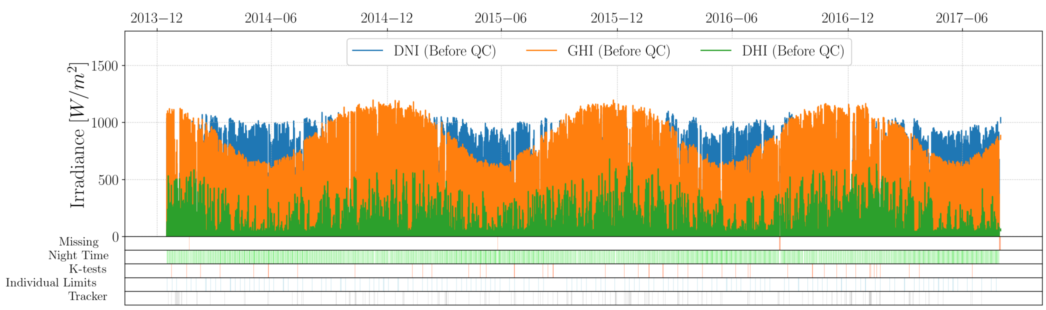

Figure 24. Pre-processed UFS station with flagged datapoints.

4.1.21. UNV

The USAid Venda (UNV) station is in Vuwani, South Africa, and Table 2 describes its geographical properties. Data is available from April 2015 onwards and is currently online [4]. Figure 25 visualises the pre-processed dataset. The dataset has missing data and unusually high DHI and DNI measurements. The dataset has more than one full year of data.

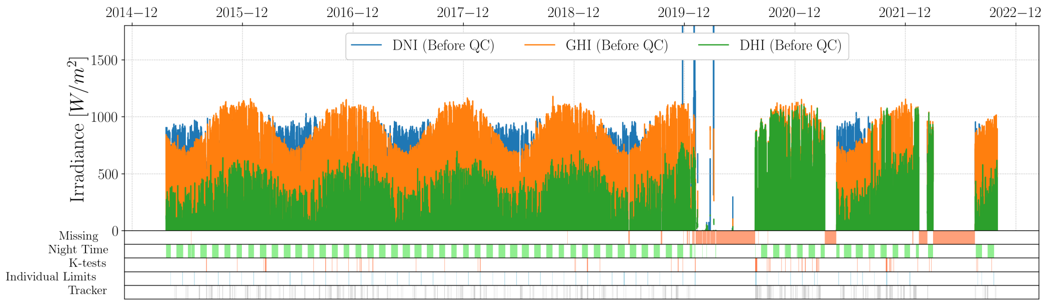

Figure 25. Pre-processed UNV station with flagged datapoints.

4.1.22. UNZ

The University of Zululand (UNZ) station is located in KwaDlangezwa, South Africa, and Table 2 describes its geographical location. Data is available from July 2014 and is currently online [4]. Figure 26 visualises the pre-processed dataset. A significant period of a tracking error is noticeable (where $DHI \approx GHI$).

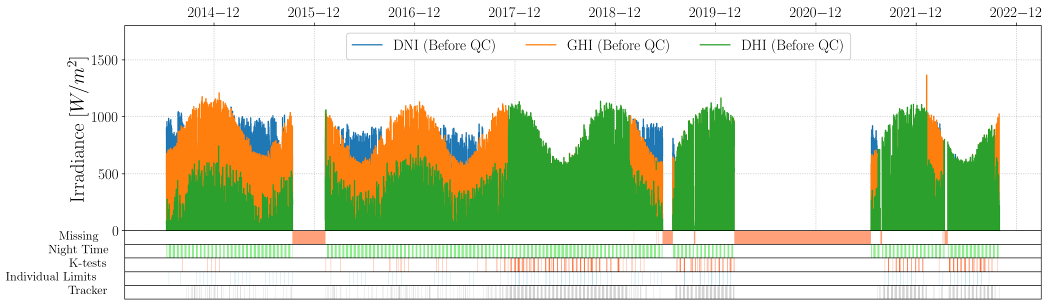

Figure 26. Pre-processed UNZ station with flagged datapoints.

4.1.23. UPR

The University of Pretoria (UPR) station is situated in Pretoria, South Africa, and its geographical properties are listed in Table 2. Data is available from September 2013 and is currently online [4]. Figure 27 visualises the pre-processed dataset where a tracking error where $DHI \approx GHI$ is noted for a short period.

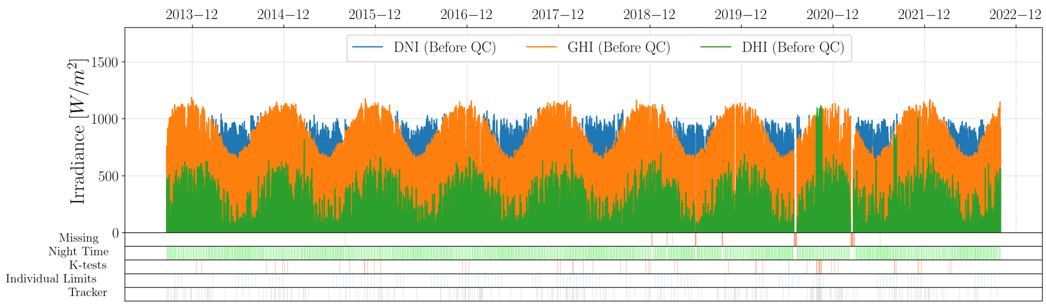

Figure 27. Pre-processed UPR station with flagged datapoints.

4.1.24. VAN

The GIZ Vanrhynsdorp (VAN) station is situated in Vanrhynsdorp, South Africa, and Table 2 describes its geographical properties. Data is available from August 2016 until July 2019, and the station is permanently offline [4]. Figure 28 visualises the pre-processed dataset. There is a tracking error visible where the $DHI \approx GHI$.

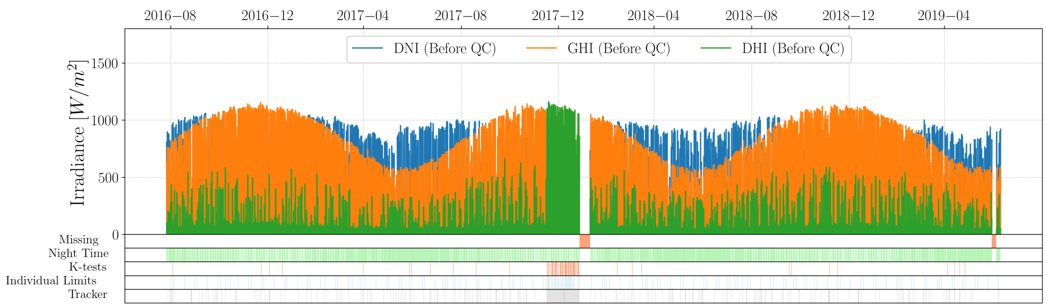

Figure 28. Pre-processed VAN station with flagged datapoints.

4.2. SAURAN Data Correlation Assessment

Table 4 shows the correlation matrix for the DNI of the different stations. The darker shades of blue in Tables 4–6 indicate a higher positive correlation. As discussed in Section 3.3, the relationships were analysed according to overlapping periods, as visualised in Figure 2. Noteworthy correlations are the CSIR and UPR stations, HLO and SUN stations and the KZH and KZW stations. These stations are very geographically close.

Table 4. DNI correlation matrix.

	CSIR	CUT	FRH	GRT	HLO	ILA	KZH	KZW	MIN	MRB	NMU	NUST	PMB	RVD	SALT	STA	SUN	SUT	UBG	UFS	UNV	UNZ	UPR	VAN
CSIR	1.00																							
CUT	0.37	1.00																						
FRH	0.20	0.34	1.00																					
GRT				1.00																				
HLO	0.01	0.18	0.19	0.44	1.00																			
ILA	0.25	0.37	0.05			1.00																		
KZH	0.33	0.29	0.24	0.19	0.07		1.00																	
KZW	0.28	0.28	0.21	0.26	0.07		0.90	1.00																
MIN	0.68	0.42	0.10			0.21			1.00															
MRB	0.27	0.09	0.49		0.36		0.18	0.15		1.00														
NMU	0.12	0.23	0.37	−0.04	0.29	0.27	0.15	0.12	0.05	0.51	1.00													
NUST	0.26	0.31	0.18		0.13	0.47	0.16	0.09	0.23	0.35	0.18	1.00												
PMB	0.41	0.40	0.10			0.21			0.38		0.18	0.18	1.00											
RVD	0.06	0.19	0.18	0.29	0.39		0.07	0.03		0.33	0.25	0.27	0.11	1.00										
SALT	0.18	0.36	0.28		0.37		0.11	0.12		0.69	0.34	0.27		0.40	1.00									
STA	0.01	0.07	0.01	−0.05	0.00		0.08	0.09		−0.12	−0.02	0.00		−0.10	0.04	1.00								
SUN	−0.05	0.11	0.13	0.29	0.92	0.21	0.03	0.01	−0.09	0.33	0.25	0.09	0.03	0.37	0.35	0.01	1.00							
SUT	0.03		0.22		0.67		0.07	0.10		0.50	0.28	0.11		0.44		−0.14	0.44	1.00						
UBG	0.40	0.42	0.21	0.26	0.11		0.27	0.24		0.29	0.12	0.23		0.16	0.28	0.00	0.07	0.15	1.00					
UFS	0.51		0.28	0.33	0.28		0.29	0.28		0.49	0.24	0.36		0.24	0.44	0.03	0.18	0.28	0.43	1.00				
UNV	0.36	0.15	0.13	0.16	−0.03	−0.07	0.14	0.14	0.41	0.13	0.02	0.16	0.14	0.03	0.13	−0.03	−0.08	0.15	0.39	0.20	1.00			
UNZ	0.35	0.23	0.18	0.16	0.07	0.03	0.58	0.57	0.21	0.12	0.11	0.12	0.52	0.05	0.13	0.01	0.00	0.24	0.31	0.25	0.28	1.00		
UPR	0.95	0.38	0.21	0.17	0.03	0.23	0.32	0.28	0.72	0.28	0.10	0.27	0.37	0.06	0.15	0.00	−0.04	0.06	0.51	0.38	0.39	0.38	1.00	
VAN	0.04	0.18	0.20		0.57		0.05	0.02		0.45	0.32	0.16		0.50	0.57	−0.04	0.58	0.66	0.08	0.19	−0.03	0.00	0.01	1.00

In Table 5, the correlation matrix for the GHI is presented. The GHI of most stations are relatively correlated; however, noteworthy-correlated groupings are CSIR-MIN-UPR, KZH-KZW-STA, CUT-MRB-SALT and VAN-HLO-RVD-SUN-SUT. The GHI correlation analysis by Farmer and Rix shows similar results [34].

Table 5. GHI correlation matrix.

	CSIR	CUT	FRH	GRT	HLO	ILA	KZH	KZW	MIN	MRB	NMU	NUST	PMB	RVD	SALT	STA	SUN	SUT	UBG	UFS	UNV	UNZ	UPR	VAN
CSIR	1.00																							
CUT	0.85	1.00																						
FRH	0.81	0.86	1.00																					
GRT				1.00																				
HLO	0.72	0.80	0.79	0.82	1.00																			
ILA	0.82	0.88	0.82			1.00																		
KZH	0.84	0.83	0.82	0.79	0.71		1.00																	
KZW	0.81	0.80	0.78	0.79	0.68		0.96	1.00																
MIN	0.94	0.84	0.81			0.81			1.00															
MRB	0.85	0.94	0.89		0.85		0.78	0.78		1.00														
NMU	0.78	0.85	0.90	0.86	0.84	0.86	0.79	0.78	0.78	0.89	1.00													
NUST	0.76	0.80	0.75		0.78	0.87	0.69	0.65	0.75	0.85	0.77	1.00												
PMB	0.86	0.83	0.85			0.76			0.85		0.79	0.68	1.00											
RVD	0.73	0.80	0.79	0.84	0.90		0.70	0.67		0.85	0.84	0.86	0.65	1.00										
SALT	0.80	0.86	0.85		0.88		0.74	0.73		0.95	0.89	0.84		0.89	1.00									
STA	0.85	0.83	0.84	0.76	0.74		0.96	0.94		0.83	0.82	0.70		0.72	0.73	1.00								
SUN	0.68	0.77	0.78	0.84	0.97	0.84	0.70	0.67	0.67	0.82	0.83	0.77	0.65	0.90	0.87	0.72	1.00							
SUT	0.73		0.79		0.95		0.72	0.70		0.88	0.85	0.78		0.91		0.75	0.91	1.00						
UBG	0.90	0.88	0.84	0.84	0.81		0.83	0.80		0.89	0.82	0.81		0.81	0.89	0.84	0.78	0.78	1.00					
UFS	0.90		0.83	0.87	0.83		0.83	0.83		0.91	0.85	0.80		0.81	0.90	0.85	0.80	0.81	0.88	1.00				
UNV	0.86	0.81	0.77	0.79	0.68	0.73	0.80	0.78	0.85	0.78	0.75	0.72	0.82	0.70	0.75	0.80	0.65	0.74	0.86	0.81	1.00			
UNZ	0.88	0.83	0.84	0.77	0.71	0.69	0.90	0.90	0.81	0.80	0.80	0.68	0.88	0.67	0.76	0.91	0.67	0.74	0.83	0.82	0.83	1.00		
UPR	0.97	0.86	0.82	0.79	0.72	0.81	0.83	0.81	0.93	0.86	0.79	0.76	0.85	0.73	0.80	0.84	0.69	0.71	0.90	0.86	0.86	0.85	1.00	
VAN	0.74	0.81	0.80		0.93		0.70	0.67		0.87	0.87	0.82		0.94	0.92	0.73	0.93	0.94	0.81	0.81	0.68	0.71	0.73	1.00

Table 6 shows the correlation matrix of the DHI of the SAURAN database. Based on the GHI and DNI correlation matrix, the CSIR is highly correlated with MIN and UPR. Further, the SUN and HLO stations are highly correlated, as well as the KZH-KZW-STA grouping.

Table 6. DHI correlation matrix.

	CSIR	CUT	FRH	GRT	HLO	ILA	KZH	KZW	MIN	MRB	NMU	NUST	PMB	RVD	SALT	STA	SUN	SUT	UBG	UFS	UNV	UNZ	UPR	VAN
CSIR	1.00																							
CUT	0.43	1.00																						
FRH	0.35	0.39	1.00																					
GRT				1.00																				
HLO	0.21	0.22	0.27	0.53	1.00																			
ILA	0.33	0.31	0.12			1.00																		
KZH	0.45	0.36	0.34	0.21	0.23		1.00																	
KZW	0.43	0.34	0.33	0.30	0.27		0.93	1.00																
MIN	0.85	0.57	0.31			0.32			1.00															
MRB	0.30	−0.10	0.49		0.22		0.16	0.18		1.00														
NMU	0.25	0.26	0.53	0.42	0.36	0.17	0.26	0.28	0.32	0.44	1.00													
NUST	0.39	0.33	0.23		0.24	0.45	0.29	0.26	0.36	0.19	0.21	1.00												
PMB	0.51	0.48	0.30			0.26			0.47		0.25	0.27	1.00											
RVD	0.19	0.17	0.20	0.23	0.35		0.18	0.17		0.22	0.21	0.24	0.09	1.00										
SALT	0.22	0.27	0.22		0.33		0.17	0.21		0.55	0.32	0.21		0.37	1.00									
STA	0.44	0.34	0.36	0.15	0.22		0.91	0.87		0.17	0.27	0.29		0.18	0.17	1.00								
SUN	0.16	0.19	0.24	0.28	0.94	0.19	0.15	0.16	0.17	0.18	0.27	0.18	0.13	0.29	0.33	0.16	1.00							
SUT	0.04		0.17		0.50		0.06	0.11		0.53	0.15	0.21		0.36		0.04	0.28	1.00						
UBG	0.62	0.50	0.30	0.27	0.21		0.40	0.43		0.23	0.25	0.41		0.19	0.26	0.39	0.14	0.15	1.00					
UFS	0.55		0.32	0.31	0.36		0.40	0.41		0.49	0.33	0.49		0.28	0.16	0.40	0.22	0.28	0.50	1.00				
UNV	0.51	0.30	0.34	0.23	0.15	−0.03	0.33	0.34	0.42	0.17	0.27	0.29	0.35	0.12	0.20	0.33	0.12	0.14	0.51	0.33	1.00			
UNZ	0.45	0.30	0.33	0.27	0.21	0.16	0.68	0.70	0.45	0.14	0.23	0.24	0.65	0.15	0.15	0.63	0.11	0.17	0.47	0.35	0.42	1.00		
UPR	0.98	0.44	0.35	0.30	0.19	0.33	0.46	0.46	0.86	0.31	0.26	0.41	0.48	0.18	0.18	0.44	0.15	0.12	0.68	0.48	0.52	0.52	1.00	
VAN	0.19	0.20	0.24		0.56		0.18	0.18		0.39	0.30	0.20		0.38	0.59	0.18	0.54	0.59	0.14	0.23	0.07	0.12	0.15	1.00

5. Discussion

5.1. Data Quality and Recommendations

Table 7 shows a summary and recommendation for each SAURAN measuring station. The following recommendation titles are: *recommended*, *use with caution* and *use with extreme caution*. In Table 7, the checkmarks (✓) and crossmark (✗) represent whether the criteria are met or unmet, respectively.

Table 7. SAURAN station recommendations.

Station	Summary			Recommendation
	Minimum One Complete Year	Minimal Missing Data	Currently Online [1]	
CSIR	✓	✗	✓	Recommended
CUT	✗	✗	✓	Use with caution
FRH	✓	✗	✗	Recommended
GRT	✓	✓	✗	Recommended
HLO	✗	✗	✗	Use with caution
ILA	✗	✓	✓	Use with caution
KZH	✓	✗	✗	Recommended
KZW	✗	✗	✗	Use with caution
MIN	✗	✗	✓	Use with caution
MRB	✗	✗	✗	Use with extreme caution
NMU	✗	✗	✗	Use with extreme caution
NUST	✓	✗	✓	Recommended
PMB	✗	✗	✓	Use with caution
RVD	✓	✓	✗	Recommended
SALT	✗	✗	✗	Use with extreme caution
STA	✗	✗	✗	Use with extreme caution
SUN	✓	✗	✓	Use with caution
SUT	✗	✓	✗	Use with extreme caution
UBG	✓	✗	✗	Use with caution
UFS	✓	✓	✗	Recommended
UNV	✓	✗	✓	Use with caution
UNZ	✓	✗	✓	Use with caution
UPR	✓	✓	✓	Recommended
VAN	✓	✗	✗	Use with caution

[1] Station status as viewed on 4 November 2022

Recommended states that the data should be used for applications after applying the QC procedure. These are datasets with minimal missing data and at least one complete year of data available. This assessment is based on all three irradiance components (DNI, DHI and GHI).

Use with caution suggests that the user should consider discrete data applications rather than time-series applications, as well as implementing shorter intervals of monitoring or modelling applications. The data is not ideal for long-term monitoring and performance modelling. These datasets have missing data or multiple incomplete years of data. As with the *Recommended* criteria, these are based on all three irradiance components.

Use with extreme caution is unusable for time-series and long-term monitoring/modelling applications due to significant gaps in the data after applying the QC procedure; however, it can be used for shorter-term applications. For time-series applications, removing incomplete months from the dataset is recommended. The user can also consider interpolating the data based on historical trends, but it can be difficult if a considerable amount of data is unavailable or faulty. At least one full year's data is required to represent the weather patterns. Incomplete annual hourly data can result in seasonally biased models. A concern worth noting is how to perform quality control on datasets that do not consist of all three irradiance measurements. Significant tracking error data can still be utilised when only the GHI is used.

Eight stations (CSIR, FRH, GRT, KZH, NUST, RVD, UFS and UPR) are recommended for most data-driven applications. Eleven stations (CUT, HLO, ILA, KZW, MIN, PMB, SUN, UBG, UNV, UNZ and VAN) are recommended with caution meaning that the user should apply a QC assessment and scrutinise the data. Five stations (MRB, NMU, SALT, STA and SUT) are recommended with extreme caution for data-driven applications. The network has ten online stations: CSIR, CUT, ILA, MIN, NUST, PMB, SUN, UNV, UNZ and UPR.

5.2. Irradiance Patterns

Figure 29 shows the distribution of the highly correlated stations across Southern Africa, highlighted in blue. The criteria for these clusters are based on the correlation of the irradiance patterns concerning each station.

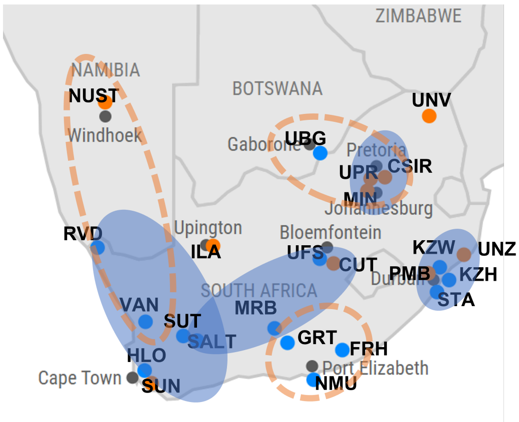

Figure 29. Highly and moderately correlated SAURAN stations indicated with blue and orange, respectively [4].

Due to its close location proximity to other nearby stations, it is assumed that the following clusters have highly correlated irradiance patterns:

- RVD-VAN-SUT-SALT-HLO-SUN;
- SUT-SALT-MRB-UFS-CUT;
- PMB-STA-KZW-KZW-UNZ;
- UPR-CSIR-MIN.

Some relationships could be closely linked based on their location. The following groupings, which have relatively highly correlated GHI but moderate correlations with DHI and DNI, are shown in Figure 29 in orange:

- GRT-FRH-NMU;
- NUST-RVD-VAN;
- UBG-UPR-CSIR.

The correlation assessment of stations can be used, for example, in the design, modelling and monitoring phases of renewable energy systems if on-site data is not available. The spatial relationships can also assist renewable energy providers in the spatial mapping of PV systems across the Southern African region.

6. Conclusions

This paper presented a case study of the SAURAN network by assessing an automated quality control procedure. The 24 stations from the SAURAN database up to 31 October 2022 were evaluated using a quality control procedure. An in-depth summary of the individual stations is discussed with usage recommendations, which can provide an overall assessment of the quality of the SAURAN database. The results showed that the automated QC methodology is sufficient in removing erroneous data. The proposed QC methodology has only been applied to scenarios where all three irradiances, DHI, GHI and DNI, are available. However, this is only feasible when expensive equipment is available for measuring and storing data. For a minimum procedure, the authors propose that if only one of the irradiance components is available, the applicable tests can only be applied to that irradiance. For example, the GHI-only tests are then applied to GHI.

The proposed QC procedure is especially beneficial in identifying faulty measurement equipment during a real-time monitoring application: consecutively flagged data, excluding night-time measurements, is a quick and easy way to identify system errors. The recommendation is that more than six consecutive hourly flagged datapoints within daylight hours can indicate that the measurement equipment is faulty and should be addressed.

The SAURAN network presents an incredible opportunity to increase the research development of PV systems in South Africa. However, many of the stations have been switched offline permanently. Reasons for switching off weather stations could be due to the lack of funding to keep these stations operational and the lack of workforce to keep the network up to date. This hinders the development of accurate models and research and development in PV systems in developing countries. South Africa is heavily investing in renewable energy generation, and advanced research will only decrease its dependency on fossil fuel-based power generation. Doing so will reduce the country's carbon footprint and its contribution to climate change. The SAURAN network significantly contributes to global research output, and the continuation of good-quality data in the network is of utmost importance.

Author Contributions: Conceptualisation, F.M.D.-D.; methodology, F.M.D.-D.; software, F.M.D.-D.; validation, F.M.D.-D.; formal analysis, F.M.D.-D.; investigation, F.M.D.-D.; resources, F.M.D.-D. and A.J.R.; data curation, F.M.D.-D.; writing—original draft preparation, F.M.D.-D.; writing—review and editing, F.M.D.-D. and A.J.R.; visualisation, F.M.D.-D.; supervision, A.J.R.; project administration, F.M.D.-D. and A.J.R.; funding acquisition, F.M.D.-D. and A.J.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gueymard, C.A.; Ruiz-Arias, J.A. Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance. *Sol. Energy* **2016**, *128*, 1–30. [CrossRef]
2. Daniel-Durandt, F.; Rix, A. The Automation of Quality Control for Large Irradiance Datasets. In Proceedings of the 2022 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Maldives, Maldives, 16–18 November 2022; pp. 1–6. [CrossRef]
3. Forstinger, A.; Wilbert, S.; Jensen, A.; Kraas, B.; Fernández-Peruchena, C.M.; Gueymard, C.; Ronzio, D.; Yang, D.; Collino, E.; Polo, J.; et al. Expert quality control of solar radiation ground data sets. In Proceedings of the ISES Solar World Congress, Virtual, 25–29 October 2021.
4. SAURAN. 2022. Available online: https://sauran.ac.za/ (accessed on 4 November 2022).
5. Fitzgerald, D. Quality Checking of Weather Data. Available online: https://sauran.ac.za/SAURAN%20Quality%20Checking%20Data.zip (accessed on4 November 2022).
6. Muneer, T.; Fairooz, F. Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases. *Build. Serv. Eng. Res. Technol.* **2002**, *23*, 151–166. [CrossRef]
7. Younes, S.; Claywell, R.; Muneer, T. Quality control of solar radiation data: Present status and proposed new approaches. *Energy* **2005**, *30*, 1533–1549. [CrossRef]
8. el Alani, O.; Ghennioui, H.; Ghennioui, A.; Saint-Drenan, Y.M.; Blanc, P.; Hanrieder, N.; Dahr, F.E. A Visual Support of Standard Procedures for Solar Radiation Quality Control. *Int. J. Renew. Energy Dev.* **2021**, *10*, 401–414. [CrossRef]
9. Brooks, M.; du Clou, S.; Van Niekerk, J.; Gauché, P.; Leonard, C.; Mouzouris, M.; Meyer, R.; Westhuizen, N.; Van Dyk, E.; Vorster, F. SAURAN: A new resource for solar radiometric data in Southern Africa. *J. Energy S. Afr.* **2015**, *26*, 2–10. [CrossRef]
10. Long, C.N.; Dutton, E.G. *BSRN Global Network Recommended QC Tests, V2.x*; Technical Report; PANGAEA: Bremerhaven, Germany, 2002. Available online: http://hdl.handle.net/10013/epic.38770.d001 (accessed on 6 April 2023).
11. Long, C.; Shi, Y. An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements. *Open Atmos. Sci. J.* **2008**, *2*, 23–37. [CrossRef]
12. Journée, M.; Bertrand, C. Quality control of solar radiation data within the RMIB solar measurements network. *Sol. Energy* **2011**, *85*, 72–86. [CrossRef]
13. Roesch, A.; Wild, M.; Ohmura, A.; Dutton, E.G.; Long, C.N.; Zhang, T. Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech.* **2011**, *4*, 339–354. [CrossRef]

14. Perez-Astudillo, D.; Bachour, D.; Martin-Pomares, L. Improved quality control protocols on solar radiation measurements. *Sol. Energy* **2018**, *169*, 425–433. [CrossRef]
15. Moreno, S.; Silva, M.; Santigosa, L. A proposed methodology for quick assessment of timestamp and quality control results of solar radiation data. *Renew. Energy* **2015**, *78*, 531–537. [CrossRef]
16. Yang, D.; Yagli, G.M.; Quan, H. Quality Control for Solar Irradiance Data. In Proceedings of the 2018 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia), Singapore, 22–25 May 2018; pp. 208–213. [CrossRef]
17. Sun, X.; Yang, X.; Wang, P. A Study of Models Combination for Global Clear Sky Irradiance Models. In Proceedings of the 2020 47th IEEE Photovoltaic Specialists Conference (PVSC), Calgary, AB, Canada, 15 June–21 August 2020; pp. 0724–0728. [CrossRef]
18. Mabhoko, M.; Fitzgerald, D.; Bekker, B. Comparing Long-Term Actual versus Simulated PV System Performance: A Review and Case Study. Technical Report, Centre for Renewable and Sustainable Energy Studies. 2020. Available online: https://www.sasec.org.za/papers2019/84.pdf (accessed on 6 April 2023).
19. Lubbe, F.; Maritz, J.; Harms, T. Evaluating the Potential of Gaussian Process Regression for Solar Radiation Forecasting: A Case Study. *Energies* **2020**, *13*, 5509. [CrossRef]
20. Masevhe, L.; Maluta, E.N. Assessment of photovoltaic power output using the estimated global solar radiation at Vuwani Science Resources Centre. *Cogent Eng.* **2022**, *9*, 2105031. [CrossRef]
21. Mutombo, N.M.A.; Mukuna, J.G.M.; Zawilska, E. Preliminary Analysis of Ground-Based Measured Solar Data From Sauran Sta Station at Umlazi, South Africa. *Int. J. Mech. Eng. Technol.* **2019**, *10*, 104–113.
22. Ineichen, P. A Study of Models Combination for Global Clear Sky Irradiance Models. In Proceedings of the 5th Southern African Solar Energy Conference, Cape Town, South Africa, 25-27 June 2018.
23. Ravele, T.; Sigauke, C.; Jhamba, L. Estimation of Extreme Quantiles of Global Horizontal Irradiance: A Comparative Analysis Using an Extremal Mixture Model and a Generalised Additive Extreme Value Model. *Math. Stat.* **2022**, *10*, 116–133. [CrossRef]
24. Gaonwe, T.; Hohne, P.; Kusakana, K. Optimal energy management of a solar-assisted heat pump water heating system with a storage system. *J. Energy Storage* **2022**, *56*, 105885. [CrossRef]
25. Mdluli, N.; Sharma, G.; Akindeji, K.; Narayanan, K.; Sharma, S. Development of Short Term Solar Radiation Forecasting Using AI Techniques. In Proceedings of the 2022 30th Southern African Universities Power Engineering Conference (SAUPEC), Durban, South Africa, 25–27 January 2022; pp. 1–6. [CrossRef]
26. Jacovides, C.; Tymvios, F.; Assimakopoulos, V.; Kaltsounides, N. Comparative study of various correlations in estimating hourly diffuse fraction of global solar radiation. *Renew. Energy* **2006**, *31*, 2492–2504. [CrossRef]
27. Le Roux, W. Optimum tilt and azimuth angles for fixed solar collectors in South Africa using measured data. *Renew. Energy* **2016**, *96*, 603–612. [CrossRef]
28. McArthur, L. Baseline Surface Radiation Network (BSRN)-Operation Manual Version 2.1. 2005. Available online: https://epic.awi.de/id/eprint/30644/1/McArthur_2005.pdf (accessed on 19 January 2023).
29. Larrañeta, M.; Fernandez-Peruchena, C.; Silva-Pérez, M.; Lillo-Bravo, I. Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility. *Sol. Energy* **2018**, *162*, 573–584. [CrossRef]
30. Sun, X.; Bright, J.M.; Gueymard, C.A.; Bai, X.; Acord, B.; Wang, P. Worldwide performance assessment of 95 direct and diffuse clear-sky irradiance models using principal component analysis. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110087. [CrossRef]
31. Nwokolo, S.C.; Obiwulu, A.U.; Ogbulezie, J.C.; Amadi, S.O. Hybridization of statistical machine learning and numerical models for improving beam, diffuse and global solar radiation prediction. *Clean. Eng. Technol.* **2022**, *9*, 100529. [CrossRef]
32. Geuder, N.; Wolfertstetter, F.; Wilbert, S.; Schüler, D.; Affolter, R.; Kraas, B.; Lüpfert, E.; Espinar, B. Screening and Flagging of Solar Irradiation and Ancillary Meteorological Data. *Energy Procedia* **2015**, *69*, 1989–1998. [CrossRef]
33. Walpole, R.E.; Myers, R.H.; Myers, S.L.; Ye, K. *Probability & Statistics for Engineers and Scientists*, 9th ed.; Pearson Education: Upper Saddle River, NJ, USA, 2012.
34. Farmer, W.; Rix, A. Mapping the spatial perturbations seen by the power system network due to intermittent renewable energy sources. In Proceedings of the 2021 Southern African Sustainable Energy Conference, Stellenbosch, South Africa, 17–19 November 2021.