

Article

Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa

Brighton Mabasa ^{1,2,*} , Meena D. Lysko ³  and Sabata J. Moloi ²

¹ Research & Development Division, South African Weather Service, Pretoria 0001, South Africa

² Department of Physics, University of South Africa, UNISA Preller Street, Muckleneuk, Pretoria 0001, South Africa; moloisj@unisa.ac.za

³ Move Beyond Consulting (Pty) Ltd., Pretoria 0081, South Africa; meena@lysko.com

* Correspondence: brighton.mabasa@weathersa.co.za; Tel.: +27-12-367-6088

Abstract: This study validates the hourly satellite based and reanalysis based global horizontal irradiance (GHI) for sites in South Africa. Hourly GHI satellite based namely: SOLCAST, Copernicus Atmosphere Monitoring Service (CAMS), and Satellite Application Facility on Climate Monitoring (CMSAF SARA) and two reanalysis based, namely, Fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5) and Modern-Era Retrospective Analysis for Research and Applications (MERRA2) were assessed by comparing in situ measured data from 13 South African Weather Service radiometric stations, located in the country's six macro climatological regions, for the period 2013–2019. The in situ data were first quality controlled using the Baseline Surface Radiation Network methodology. Data visualization and statistical metrics relative mean bias error (rMBE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and the coefficient of determination (R^2) were used to evaluate the performance of the datasets. There was very good correlation against in situ GHI for the satellite based GHI, all with R^2 above 0.95. The R^2 correlations for the reanalysis based GHI were less than 0.95 (0.931 for ERA5 and 0.888 for MERRA2). The satellite and reanalysis based GHI showed a positive rMBE (SOLCAST 0.81%, CAMS 2.14%, CMSAF 2.13%, ERA5 1.7%, and MERRA2 11%), suggesting consistent overestimation over the country. SOLCAST satellite based GHI showed the best rRMSE (14%) and rMAE (9%) combinations. MERRA2 reanalysis based GHI showed the weakest rRMSE (37%) and rMAE (22%) combinations. SOLCAST satellite based GHI showed the best overall performance. When considering only the freely available datasets, CAMS and CMSAF performed better with the same overall rMBE (2%), however, CAMS showed slightly better rRMSE (16%), rMAE (10%), and R^2 (0.98) combinations than CMSAF rRMSE (17%), rMAE (11%), and R^2 (0.97). CAMS and CMSAF are viable freely available data sources for South African locations.



Citation: Mabasa, B.; Lysko, M.D.; Moloi, S.J. Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa. *Geomatics* **2021**, *1*, 429–449. <https://doi.org/10.3390/geomatics1040025>

Academic Editor: Naser El-Sheimy

Received: 9 September 2021

Accepted: 3 November 2021

Published: 5 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: satellite; reanalysis; global horizontal irradiance; SOLCAST; Copernicus Atmosphere Monitoring Service (CAMS); Satellite Application Facility on Climate Monitoring (CMSAF); fifth generation European Center for Medium-Range Weather Forecasts atmospheric reanalysis (ERA5); Modern-Era Retrospective Analysis for Research and Applications (MERRA2)

1. Introduction

Solar radiation is the electromagnetic radiation or energy emitted from the surface of the Sun because of the fusion of atoms inside the sun [1]. Global horizontal Irradiance (GHI) is the electromagnetic radiation that reaches the Earth's horizontal surface after passing through the atmosphere and is the sum of direct normal irradiance (DNI), which is the incident radiative flux on the surface without interacting with the atmosphere and diffuse horizontal irradiance (DIF), which is because of the scattering of radiation by the atmospheric constituents [2]. Accurate knowledge of GHI is important for the technical and economic evaluation of solar energy technologies [3–8] and in the development and validation of empirical models [6]. Amongst the myriad of applications, GHI is important

in climate change and environmental studies, agricultural sciences, hydrology, atmospheric research [6,7], and in astronomy [6]. GHI is also important in assessing ultraviolet effects on health as well as in material science [6], and in the development of a typical meteorological year of a country. Therefore, obtaining true solar measurements at a location is important.

GHI measurements taken from radiometric stations using at least a good quality (broadband) operational pyranometer remain the most accurate way to collect GHI data. However, GHI monitoring stations are sparse and expensive to install and maintain. As a result, data are only available for a limited number of locations [6,8–14].

Alternative sources of GHI data include models such as the Ångström–Prescott model. Though the model with calibrated coefficients can be used to accurately estimate GHI data [13–15], the drawback might be the unavailability of sunshine duration data that are needed as an input at some areas. Another limitation of the Ångström–Prescott model is that the highest possible temporal resolution of estimated GHI is the daily average, so the model is not capable of estimating hourly averages. According to Žák et al. [16], GHI data can also be generated by interpolation of measured in situ GHI. The drawback of the interpolation method is the biases that are introduced by interpolation, and the additional errors introduced by using sparsely distributed in situ stations.

Given that GHI datasets are critical for better understanding of wider coverage of solar radiation [10], satellite and reanalysis based GHI datasets can be used to provide reliable alternative GHI data and compensate for the scarcity of monitoring stations by increasing the density of GHI data. The satellite or reanalysis-based datasets must first be validated by using GHI data from a good quality pyranometer [9,11,12,16–20] to obtain proof of their reliability before they are used in different applications.

The satellite based, reanalysis based, and in situ measurements differ in spatial and temporal resolutions. This creates challenges when using satellite based and reanalysis based GHI datasets as alternative GHI sources [21]. To address these challenges, there has been an improvement in the spatio-temporal resolution of satellites and reanalysis-based datasets in the past few years [21,22]. According to Slater (2016) [22], the improvement was due to the advances in modeling and data assimilation systems. However, there are still challenges due to limited spatio-temporal coverage of observation data in some areas as required by the models, for example, in South Africa, there is only one Baseline Surface Radiation Network (BSRN) station. Baseline Surface Radiation Network (BSRN) was established in 1992, and is a centralized database that archives one minute temporal-resolution in situ radiation measurements from 59 stations worldwide. The archived data are used for the validation of satellite data and improvement in radiative transfer calculations in climate models [23].

The algorithms that are used to convert satellite images to estimate GHI data depend on inputs of meteorological parameters (albedo, cloud thickness, aerosols, water vapor, and ozone content). When the parameters have not been measured in some areas, estimated or monthly climatologies are used. Climatological values might not fully represent changes in atmospheric constituents, and as a result, introduce errors in the estimated GHI data when used as inputs.

This study aims to contribute to the reviewed literature by quantifying the errors between the in situ measured GHI and estimated gridded datasets such as to validate satellite-based datasets (SOLCAST, CAMS, and CMSAF SARA) and reanalysis-based datasets (ERA5 and MERRA2) relative to quality controlled in situ data from 13 reference stations managed by the South African Weather Services (SAWS). The validation was conducted on an hourly temporal scale over all six macro climatic zones of South Africa. This can be applied in studies for overall local accuracy of the datasets to be evaluated.

Validated datasets could be used to estimate GHI in the long-term and over a wide spatial resolution in South Africa. This will enable climate studies, which is generally not possible with ground observed data because there are no continuous long-term records covering decades and covering all areas of the country. The validated datasets could also be used as an additional quality control parameter of the measured data. The bias information

from different sources of GHI in different areas of the country can be used as a basis for bias correction. The bias corrected data sources could also be merged with measured in situ data by applying interpolation with external drift kriging to produce most accurate GHI maps than when using an individual data source, as shown by [24,25].

2. Literature Review

GHI varies from point to point in an area and from time to time. Validation studies have been carried out worldwide to quantify the errors between in situ measured GHI and gridded data sources and is an ongoing process. This is because the algorithms and inputs that are used to generate the gridded data sources are continuously evolving. The findings, the challenges, and recommendations of some of the studies that have been carried out are summarized in Table 1.

The study by Bright [26] validated hourly SOLCAST GHI data, which is satellite based, against 48 BSRN stations by considering individual stations and grouping stations into global climates. There was a good agreement across all climates (Table 1). The author emphasized that a comparative study against freely available alternative satellite based and reanalysis GHI datasets is necessary to gauge its performance. Yang and Bright [17] validated six satellite based and two reanalysis based GHI data sources using hourly data from 57 BSRN stations spread across all continents. They found that reanalysis products did not perform well compared to satellite-based products, since they overestimated irradiance in most sites (Table 1). The authors found that each gridded product had a site where it performed better, so testing different available gridded datasets at different sites was emphasized. It was also shown that SOLCAST, which is a commercial satellite-based product, did not outperform the freely available products at all sites, but overall, it was the best performing product.

Merchand et al. [27] validated the CAMS satellite-based dataset against hourly GHI from 16 stations from the Royal Meteorological Institute KNMI in the Netherlands, a temperature climate without a dry season and warm winter. The reference stations were located inland and along the coast. They found that CAMS satellite-based datasets could very well estimate the hourly-to-hourly variation in GHI (Table 1). The biases reported were suspected to be due to the McClear model, described by Lefèvre et al. in [28], and an input to the Heliosat-4 method, described by Qu et al. in [29], but failed to identify actual cloud free conditions in some regions due to errors in aerosols used as inputs to the McClear model. Negative rMBEs were reported from stations located along the southern coast and positive rMBE were reported from inland stations (Table 1). The authors concluded that the data were of low moderate quality and disagreed with some studies, for example, in [9], which found the data to be of moderate quality. The disagreement was suspected to be due to the low number of cloud free days in the Netherlands.

Thomas et al. [30] used the CAMS satellite based against hourly GHI data from 42 stations in Brazil to assess the quality of the satellite datasets in Brazil. Prior to their validation exercise, the datasets were only validated in Europe and North Africa. They found that the CAMS satellite-based dataset could estimate hourly GHI (Table 1). The biases increased with the viewing angle of the satellite. Additionally, the biases were high in tropical climates because of high humidity. CAMS overestimated GHI, but it was deemed suitable for solar energy applications.

Ameen et al. [20] validated CAMS satellite-based dataset in Northeast Iraq, which has a complex topography, using hourly observation data from nine stations. It was found that the CAMS satellite-based dataset captured the spatio-temporal trends of the measured data in clear sky, cloudy sky, and all sky conditions (combination of clear sky and cloud conditions). The performance was found to be better for clear sky and all sky conditions, but worse in cloudy sky conditions. For all sky, the results were as shown in Table 1. The dataset was recommended for use in solar resource applications. Further validation of the CAMS satellite-based data in other areas was recommended.

Table 1. Summary of the validation results from the literature review (the rMBE and rRMSE results were rounded off to the nearest whole number).

Study	Dataset	rMBE	rRMSE	R^2
Bright [26]	SOLCAST (climates)	−0.1% to 1%	-	-
	SOLCAST (individual stations)	−18% to 6%	6% to 44%	0.42 to 0.97
Yang and Bright [17]	SOLCAST	−5% to 3%	9% to 30%	-
	CAMS	−14% to 30%	9% to 45%	-
	CMSAF	27% to 40%	10% to 80%	-
	ERA5	15% to 72%	8% to 120%	-
	MERRA2	20% to 76%	10% to 128%	-
Merchand et al. [27]	CAMS	−4% to 10%	20% to 28%	0.94 to 0.97
	CAMS (inland)	1% to 10%	20% to 28%	0.94 to 0.97
	CAMS (southern coast)	−4% to −3%	23% to 24%	0.96
Thomas et al. [30]	CAMS	2% to 16%	17% to 35%	0.89 to 0.97
Ameen et al. [20]	CAMS (all sky conditions)	−5% to 5.3%	14% to 20%	0.92 to 0.96
Marchand et al. [9]	CAMS	−4% to 7%	11% and 21%	0.92 to 0.98
Trolliet et al. [31]	CMSAF SARA	2% to 12%	-	0.92 to 0.98
	CAMS	2% to 8%	-	0.93 to 0.97
	ERA5	−2% and 5%	-	0.88 to 0.93
	MERRA2	−10% to 4%	-	0.83 to 0.91

Marchand et al. [9] validated the CAMS satellite-based dataset against hourly observed GHI from five stations located in northern and central parts of Morocco to investigate how the bias of the stations located in the same climate varied. They found that the dataset was capable of estimating GHI (Table 1) and CAMS slightly overestimated GHI. Overall, there was a variation from site to site, but the dataset was recommended as a reliable source of estimated GHI data.

Trolliet et al. [31] validated satellite-based datasets (CMSAF SARA and CAMS) and reanalysis based (ERA5 and MERRA2) against hourly observed GHI data in the tropical Atlantic Ocean collected by five buoys of the Prediction and Research Moored Array in the Tropical Atlantic (PIRATA) network. The PIRATA network was established in the mid-1990s to study ocean–atmosphere interactions in the tropical Atlantic that affect regional weather and climate variability [32]. It was found that satellite-based datasets performed well (Table 1). Reanalysis datasets were found to report clear sky conditions while the actual conditions were cloudy and vice-versa. The reanalysis cloud parameterization scheme was suspected to be weak and contributed to large biases reported. ERA5 and MERRA2 were not recommended to estimate hourly temporal variability of GHI in the tropical Atlantic Ocean, but that they could be used in annual variability studies. The limitations of the study were that some buoys gathered African dust and the pyranometers may have been affected by the motion of the buoys; these factors might have contributed to the biases reported.

Since the satellite and reanalysis gridded datasets considered were generated by applying different algorithms, a variation in performance was expected in different stations, hence the need for quantifying errors in different locations.

3. Materials and Methods

3.1. Observation or Reference GHI Data

The South African Weather Service (SAWS) manages a radiometric network of 13 stations that are distributed in all six macro-climate regions in South Africa (with location and

climate as shown in Figure 1 and given in Table 2). The color shaded climate zones shown in Figure 1 are based on SAWS macro classification and the climate codes, which are indicated in square brackets, are the “micro” zones based on the Council for Scientific and Industrial Research (CSIR) Köppen–Geiger climate classification (KGCC), as given by Conradie in [33].

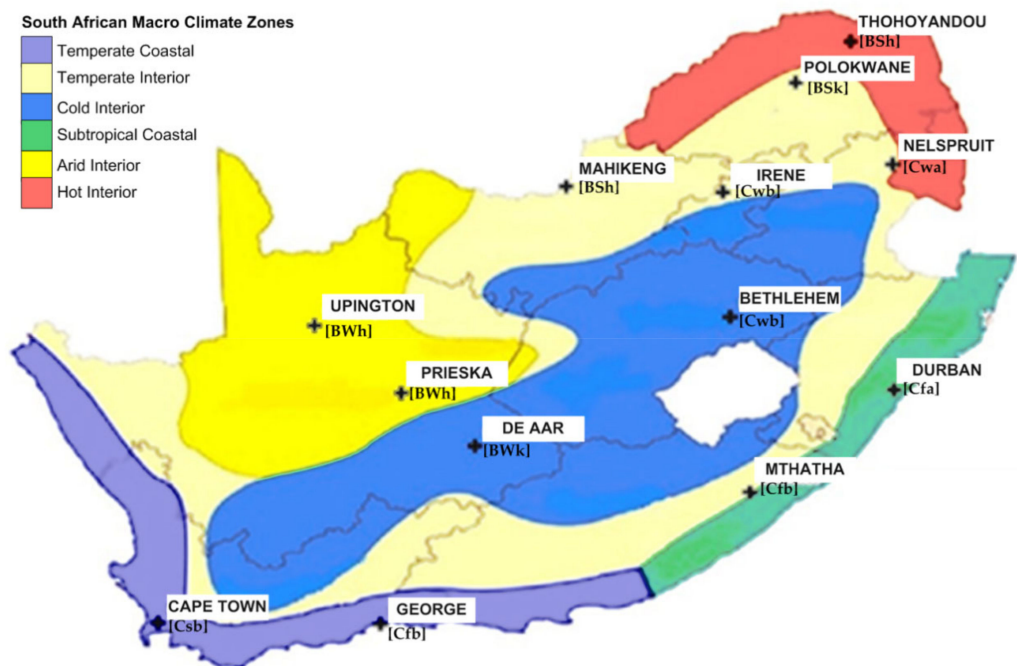


Figure 1. A map showing the location of the South African Weather Service’s radiometric station, SAWS macro climate zones, and with CSIR Köppen–Geiger “micro” climate in square brackets (Adapted from [15,34,35]).

Table 2. South African Weather Services radiometric stations with Köppen–Geiger climate classification (KGCC), altitude, latitude, climatic zones, average number of clear sky days per year (percentage of clear sky days per year), annual aggregated diffuse fraction, humidity, and the percentage of data outliers removed per station (colors are described in Table 3).

Station	KGCC	Altitude (m)	Latitude (°)	GHI Observation Period	Clear Sky Days	Diffuse Fraction	Humidity	Outliers (%)
Upington	BWh	848	−28.48	1 February 2014 to 30 November 2019	97 (27)	0.18	35.4	4.47
Prieska	BWh	989	−29.68	1 September 2013 to 31 August 2019	78 (21)	0.18	38	3.88
De Aar	BWk	1284	−30.67	1 May 2014 to 31 December 2019	58 (16)	0.2	44.5	4.23
Bethlehem	Cwb	1688	−28.25	1 January 2015 to 31 December 2019	43 (12)	0.31	59.1	6.57
Irene	Cwb	1524	−25.91	1 March 2014 to 31 December 2019	62 (17)	0.3	54.9	2.99
Mahikeng	BSh	1289	−25.81	1 January 2016 to 31 December 2019	77 (21)	0.24	43.9	6.4
Polokwane	BSk	1233	−23.86	1 March 2015 to 31 December 2019	49 (13)	0.31	58.2	5.12
Nelspruit	Cwa	870	−25.39	1 February 2014 to 31 December 2019	39 (11)	0.4	62	5.85
Thohoyandou	BSh	619	−23.08	1 March 2015 to 31 October 2017	50 (14)	0.34	60.8	4.06
Mthatha	Cfb	744	−31.55	1 July 2014 to 31 December 2019	19 (5)	0.33	68.1	4.98
Durban	Cfa	91	−29.61	1 March 2015 to 31 December 2019	20 (5)	0.39	72.8	5.34
Cape Point	Csb	86	−34.35	1 February 2015 to 31 December 2019	12 (3)	0.34	77.2	4.96
George	Cfb	192	−34.01	1 January 2015 to 31 December 2019	11 (3)	0.36	79.2	2.75

Table 3. The coding representing the different levels of clear sky days, diffuse fraction, and humidity.

Parameter/Colour	Humidity (H)	Clear Sky Days (CL)	Diffuse Fraction (DF)
Green	$H < 50\%$	$CL > 20\%$	$DF < 0.2$
Yellow	$50\% < H < 60\%$	$15\% < CL < 20\%$	$0.2 < DF < 0.25$
Blue	$60\% < H < 70\%$	$10\% < CL < 15\%$	$0.25 < DF < 0.32$
Orange	$H > 70\%$	$CL < 5\%$	$DF > 0.32$

Information on each of the stations (including site parameters, instrumentation, datalogging, and quality control) is given in [15,34,35]. The area of study included all 13 South African Weather Service radiometric sites.

In Table 2, the cells for clear sky days, diffuse fraction, and humidity are color coded. The coding represents different levels of clear sky days, diffuse fraction, and humidity. The meteorological information is used to explain the reason behind the performance of satellite-based and reanalysis datasets in different stations. The color-coding limits and distributions are summarized in Table 3.

ERA5 hourly cloud data from 2013–2019 were used to calculate the average clear sky days per year for each station, and this information was used to evaluate the performance of datasets based on the frequency of cloud occurrences. Diffuse fraction and relative humidity information for the study site were from Mabasa et al. (2018) [35].

3.2. Reanalysis Data

The choice of reanalysis datasets was based on the general free availability. The spatial and temporal characteristics as well as the level of accessibility and applicable regions of the reanalysis datasets are given in Table 4. Reanalysis datasets are generated by assimilating historical observation data from various platforms (ground observation, satellites, ships, and aircrafts) and numerical weather prediction models using a consistent algorithm. Reanalysis datasets have an advantage of global spatial coverage and long-time series [36].

ERA5, as described by Hersbach et al. (2020) [37], is the fifth generation European Centre for Medium-Range Weather Forecasts (ECMWF, Reading, United Kingdom) atmospheric reanalysis of the global climate. ERA5 has a spatial resolution of $0.25^\circ \times 0.25^\circ$ and has an hourly time resolution. It uses a new advanced model cycle for Integrated Forecasting System (IFS Cycle 41r2) for data assimilation, which increased the computational efficiency and forecast accuracy. ERA5 uses climatological aerosol information, and the Global Ozone Chemistry Aerosol Radiation and Transport (GEOCART) stratospheric sulfate aerosol from volcanic eruptions is also included [37]. The dataset covers a period from 1979 to the present, hourly data are available in Joule per square meter (J/m^2), so to obtain the hourly values in watt per square meter (W/m^2), values should be divided by 3600 (number of seconds in one hour).

The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA2), which was introduced by Gelaro et al. (2017) in [38], assimilates space-based observations of aerosols and represents their interactions with other physical processes in the climate system. MERRA2 products are generated using the GEOS 5.12.4 model, which uses real time aerosols as inputs. MERRA2 data are available from 1980 to the present with two months delay, has a spatial resolution of $0.5^\circ \times 0.625^\circ$, and has an hourly resolution for surface irradiance variables [38].

3.3. Satellite-Based Datasets

The choice of satellite-based datasets was based on the general free availability and availability for research purposes. Information on the satellite datasets for this study is given in Table 4. The relationship between the satellite images and the actual ground GHI was established with an algorithm that combines cloud information, which is generated from a clear sky model and cloud index. Aerosol, water vapor, linke turbidity, and

ozone information were also added as inputs to an algorithm, which will then generate estimated GHI at the surface. The technique is also described in the literature including Ameen et al. (2018) [20].

The Surface Solar Radiation Data Record—Heliosat Edition 2 (SARAH) dataset, as described by Pfeifroth et al. (2019) [39] is provided by the EUMETSAT Satellite Application Facility on Climate Monitoring (CMSAF) and covers the time period from 1983 to present, with the temporal resolution ranging from 30 min instantaneous values. Amongst the other products, the dataset provides GHI. CMSAF SARAH datasets are derived from the geostationary METEOSAT satellite service of the first and second generation, which are geostationary over Europe, Africa, and a small part of South America. In this case, the data were retrieved by using the Heliosat method to estimate the cloud index, clear sky radiative transfer model, and several climatological parameters (precipitable water vapor, monthly AOD climatology, monthly ozone climatology, and ground albedo). CMSAF SARAH data have a higher stability in early years due to the removal of erroneous satellite images during the transition from the first to the second generation METEOSAT satellite. CMSAF, as also given in Table 4, has a spatial resolution of $0.05^\circ \times 0.05^\circ$ [39].

Table 4. Summary of the satellite-based and reanalysis based datasets.

Data	Data Derived from	Time Period	Spatial Resolution	Temporal Resolution	Data Availability	Region Available
SOLCAST [40]	satellite	2007 to present	1–2 km	1 h	Not free	Almost Global (except Polar regions and oceans)
CMSAF SARAH [37]	satellite	1983 to present	$0.05^\circ \times 0.05^\circ$ (5 km)	1/2 h, 1 day	Free	Europe, Africa, and a small part of South America
CAMS [39]	satellite	2004 to (current day—2 days)	* Interpolated to a point of interest	1 min, 15 min, 1 h, 1 day, 1 month	Free	Europe, Africa, Middle East, Eastern part of South America and Atlantic Ocean
ERA5 [35]	reanalysis	1979 to present	$0.25^\circ \times 0.25^\circ$ (31 km)	1 h	Free	Global
MERRA2 [36]	reanalysis	1980 to (present—2 months)	$0.625^\circ \times 0.5^\circ$ (50 km)	1 h	Free	Global

* (3–5 km in Southern Africa).

CMSAF SARAH was validated using twelve BSRN stations from three continents, eight stations from Europe, one station from South America, and three from Africa (South Africa, Algeria, and Namibia). The validating irradiance datasets were first quality controlled using BSRN methodology and outliers were discarded [40].

The Copernicus Atmosphere Monitoring Service (CAMS) radiation service is part of the Copernicus Program, an Earth observation program coordinated and managed by the European Commission in partnership with the European Space Agency. The CAMS radiation service is available for free via the CAMS [41] and solar radiation data (SoDa) [42]. It makes use of the Heliosat-4 method [29], which models the radiative transfer in the atmosphere to compute the solar radiation parameters. The McClear model [28] is used to estimate clear sky irradiance. The AVHRR Processing scheme Over cLOUDs, Land and Ocean (APPOLLO) method is used to process the satellite images from the German Aerospace Center database to yield information (cloud coverage, cloud level, and cloud type) for each pixel (3 km at nadir) every 15 min [29]. Ground albedo from Moderate Resolution Imaging Spectroradiometer (MODIS) is used, as given in Qu et al. (2017) [29]. The CAMS radiation service provides a time series of GHI with a spatial coverage of -66° to 66° in both latitudes and longitudes (i.e., North to South and East to West directions). The data are interpolated to the point of the user's interest. The time coverage of data was from 1 February 2004 to date with a two day delay and with a temporal resolution

that ranged from 1-min average, 15-min averages, hourly averages, daily averages, and monthly averages [29,41,42].

The CAMS satellite dataset is regularly validated using global BSRN and non BSRN stations. The validating in situ irradiation datasets were first quality controlled using the BSRN methodology and outliers were discarded. The recent validation process in 2020 used 32 in situ stations and only one station from Southern Africa, Namibia (Gobabeb), was used in the recent validation process [43].

SOLCAST is a commercial company [17,26,44]. The SOLCAST method estimates solar irradiance from satellites by detecting cloud cover and characterizing cloud cover in terms of its impact on solar radiance, modeling the available solar irradiance under clear skies and then combining the estimate of the amount of solar irradiance reaching the Earth's surface after it passes through the clouds. In Africa, the EUMETSAT Meteosat satellite and REST2v5 clear sky model [45] with MERRA2 reanalysis inputs (for metadata) were used in SOLCAST GHI data estimation. The temporal resolution ranged from basic time series (hourly averages) and alternative time series (5 min, 10 min, 15 min, and 30 min). One-minute data are also available on request. The data are available from January 2007 to date with a seven day delay, through the SOLCAST website [44]. SOLCAST provides global coverage of the data, except for ocean and polar regions, with a spatial resolution of 1–2 km [17,26,44]. According to the website on SOLCAST validation and accuracy [46], SOLCAST has considered 46 of the BSRN sites for validation of GHI and reports a maximum bias deviation of 2.01%. The BSRN sites for steep mountain areas, oceanic, and polar sites were excluded in the SOLCAST validation.

3.4. Methodology

The methodology used in this study is summarized in the flow chart in Figure 2. The methodology consists of the following steps: (1) Preprocessing and quality control of in-situ GHI data; (2) Averaging one minute to 15 min; (3) Averaging four slots of 15 min to obtain hourly averages; (4) Gathering and preprocessing of satellite-based and reanalysis datasets; (5) Matching common time steps of in situ GHI and satellite/reanalysis-based datasets; (6) Calculation of hourly zenith angles and removing datasets that fall on points where the zenith angle is greater than 90 degrees; and (7) Calculation of statistical metrics.

3.5. Pre-Processing of Observation Data and Validation Process

One-minute average GHI data recorded from each of the 13 stations using a CMP11 pyranometer from Kipp and Zonen were pre-processed, subjected to “physical possible” limit check, which is aimed at detecting extremely large errors in the radiation data and “extremely rare” values checks of the BSRN QC test [23,47,48], and the outliers were removed. The minute data that passed the BSRN QC test were converted to 15 min averages; then hourly averages were calculated from four slots of 15-min averages. This methodology was used in [5,15,30,34,35,48,49]. All the datasets were synchronized to South African standard time (SAST) to avoid the misalignment of time series datasets.

Hourly average GHI observation data were also further processed by removing all the data points recorded on hours when the solar zenith angle was greater than 90°. Hourly averages were then averaged to daily average values. Daily averages were then subjected to HelioClim model QC, described by Geiger et al. in [50]. Hourly data on days that failed the HelioClim QC test were removed before any further analysis. The percentage of outliers removed from each station is given in the last column in Table 2.

Solar zenith angles were calculated using the solar position algorithm (SPA) on Python PV_LIB [51,52]. Hourly average values were then compared to corresponding hourly average CAMS, CMSAF SARA, SOLCAST, ERA5, and MERRA2.

SOLCAST, ERA5, MERRA2, and CAMS were sourced as hourly averages while CMSAF SARA was sourced as instantaneous 30 min averages and then 2-time steps were averaged to obtain hourly averages.

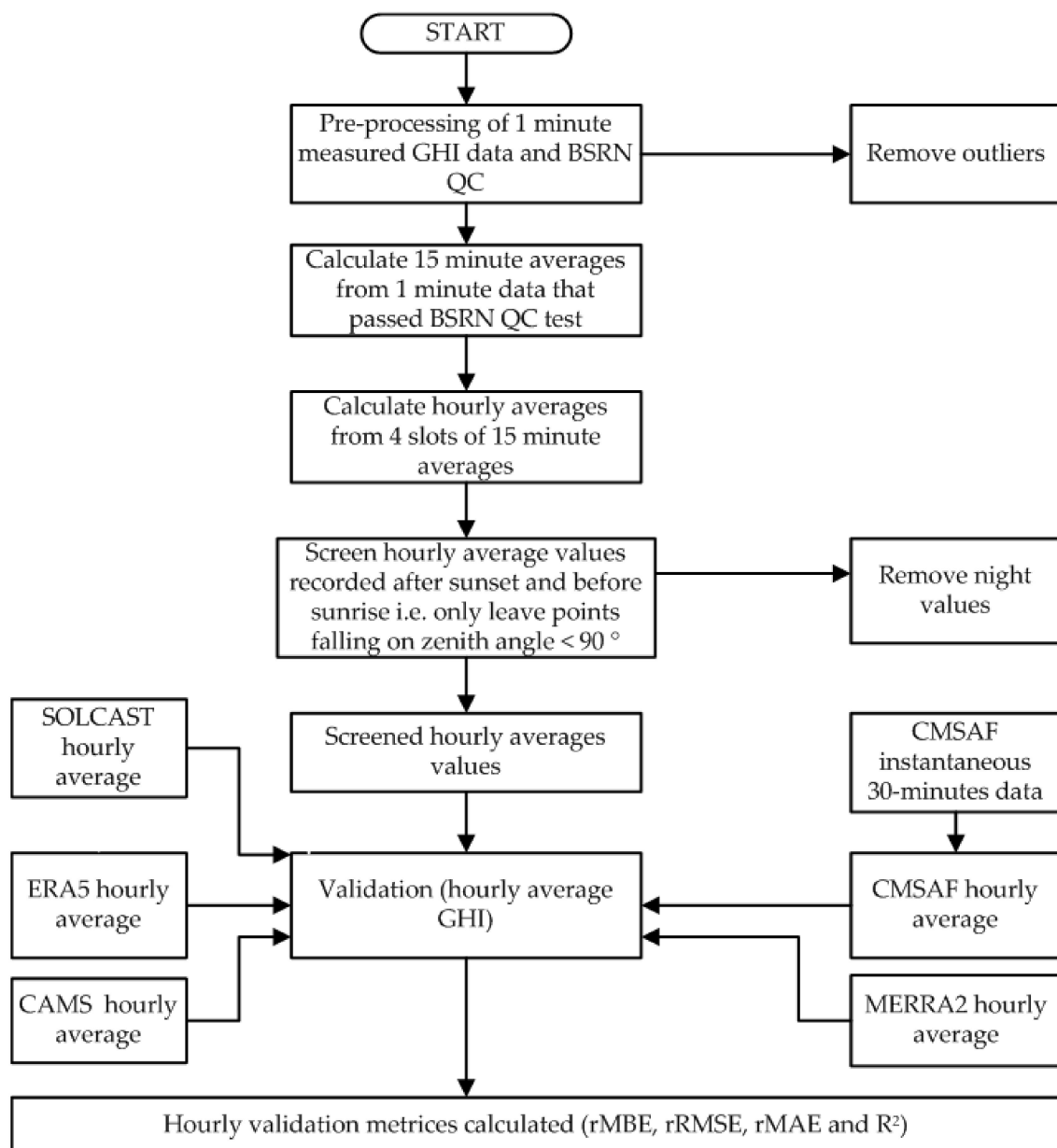


Figure 2. The flowchart summarizes the approach used from data preprocessing to data validation.

3.6. Statistical Metrics

The statistical metrics that were used to quantify the difference between hourly estimated and hourly measured GHI are relative mean bias error (rMBE), relative root mean square error (rRMSE), relative mean absolute error (rMAE), and the coefficient of determination (R^2). These statistical metrics are given in the literature [53,54] and have also been described and applied in the authors' previous studies [15,34].

3.7. Most Feasible Gridded Dataset

The most feasible option from the satellite-based and reanalysis datasets at each of the 13 stations is determined by first considering the best performing model for each metric. The most feasible option at a station is then the model with the maximum count across all the metrics. A numerical relative rating of the optimal metric is calculated as the maximum count relative to the four metrics (rMBE, rRMSE, rMAE, and R^2).

4. Results

The threshold or the range of the statistical metrics that are used to benchmark the optimal and/or the best applicable hourly gridded products based on their performance when compared to measured hourly data in the study area are summarized in (Table 5). The interval metrics were based on the distribution of the results.

Table 5. Range of the statistical metrics used to benchmark the applicability of the gridded datasets. Green represents excellent performance, blue good performance, and orange poor performance).

Skill	rMBE	rRMSE	rMAE	R^2
Poor	$rMBE > \pm 10 \%$	$rRMSE > 20\%$	$rMAE > 15\%$	$R^2 < 90$
Good	$ \pm 5 \% < rMBE \leq \pm 10 \%$	$10\% < rRMSE \leq 20\%$	$10\% < rMAE \leq 15\%$	$90 < R^2 < 95$
Excellent	$rMBE \leq \pm 5 \%$	$rRMSE \leq 10\%$	$rMAE \leq 10\%$	$R^2 > 95$

4.1. CAMS

From Figure 3, CAMS underestimated GHI in Upington (−3%), Prieska (−0.1%), De Aar (−2%), and Cape Point (−0.1%); and overestimated GHI at the remaining nine stations, rMBE ranged from 2%–6%. From Figure 4, hourly rRMSE was less than and slightly above 20% for all stations. When considering the correlation R^2 results in Table 6, CAMS had $R^2 > 0.96$ in all stations, suggesting that measured and estimated data correlate well. Good correlation was also demonstrated by scatterplots in Figures S1–S13, given in the Supplementary Materials.

From Figures 3–5 and Table 6, the hourly metrics for CAMS varied as given below:

- $-3\% \leq \text{hourly rMBE} \leq 6\%$;
- $10\% \leq \text{hourly rRMSE} \leq 21\%$;
- $6\% \leq \text{hourly rMAE} \leq 13\%$; and
- $0.962 \leq \text{hourly } R^2 \leq 0.995$.

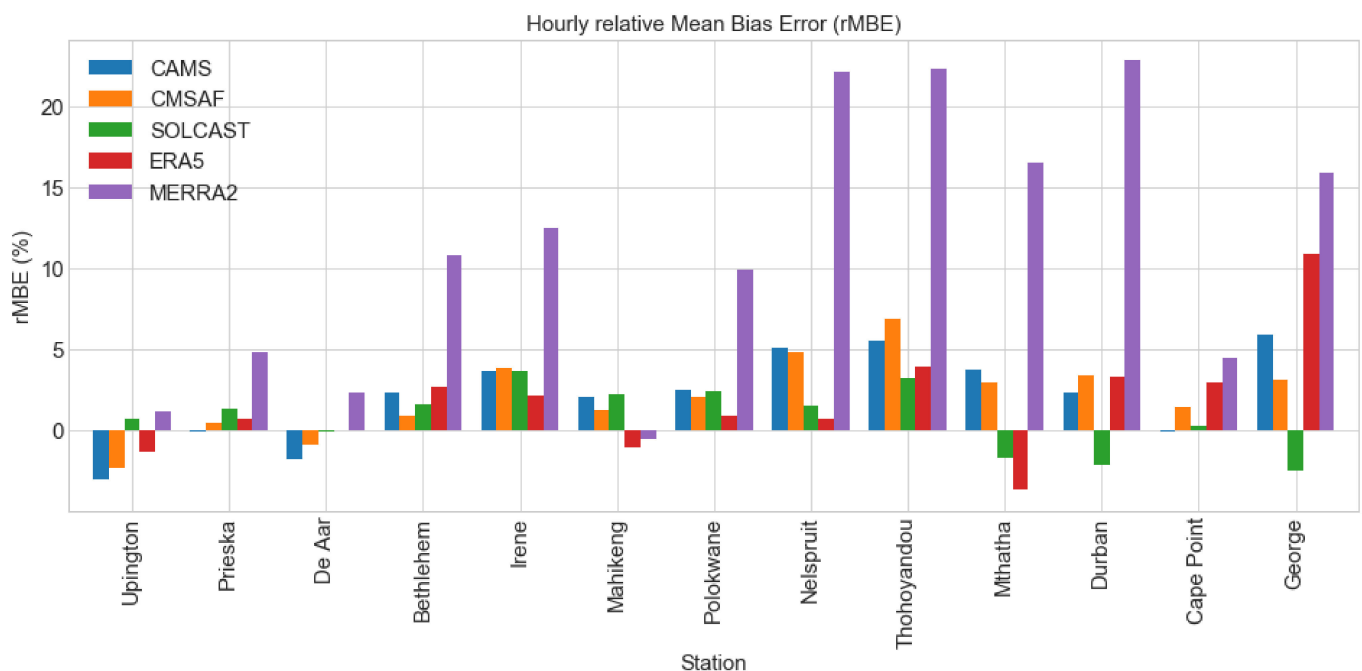


Figure 3. Hourly relative mean bias error of gridded datasets against measured in situ GHI.

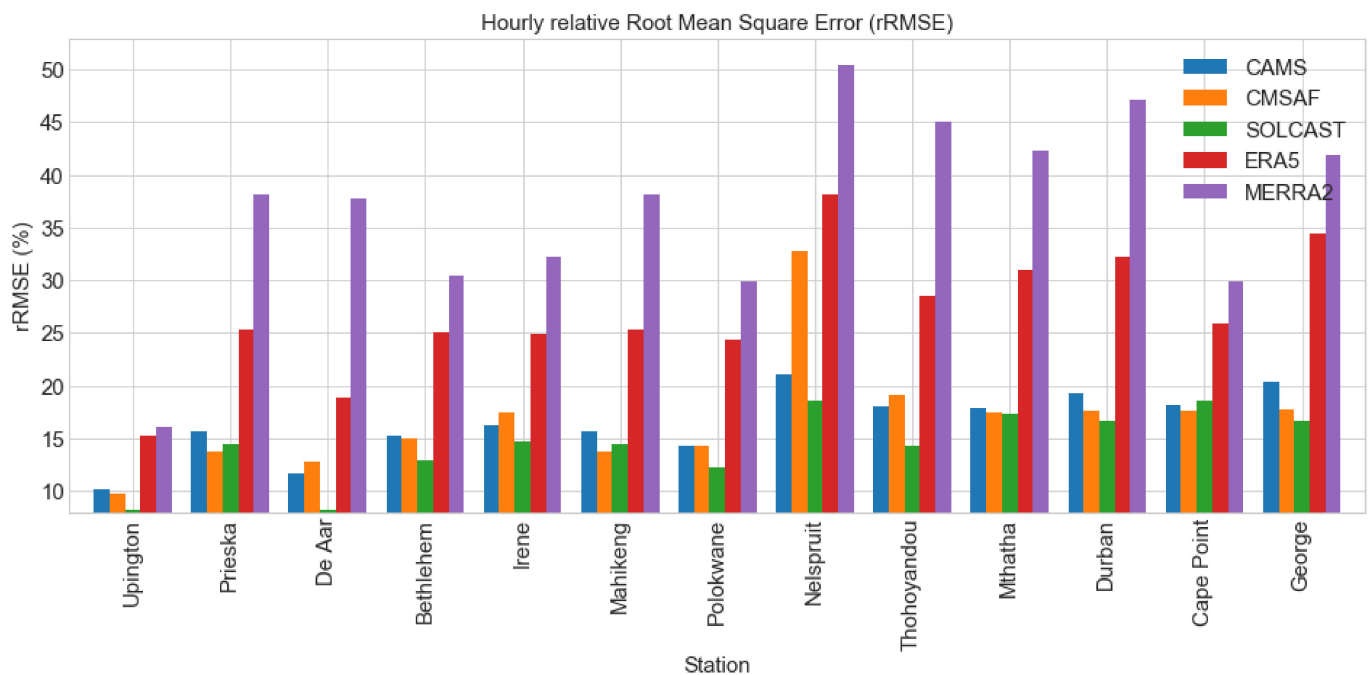


Figure 4. Hourly relative root mean square error of gridded datasets against measured in situ GHI.

Table 6. Hourly mean measured GHI (W/m^2) and correlation (R^2) of gridded datasets against measured in situ GHI. Green color represents the best ($R^2 > 95$), blue color represents the intermediate ($90 < R^2 < 95$), and orange color represents the poor correlation ($R^2 < 90$).

Station	Mean GHI (W/m^2)	CAMS	CMSAF	SOLCAST	ERA5	MERRA2
Upington	522.86	0.995	0.995	0.996	0.987	0.985
Prieska	490.95	0.986	0.983	0.991	0.962	0.955
DeAar	497.67	0.984	0.981	0.992	0.958	0.833
Bethlehem	459.50	0.976	0.976	0.983	0.932	0.914
Irene	458.15	0.973	0.969	0.979	0.932	0.907
Mahikeng	492.13	0.972	0.978	0.977	0.923	0.826
Polokwane	466.63	0.978	0.977	0.985	0.932	0.910
Nelspruit	404.24	0.962	0.908	0.970	0.868	0.823
Thohoyandou	408.46	0.972	0.970	0.983	0.922	0.856
Mthatha	382.77	0.975	0.976	0.975	0.915	0.873
Durban	364.64	0.972	0.977	0.979	0.919	0.871
Cape Point	415.41	0.970	0.972	0.969	0.938	0.920
George	387.46	0.968	0.974	0.977	0.909	0.875

4.2. CMSAF

From Figure 3, CMSAF underestimated GHI in Upington (−2%) and De Aar (−1%) and overestimated GHI at the remaining 11 stations; rMBE ranged from 0–7%. In Figure 4, when considering rRMSE, CMSAF showed a very good performance at Upington (10%), the worst performance at Nelspruit (33%), and the rest of the stations had a rRMSE between 10% and 20%.

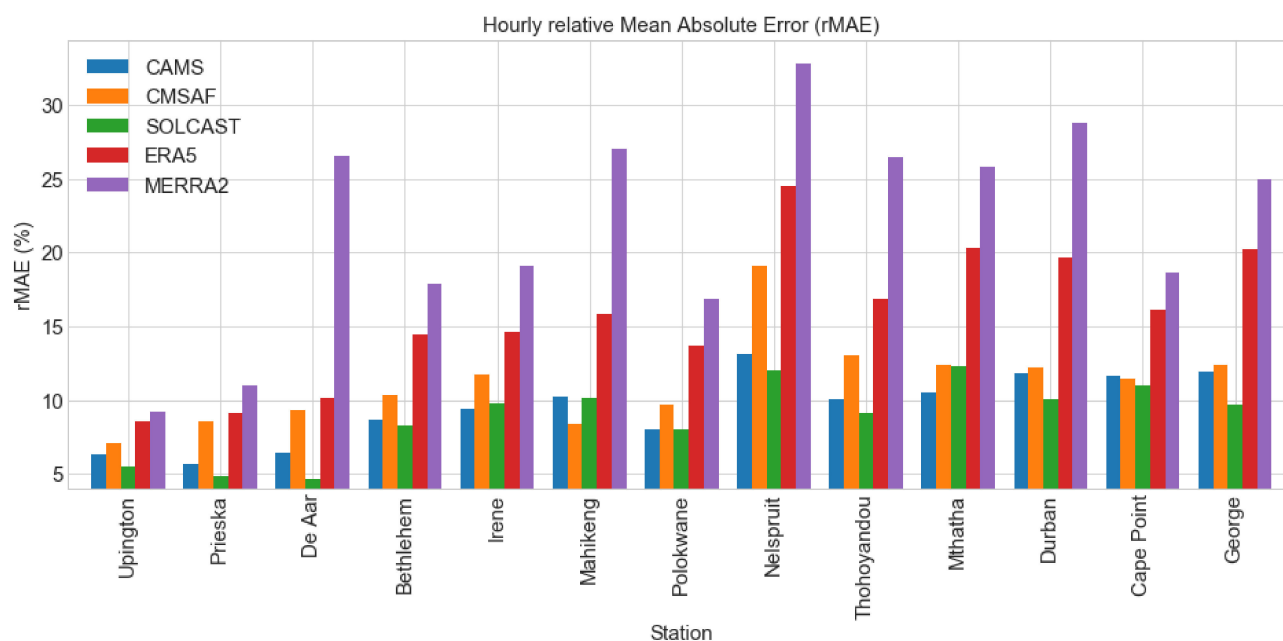


Figure 5. Hourly relative mean absolute error of gridded datasets against measured in situ GHI.

From Table 6, when considering the hourly correlation R^2 results, CMSAF had $R^2 > 0.96$ at all stations except in Nelspruit (0.908), which shows a very good correlation between the in situ observed and estimated data. This is also demonstrated by scatterplots in Figures S1–S13, given in the Supplementary Materials.

From Figures 3–5 and Table 6, the hourly metrics for CMSAF varied as given below:

- $-2\% \leq \text{hourly rMBE} \leq 7\%$;
- $10\% \leq \text{hourly rRMSE} \leq 33\%$;
- $7\% \leq \text{hourly rMAE} \leq 19\%$; and
- $0.908 \leq \text{hourly } R^2 \leq 0.995$.

4.3. SOLCAST

From Table 7, which gives the most feasible hourly dataset per station by combining all the hourly metrics, SOLCAST was most prominent in 10 out of 13 stations for the hourly aggregates. From Figure 3, the SOLCAST dataset slightly underestimated hourly GHI at De Aar (-0.1%), Mthatha (-2%), George (-3%), and Durban (-2%) and it overestimated hourly GHI at the rest of the stations; rMBE ranged from 1% to 4%. From Table 6, SOLCAST had the correlation $R^2 > 0.96$ in all 13 stations, showing a very good agreement between the measured and estimated data. This was also demonstrated by the scatterplots in Figures S1–S13, given in the Supplementary Materials.

SOLCAST hourly rRMSE was less than 15% at eight stations and more than 15% in five stations George (17%), Cape Point (19%), Durban (17%), Mthatha (17%), and Nelspruit (19%).

From Figures 3–5 and Table 6, the hourly metrics for SOLCAST varied as given below:

- $-3\% \leq \text{hourly rMBE} \leq 4\%$;
- $8\% \leq \text{hourly rRMSE} \leq 19\%$;
- $5\% \leq \text{hourly rMAE} \leq 12\%$;
- $0.969 \leq \text{hourly } R^2 \leq 0.996$;

Table 7. Best performing gridded dataset per hourly metric, most feasible model, and level of performance (rating). The colors were used to show the best performing gridded dataset based on the hourly statistical metric per station and the most feasible dataset per station based level of rating out of 4 (the number of statistical metrics used). Green represents SOLCAST, yellow CAMS, blue CMSAF, orange ERA5, and red MERRA2.

Hourly	Minimum rMBE	Minimum rRMSE	Minimum rMAE	Maximum R^2	Most Feasible	Rating
Upington	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Prieska	CAMS	CMSAF	SOLCAST	SOLCAST	SOLCAST	2/4
De Aar	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Bethlehem	CMSAF	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Irene	ERA5	SOLCAST	CAMS	SOLCAST	SOLCAST	2/4
Mahikeng	MERRA2	CMSAF	CMSAF	CMSAF	CMSAF	3/4
Polokwane	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Nelspruit	ERA5	SOLCAST	SOLCAST	SOLCAST	SOLCAST	3/4
Thohoyandou	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Mthatha	SOLCAST	SOLCAST	CAMS	CMSAF	SOLCAST	2/4
Durban	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4
Cape Point	CAMS	CMSAF	SOLCAST	CMSAF	CMSAF	2/4
George	SOLCAST	SOLCAST	SOLCAST	SOLCAST	SOLCAST	4/4

4.4. ERA5

From Figure 4, two stations Upington (15%) and De Aar (19%) had a rRMSE less than 20%. Seven stations had rRMSE between 20% and 30%, and four stations, namely Mthatha (31%), Durban (32%), George (34%), and Nelspruit (38%) had rRMSEs greater than 30%. When considering hourly R^2 results, three stations namely Upington (0.987), Prieska (0.962), and De Aar (0.958) had $R^2 > 0.95$, and nine stations had $0.9 < R^2 < 0.95$. Nelspruit (0.868) was the only station with hourly $R^2 < 0.9$. The poor correlation of ERA5 reanalysis hourly data was also demonstrated by scatterplots in Figures S1–S13, given in the Supplementary Materials, with data points not elongated along the 1:1 line. From Figure 3, ERA5 reanalysis data underestimated hourly GHI in Upington (−1%), Mahikeng (−1%), and Mthatha (−4%) and overestimated in 10 stations with rMBE ranging from 0% to 11%.

From Figures 3–5 and Table 6, the hourly metrics for ERA5 varied as given below:

- $-4\% \leq \text{hourly rMBE} \leq 11$
- $15\% \leq \text{hourly rRMSE} \leq 38\%$;
- $9\% \leq \text{hourly rMAE} \leq 25\%$;
- $0.868 \leq \text{hourly } R^2 \leq 0.987$;

4.5. MERRA2

From Figure 3, the MERRA2 reanalysis hourly dataset overestimated GHI in 12 stations, rMBE ranged from 1% to 23%, and slightly underestimated GHI in Mahikeng (−0.54%). From the rRMSEs given in Figure 4, Upington (16%), Cape Point (29.86%), and Polokwane (29.89%) were the only stations with an hourly rRMSE less than 30%. Ten other stations had hourly rRMSE ranging from 30% to 50%. When considering hourly R^2 results, two stations, namely Upington (0.985) and Prieska (0.955) had $R^2 > 0.95$, three stations had $0.9 < R^2 < 0.95$. Seven stations had $R^2 < 0.9$. The poor correlation of hourly MERRA2 reanalysis data was also demonstrated by scatterplots in Figures S1–S13, given in the

Supplementary Materials, where the data points were not elongated along the 1:1 line and were mostly above the 1:1 line, indicating overestimation.

From Figures 3–5 and Table 6, the hourly metrics for MERRA2 varied as given below:

- $-1\% \leq \text{hourly rMBE} \leq 23\%$;
- $16\% \leq \text{hourly rRMSE} \leq 50\%$;
- $9\% \leq \text{hourly rMAE} \leq 32\%$; and
- $0.823 \leq \text{hourly } R^2 \leq 0.985$.

Figures 6–9, which shows the aggregated hourly averages, demonstrates that all the gridded datasets could capture the temporal variability of GHI in different sites nonetheless, with varying accuracy. The MERRA2 reanalysis dataset overestimated GHI and it did not perform well as its line diverged from the reference measured dataset as well as the other datasets, in almost all the stations. ERA5 reanalysis also overestimated, and it diverged from observation and satellite datasets lines in most stations. SOLCAST, CMSAF, and CAMS satellite-based datasets lines were adjacent to the observation in most stations, which showed good performance.

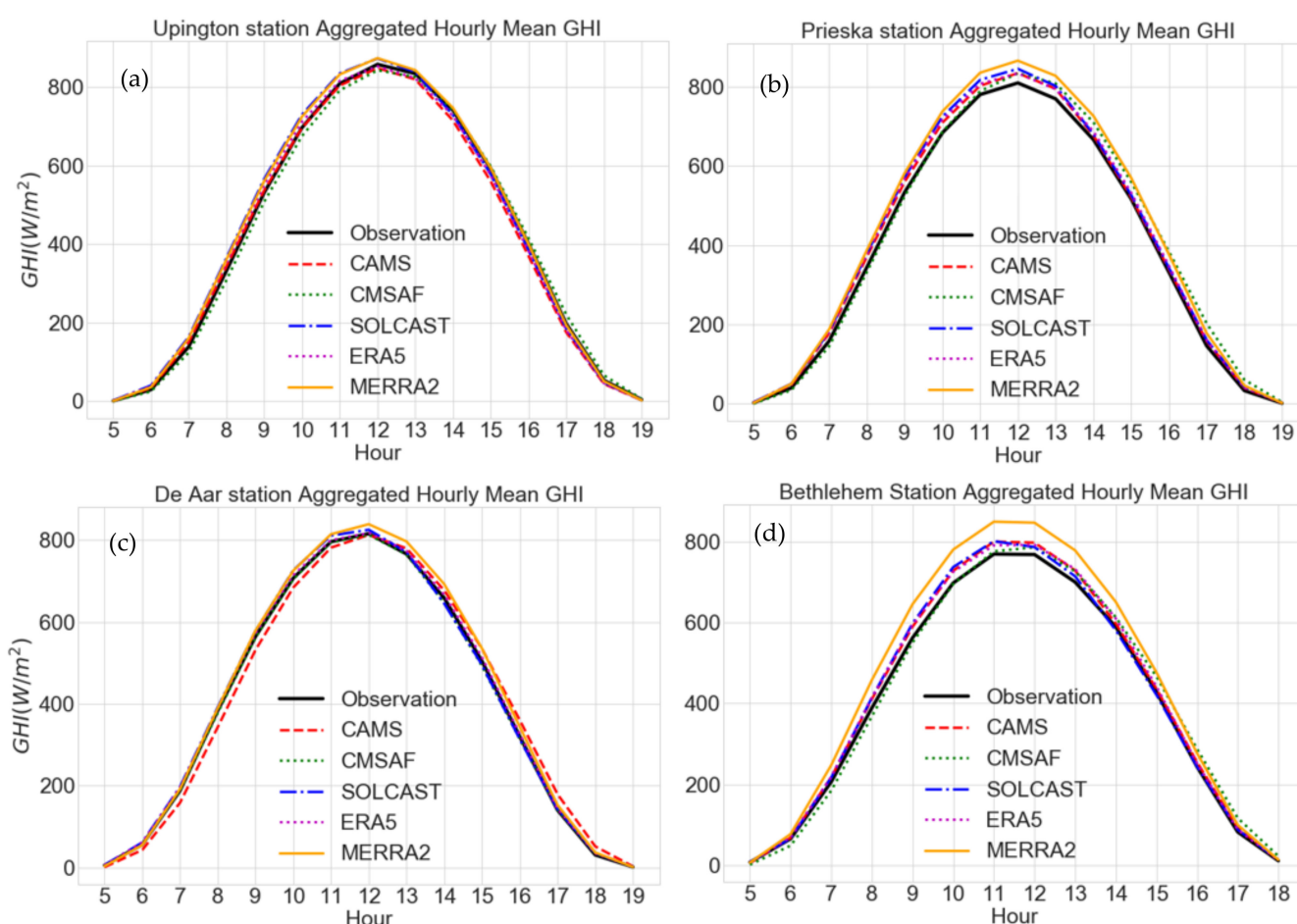


Figure 6. Aggregated measured and estimated hourly GHI values in De Aar (a), Bethlehem (b), Prieska (c), and Uppington (d). The aggregated GHI Observation period for each station is given in Table 2.

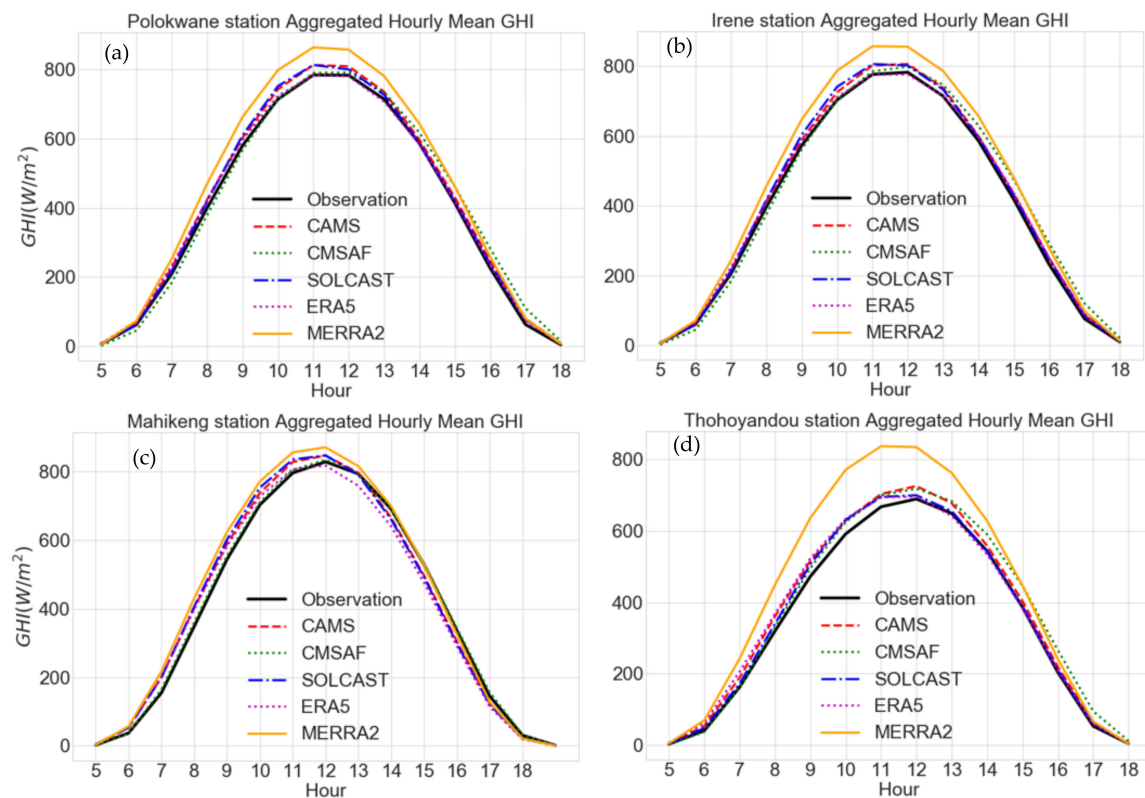


Figure 7. Aggregated measured and estimated hourly GHI values in Polokwane (a), Irene (b), Mahikeng (c), and Thohoyandou (d). The aggregated GHI Observation period for each station is given in Table 2.

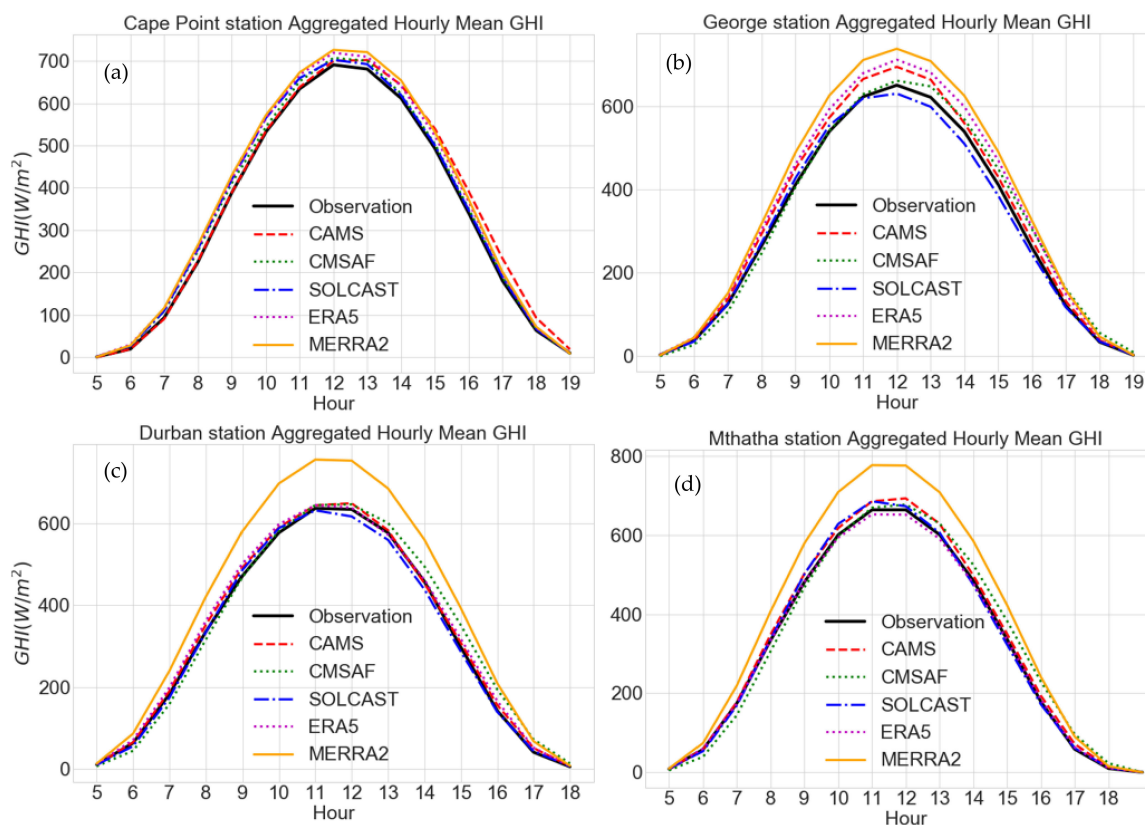


Figure 8. Aggregated measured and estimated hourly GHI values in values in Cape Point (a), George (b), Durban (c), and Mthatha (d). The aggregated GHI Observation period for each station is given in Table 2.

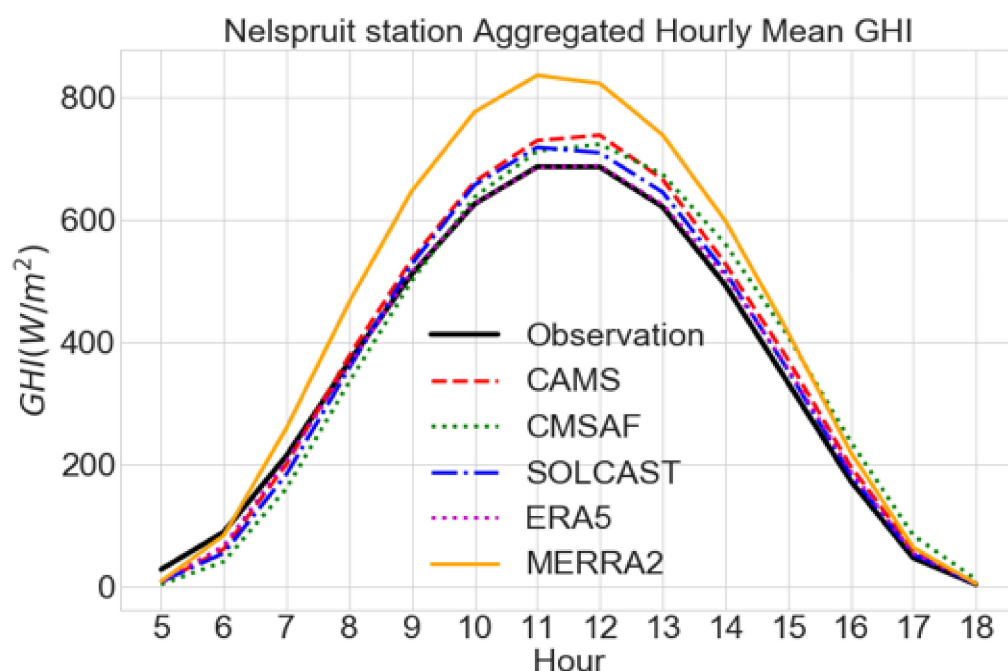


Figure 9. Aggregated measured and estimated hourly GHI values in values in Nelspruit. The aggregated GHI Observation period for each station is given in Table 2.

5. Discussion

5.1. CAMS

The above statistical metrics suggest that the CAMS dataset has good performance in South Africa. The findings are similar to the studies by Marchand et al. (2018) [9] in Morocco; Ameen et al. (2018) [20] in Iraq; Yang and Bright [17] for 57 BSRN stations; Thomas et al. [30] in Brazil; and Trolliet et al. [31] in the tropical Atlantic Ocean, which found that CAMS could accurately estimate the hourly GHI and can therefore be used with quantitative confidence as a reliable alternative source of estimated GHI data. The relatively high rRMSE observed for Nelspruit, Thohoyandou, Mthatha, Durban, Cape Point, and George may be due to the characteristic high annual humidity (greater than 60% in Tables 2 and 3). The same tendency was also observed by Thomas et al. [30]. George and Nelspruit stations also had a high diffuse fraction ($DF > 0.32$). This implies that the performance of CAMS is affected by high aerosols, high humidity, and hence, many days with diffuse skies. When considering areas where there is infrequent cloud occurrence, there was no significance difference in bias, meaning that McClear can accurately estimate clear sky conditions in the study area. The previous study by Mabasa et al. [32] showed that the McClear clear sky model had a good performance in South Africa. For CAMS, cloud properties were derived from Meteosat satellites with 15-min temporal resolution.

For CAMS, the main inputs to Heliosat-4 are aerosol properties, total column water vapor, and ozone content, as provided by the CAMS global services every three hours. The lower temporal resolution in aerosol, total column vapor, and ozone content may account for the observed biases between ground truth and CAMS estimated GHI. The overall good performance of CAMS datasets might be attributed to having a high spatial resolution 3–5 km in Southern Africa.

5.2. CMSAF

The overall results show that there is relatively good performance by the CMSAF satellite-based dataset, which suggests that CMSAF is a viable tool to estimate GHI for sites such as the 13 stations in this study. CMSAF satellite-based dataset showed a relatively poor performance at Nelspruit Station. Nelspruit is a station with the highest diffuse

fraction (Tables 2 and 3). CMSAF satellite-based dataset uses aerosol climatology as input to satellite retrieval algorithms as given by Riihelä et al. [49], however, aerosol climatology might not capture the aerosol climate variability. Mueller et al. [55] showed that aerosol climatologies used in CMSAF satellite retrievals algorithms were underestimated when compared to real aerosol measurements. The poorer metrics for Nelspruit might be due to the use of aerosol climatology information in the CMSAF satellite retrieval algorithm. No significant bias has been observed for the stations with high humidity and more frequent cloud occurrence compared to areas with low humidity and less frequent cloud occurrence. This means that the CMSAF cloud and water vapor parameterization scheme is effective in South Africa.

The CMSAF satellite-based dataset outperformed CAMS, ERA5, and MERRA2 at all 13 stations under this study (Table 7). CMSAF was outperformed by SOLCAST in 11 stations, this might be due to the use of hourly average from only two intervals (half hour and hour) instead of four intervals (15 min, 30 min, 45 min, and hour). The good performance of CMSAF datasets might also be attributed to having a high spatial resolution 5 km.

5.3. SOLCAST

These results show an overall very good performance of SOLCAST satellite-based dataset in all 13 stations in this study. The results agreed with the study by Yang and Bright [17] and Bright [26], who found that the SOLCAST satellite-based dataset performed well. Overall SOLCAST was outperformed by CMSAF (from Table 7) at Mahikeng and Cape Point Stations; this was similar to the findings by Yang and Bright [17], where SOLCAST did not outperform some freely available products at all sites. The stations where rRMSE was greater than 15% when referencing from Tables 2 and 3. They all had a low number of clear sky days (less than 5%), high humidity (greater than 60%), and high diffuse fraction (greater than 0.33), except Nelspruit Station with (more than 10%) number of clear sky days. This implies that frequent cloud occurrence, higher humidity, and higher diffuse fraction slightly affected the performance of the SOLCAST satellite-based dataset.

The excellent performance of the SOLCAST dataset might be due to the use of the REST2v5 [45] clear sky model to calculate the clear sky index when converting the satellite image to GHI. Sun X et al. [56] found that the REST2v5 clear sky model had an excellent worldwide performance. The use of very high spatial resolution satellite images 1–2 km enabled almost all features (e.g., terrain difference) in an area of interest or a grid to be properly identified and properly interpolated.

5.4. ERA5

Overall, the ERA5 reanalysis dataset showed a poor performance in estimating GHI in South Africa. The results were similar to Yang and Bright [17] and Trolliet et al. [31], who found that ERA5 datasets had poor performance, overestimated GHI for most sites, and were outperformed by the satellite-based dataset; rMBE was 72% (Table 1). ERA5 reanalysis data showed a very poor performance in areas with frequent cloud occurrences, high humidity, and high diffuse fraction. From Supplementary Materials Figures S1–S13, ERA5 estimates tended to estimate cloud conditions while observations showing actual conditions as clear conditions was shown by irradiance values, this might be contributing to the higher biases. Basically, ERA5 cloud models struggle to differentiate non cloud and cloud conditions. ERA5 uses climatological aerosol information [37] instead of measured aerosols which captures changes in atmospheric constituents. This might also be one of the reasons for poor performance. The low spatial resolution ($0.25^\circ \times 0.25^\circ$) of the ERA5 reanalysis data might also be a contributing factor to poor performance of the ERA5 datasets in South Africa.

5.5. MERRA2

Overall, MERRA2 reanalysis showed a very poor performance and it overestimated hourly GHI in all 13 sites under study. The performance of MERRA2 in South Africa was

similar to the findings by Yang and Bright [17] for 57 BSRN stations and Trollet et al. [31] in the tropical Atlantic Ocean (i.e., poor performance, overestimating GHI, and being outperformed by satellite-based datasets; rMBE was 76%) (Table 1). MERRA2 data showed a very poor performance in areas with frequent cloud occurrences, high humidity, and high diffuse fraction. The very low spatial resolution ($0.625^\circ \times 0.5^\circ$) of MERRA2 reanalysis data might also be a contributing factor to poor performance of MERRA2, as it was the overall worst performing dataset in the study.

6. Conclusions

The study validated hourly global horizontal irradiance (GHI) from three satellite-based GHI datasets (namely SOLCAST, CAMS, and CMSAF SARA) and two reanalysis based GHI datasets (namely ERA5 and MERRA2) against quality-controlled hourly in situ GHI recorded at 13 radiometric stations in South Africa. The study demonstrated that GHI from the satellite-based datasets had better performance than reanalysis-based datasets in South Africa. The overall statistical metrics used to gauge the performance of the datasets varied, as tabulated below (Table 8).

Table 8. Summary of the overall validation results. Colors are used to show the overall performance ranking of the five datasets with green (1/5), yellow (2/5), blue (3/5), orange (4/5), and red (5/5).

Dataset	rMBE	rRMSE	rMAE	R ²
SOLCAST	−3% to 4%	8% to 19%	5% to 12%	0.969 to 0.996
CAMS	−3% to 6%	10% to 21%	6% to 13%	0.962 to 0.995
CMSAF	−2% to 7%	10% to 33%	7% to 19%	0.908 to 0.995
ERA5	−4% to 11%	15% to 38%	9% to 25%	0.868 to 0.987
MERRA2	−1% to 23%	16% to 50%	9% to 32%	0.823 to 0.985

SOLCAST was the best performing overall, while MERRA2 was the overall worst performing dataset. Freely available satellite-based datasets (CAMS and CMSAF) are recommended for use with quantitative confidence in diverse solar energy applications that require GHI data. Reanalysis based GHI datasets (ERA5 and MERRA2) are not good enough to be used in South Africa. Low spatial resolution, weak cloud parameterization schemes, and the use of climatological inputs instead of real in situ measurement in reanalysis GHI deriving algorithms might be some of the reasons behind the poor performance of reanalysis based GHI estimates in the study.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/geomatics1040025/s1>. Figures S1–S13: Hourly measured and estimated GHI correlation graphs.

Author Contributions: Conceptualization, B.M.; Methodology, B.M. and M.D.L.; Software, B.M. and M.D.L.; Validation, M.D.L.; Formal analysis, B.M. and M.D.L.; Investigation, B.M.; Data curation, B.M.; Writing—original draft preparation, B.M.; Supervision, M.D.L. and S.J.M.; Writing—review and editing, B.M., M.D.L. and S.J.M. All authors have read and agreed to the published version of the manuscript.

Funding: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The datasets used in the study can be accessed as follows: (1) Reference in situ global horizontal irradiance (GHI) data: South African Weather Services (SAWS) measurements are performed every one minute from 13 stations. Data are available on request from (<https://www.weathersa.co.za/home/network>, last access: 1 November 2021). (2) SOLCAST is a commercial data supplier with support for researchers available on request from at (<https://solcast.com/>, last

access: 1 November 2021). (3) CMSAF SARA time series data were extracted from the gridded datasets available at an open source (https://wui.cmsaf.eu/safira/action/viewDoiDetails?acronym=SARAH_V002, last access: 1 November 2021). (4) CAMS radiation data were downloaded from the SoDa service website, an open source (<http://www.soda-pro.com>, last access: 1 November 2021). (5) MERRA-2 times-series were extracted from the gridded datasets available at (<https://goldsmr4.gesdisc.eosdis.nasa.gov/data/MERRA2/>, last access: 1 November 2021). (6) ERA5 times-series data were extracted from the gridded datasets available at (<http://apps.ecmwf.int/data-catalogues/era5/?class=ea&stream=enda&expver=1>, last access: 1 November 2021).

Acknowledgments: The authors thank the South African Weather Services (SAWS) for providing the Global Horizontal Irradiance (GHI) data used as a reference for satellite and reanalysis data validation and for providing the humidity data. The authors thank the Copernicus Atmosphere Monitoring Service (CAMS) Solar Radiation Data (SoDa) for freely providing the CAMS hourly GHI data on their website; European Center for Medium-Range Weather Forecasts (ECMWF) for providing freely available ERA5 hourly cloud data and ERA5 hourly radiation on their website; the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) for providing freely available surface Satellite Application Facility on Climate Monitoring (CMSAF SARA time series) data on their website, and the National Aeronautics and Space Administration (NASA) Global Modeling and Assimilation Office (GMAO) for freely providing hourly Modern-Era Retrospective Analysis for Research and Applications (MERRA2) GHI data on their website. The authors thank SOLCAST for providing the SOLCAST satellite derived GHI data used in the study.

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

References

- Liou, K.N. *An Introduction to Atmospheric Radiation*; Elsevier: Amsterdam, The Netherlands, 2002.
- Dal Pai, A.; Escobedo, J.F.; Martins, D.; Teramoto, É.T. Analysis of hourly global, direct and diffuse solar radiations attenuation as a function of optical air mass. *Energy Procedia* **2014**, *57*, 1060–1069. [[CrossRef](#)]
- Schwandt, M.; Chhatbar, K.; Meyer, R.; Mitra, I.; Vashistha, R.; Giridhar, G.; Gomathinayagam, S.; Kumar, A. Quality check procedures and statistics for the Indian SRRA solar radiation measurement network. *Energy Procedia* **2014**, *57*, 1227–1236. [[CrossRef](#)]
- Zawilska, E.; Brooks, M. An assessment of the solar resource for Durban, South Africa. *Renew. Energy* **2011**, *36*, 3433–3438. [[CrossRef](#)]
- Urraca, R.; Gracia-Amillo, A.M.; Huld, T.; Martinez-De-Pison, F.J.; Trentmann, J.; Lindfors, A.V.; Riihelä, A.; Sanz-Garcia, A. Quality control of global solar radiation data with satellite-based products. *Sol. Energy* **2017**, *158*, 49–62. [[CrossRef](#)]
- Zawilska, E.; Brooks, M.J.; Meyer, A.J. A review of solar resource assessment initiatives in South Africa: The case for a national network. In Proceedings of the World Renewable Energy Forum, Denver, CO, USA, 13–17 May 2012.
- Jiang, H.; Yang, Y.; Bai, Y.; Wang, H. Evaluation of the total, direct, and diffuse solar radiations from the ERA5 reanalysis data in China. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 47–51. [[CrossRef](#)]
- Sianturi, Y.; Marjuki; Sartika, K. Evaluation of ERA5 and MERRA2 reanalyses to estimate solar irradiance using ground observations over Indonesia region. *AIP Conf. Proc.* **2020**, *2223*, 020002. [[CrossRef](#)]
- Marchand, M.; Ghennioui, A.; Wey, E.; Wald, L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Adv. Sci. Res.* **2018**, *15*, 21–29. [[CrossRef](#)]
- Martín-Pomares, L.; Romeo, M.G.; Polo, J.; Frías-Paredes, L.; Fernández-Peruchena, C.M. Sampling design optimization of ground radiometric stations. In *Solar Resources Mapping*; Springer International Publishing: Cham, Switzerland, 2019; pp. 253–281.
- Wilbert, S.; Stoffel, T.; Myers, D.; Wilcox, S.; Habte, A.; Vignola, F.; Wood, J.; Pomares, L.M. Measuring solar radiation and relevant atmospheric parameters. In *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*; National Renewable Energy Laboratory: Golden, CO, USA, 2017. Available online: <https://hal-mines-paristech.archives-ouvertes.fr/hal-01184753> (accessed on 12 March 2020).
- Moradi, I. Quality control of global solar radiation using sunshine duration hours. *Energy* **2009**, *34*, 1–6. [[CrossRef](#)]
- Angstrom, A. Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation. *Q. J. R. Meteorol. Soc.* **1924**, *50*, 121–126. [[CrossRef](#)]
- Prescott, J.A. Evaporation from a water surface in relation to solar radiation. *Trans. R. Soc. S. Aust.* **1940**, *46*, 114–118.
- Mabasa, B.; Lysko, M.D.; Tazvinga, H.; Mulaudzi, S.T.; Zwane, N.; Moloi, S.J. The ångström–prescott regression coefficients for six climatic zones in South Africa. *Energies* **2020**, *13*, 5418. [[CrossRef](#)]
- Žák, M.; Mikšovský, J.; Pišoft, P. CMSAF radiation data: New possibilities for climatological applications in the Czech Republic. *Remote Sens.* **2015**, *7*, 14445–14457. [[CrossRef](#)]
- Yang, D.; Bright, J.M. Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years. *Sol. Energy* **2020**, *210*, 3–19. [[CrossRef](#)]

18. Esterhuyse, D.J. Establishment of the South African Baseline Surface Radiation Network Station at De Aar. Ph.D. Thesis, University of Pretoria, Pretoria, South Africa, 2006. Available online: <https://repository.up.ac.za/handle/2263/23761> (accessed on 10 May 2020).
19. Yang, D.; Perez, R. Can we gauge forecasts using satellite-derived solar irradiance? *J. Renew. Sustain. Energy* **2019**, *11*, 023704. [[CrossRef](#)]
20. Ameen, B.; Balzter, H.; Jarvis, C.; Wey, E.; Thomas, C.; Marchand, M. Validation of hourly global horizontal irradiance for two satellite-derived datasets in Northeast Iraq. *Remote Sens.* **2018**, *10*, 1651. [[CrossRef](#)]
21. Sancho, J.M. Comparison of global irradiance measurements of the official Spanish radiometric network for 2006 with satellite estimated data. *J. Mediterr. Meteorol. Climatol. Tethys* **2011**, *8*, 43–52. [[CrossRef](#)]
22. Slater, A. Surface solar radiation in North America: A comparison of observations, reanalyses, satellite, and derived products. *J. Hydrometeorol.* **2015**, *17*, 401–420. [[CrossRef](#)]
23. Driemel, A.; Augustine, J.; Behrens, K.; Colle, S.; Cox, C.; Cuevas-Agulló, E.; Denn, F.M.; Duprat, T.; Fukuda, M.; Grobe, H.; et al. Baseline surface radiation network (BSRN): Structure and data description (1992–2017). *Earth Syst. Sci. Data* **2018**, *10*, 1491–1501. [[CrossRef](#)]
24. Journée, M.; Bertrand, C. Geostatistical merging of ground-based and satellite-derived data of surface solar radiation. *Adv. Sci. Res.* **2011**, *6*, 1–5. [[CrossRef](#)]
25. Dirksen, M.; Meirink, J.F.; Sluiter, R. Quality assessment of high-resolution climate records of satellite derived solar irradiance. *Energy Procedia* **2017**, *125*, 221–229. [[CrossRef](#)]
26. Bright, J.M. Solcast: Validation of a satellite-derived solar irradiance dataset. *Sol. Energy* **2019**, *189*, 435–449. [[CrossRef](#)]
27. Marchand, M.; Lefèvre, M.; Saboret, L.; Wey, E.; Wald, L. Verifying the spatial consistency of the CAMS radiation service and HelioClim-3 satellite-derived databases of solar radiation using a dense network of measuring stations: The case of The Netherlands. *Adv. Sci. Res.* **2019**, *16*, 103–111. [[CrossRef](#)]
28. Lefèvre, M.; Oumbe, A.; Blanc, P.; Espinar, B.; Gschwind, B.; Qu, Z.; Wald, L.; Schroedter-Homscheidt, M.; Hoyer-Klick, C.; Arola, A.; et al. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos. Meas. Tech.* **2013**, *6*, 2403–2418. [[CrossRef](#)]
29. Qu, Z.; Oumbe, A.; Blanc, P.; Espinar, B.; Gesell, G.; Gschwind, B.; Klüser, L.; Lefèvre, M.; Saboret, L.; Schroedter-Homscheidt, M.; et al. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Meteorol. Z.* **2017**, *26*, 33–57. [[CrossRef](#)]
30. Thomas, C.; Wey, E.; Blanc, P.; Wald, L. Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil. *Adv. Sci. Res.* **2016**, *13*, 81–86. [[CrossRef](#)]
31. Trolliet, M.; Walawender, J.P.; Bourlès, B.; Boilley, A.; Trentmann, J.; Blanc, P.; Lefèvre, M.; Wald, L. Downwelling surface solar irradiance in the tropical Atlantic Ocean: A comparison of re-analyses and satellite-derived data sets to PIRATA measurements. *Ocean Sci.* **2018**, *14*, 1021–1056. [[CrossRef](#)]
32. PIRATA Network. Available online: <https://www.pmel.noaa.gov/gtmba/pmel-theme/atlantic-ocean-pirata> (accessed on 15 May 2021).
33. Conradie, D. Köppen-Geiger Climate Classification. Available online: https://stepsa.org/climate_koppen_geiger.html#Development (accessed on 8 May 2021).
34. Mabasa, B.; Lysko, M.; Tazvinga, H.; Zwane, N.; Moloi, S. The performance assessment of six global horizontal irradiance clear sky models in six climatological regions in South Africa. *Energies* **2021**, *14*, 2583. [[CrossRef](#)]
35. Mabasa, M.B.; Botai, J.; Ntsangwane, M.L. Update on the re-establishment of the south african weather services (SAWS) radiometric network in all six climatological regions and the quality of the data. In Proceedings of the South African Solar Energy Conference (SASEC), Blue Waters Hotel, KwaZulu-Natal, South Africa, 25–27 June 2018. Available online: https://www.sasec.org.za/full_papers/68.pdf (accessed on 20 January 2020).
36. Peng, X.; She, J.; Zhang, S.; Tan, J.; Li, Y. Evaluation of multi-reanalysis solar radiation products using global surface observations. *Atmosphere* **2019**, *10*, 42. [[CrossRef](#)]
37. Hersbach, H.; Bell, B.; Berrisford, P.; Hirahara, S.; Horanyi, A.; Muñoz-Sabater, J.; Nicolas, J.; Peubey, C.; Radu, R.; Schepers, D.; et al. The ERA5 global reanalysis. *Q. J. R. Meteorol. Soc.* **2020**, *146*, 1999–2049. [[CrossRef](#)]
38. Gelaro, R.; McCarty, W.; Suárez, M.J.; Todling, R.; Molod, A.; Takacs, L.; Randles, C.A.; Darmenov, A.; Bosilovich, M.G.; Reichle, R.; et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *J. Clim.* **2017**, *30*, 5419–5454. [[CrossRef](#)]
39. Pfeifroth, U.; Kothe, S.; Trentmann, J.; Hollmann, R.; Fuchs, P.; Kaiser, J.; Werscheck, M. *Surface Radiation Data Set—Heliosat (SARAH)*, 2nd ed.; Satellite Application Facility on Climate Monitoring: Offenbach, Germany, 2019. [[CrossRef](#)]
40. Pfeifroth, U.; Kothe, S.; Trentmann, J. EUMETSAT Satellite Application Facility on Climate Monitoring, Validation Report. Meteosat Solar Surface Sarah 2. Available online: https://www.cmsaf.eu/SharedDocs/Literatur/document/2016/saf_cm_dwd_val_meteosat_hel_2_1_pdf.pdf?__blob=publicationFile (accessed on 17 January 2020).
41. Copernicus Portal. Available online: <https://atmosphere.copernicus.eu/data> (accessed on 16 March 2020).
42. Solar Radiation Data (SoDa) Service. Available online: <http://solar.atmosphere.copernicus.eu/cams-radiation-service> (accessed on 2 March 2020).

43. CAMS Validation. Available online: https://atmosphere.copernicus.eu/sites/default/files/custom-uploads/EQC-solar/CAMS7_2_2018SC2_D72.1.3.1-2021Q2_RAD_validation_report_SON2020_v1.pdf (accessed on 18 January 2021).
44. SOLCAST. Available online: <https://solcast.com> (accessed on 15 April 2020).
45. Gueymard, C.A. REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation—Validation with a benchmark dataset. *Sol. Energy* **2008**, *82*, 272–285. [[CrossRef](#)]
46. SOLCAST Historical Data Validation. Available online: <https://solcast.com/historical-and-tmy/validation-and-accuracy> (accessed on 15 April 2020).
47. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2. 2010. Available online: https://epic.awi.de/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 11 December 2019).
48. 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](#)]
49. Riihelä, A.; Kallio, V.; Devraj, S.; Sharma, A.; Lindfors, A.V. Validation of the SARA-E satellite-based surface solar radiation estimates over India. *Remote Sens.* **2018**, *10*, 392. [[CrossRef](#)]
50. Geiger, M.; Diabaté, L.; Ménard, L.; Wald, L. A web service for controlling the quality of measurements of global solar irradiation. *Sol. Energy* **2002**, *73*, 475–480. [[CrossRef](#)]
51. Reda, I.; Andreas, A. Solar position algorithm for solar radiation applications. *Sol. Energy* **2004**, *76*, 577–589. [[CrossRef](#)]
52. Holmgren, W.F.; Hansen, C.W.; Mikofski, M.A. Pvlb python: A python package for modeling solar energy systems. *J. Open Source Softw.* **2018**, *3*, 884. [[CrossRef](#)]
53. Gueymard, C.A. A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects. *Renew. Sustain. Energy Rev.* **2014**, *39*, 1024–1034. [[CrossRef](#)]
54. Fernández-Peruchena, C.M.; Gastón, M. A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series. *Renew. Energy* **2016**, *86*, 375–383. [[CrossRef](#)]
55. Mueller, R.; Behrendt, T.; Hammer, A.; Kemper, A. A new algorithm for the satellite-based retrieval of solar surface irradiance in spectral bands. *Remote Sens.* **2012**, *4*, 622–647. [[CrossRef](#)]
56. Sun, X.; Bright, J.M.; Gueymard, C.A.; Acord, B.; Wang, P.; Engerer, N.A. Worldwide performance assessment of 75 global clear-sky irradiance models using principal component analysis. *Renew. Sustain. Energy Rev.* **2019**, *111*, 550–570. [[CrossRef](#)]