

Hourly global horizontal irradiance over West Africa: A case study of one-year satellite- and reanalysis-derived estimates vs. in situ measurements

Windmanagda Sawadogo^{a,*}, Jan Bliefernicht^a, Benjamin Fersch^b, Seyni Salack^c, Samuel Guug^c, Belko Diallo^c, Kehinde.O. Ogunjobi^c, Guillaume Nakoulma^d, Michael Tanu^e, Stefanie Meilinger^f, Harald Kunstmann^{a,b,**}

^a Institute of Geography, University of Augsburg, 86159, Augsburg, Germany

^b Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany

^c West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL) Competence Centre, Ouagadougou, Burkina Faso

^d Burkina Faso National Meteorological Agency (ANAM), Burkina Faso

^e Ghana Meteorological Agency (GMet), Ghana

^f International Centre for Sustainable Development (IZNE), University of Applied Sciences Bonn-Rhein-Sieg, 53757, Sankt Augustin, Germany

ARTICLE INFO

Keywords:

Solar energy
Global horizontal irradiance
Reanalysis
Satellite
West Africa

ABSTRACT

Estimates of global horizontal irradiance (GHI) from reanalysis and satellite-based data are the most important information for the design and monitoring of PV systems in Africa, but their quality is unknown due to the lack of in situ measurements. In this study, we evaluate the performance of hourly GHI from state-of-the-art reanalysis and satellite-based products (ERA5, MERRA-2, CAMS, and SARA-2) with 37 quality-controlled in situ measurements from novel meteorological networks established in Burkina Faso and Ghana under different weather conditions for the year 2020. The effects of clouds and aerosols are also considered in the analysis by using common performance measures for the main quality attributes and a new overall performance value for the joint assessment. The results show that satellite data performs better than reanalysis data under different atmospheric conditions. Nevertheless, both data sources exhibit significant bias of more than 150 W/m² in terms of RMSE under cloudy skies compared to clear skies. The new measure of overall performance clearly shows that the hourly GHI derived from CAMS and SARA-2 could serve as viable alternative data for assessing solar energy in the different climatic zones of West Africa.

1. Introduction

Global horizontal irradiance (GHI) also called surface shortwave downward radiation or solar irradiance, is defined as the amount of sunlight received from the Sun at the surface. It plays a vital role in the dynamics of the Earth's surface and drives physical processes in the atmosphere and on the land surface [1]. In addition, knowledge of the values of GHI in the solar energy sector is crucial to installing photovoltaic (PV) systems at a given location. The West Africa region receives abundant GHI throughout the year; and the daily average is estimated to be around 5–6 kWh/m² [2]. In recent years, the capacities of solar PV

technology in off-grid (rural and urban) and grid-connected systems strongly increased. For instance, between 2016 and 2018, the installed PV capacity almost tripled, and this trend is expected to continue in the coming years [3]. However, the long-term profitability of solar energy plants based on the PV technology requires an accurate GHI estimation.

Ground-based measurements from state-of-the-art pyranometers according to the WMO (World Meteorology Organization) standards are still the best data source for GHI observations [4]. However, GHI observations and related information such as sunshine duration from meteorological stations are often not accessible from African meteorological agencies due to a poor station network, national data regulations

* Corresponding author. Institute of Geography, University of Augsburg, 86159 Augsburg, Germany.

** Corresponding author. Institute of Meteorology and Climate Research (IMK-IFU), Karlsruhe Institute of Technology, Campus Alpin, Garmisch-Partenkirchen, Germany.

E-mail addresses: windmanagda.sawadogo@geo.uni-augsburg.de (W. Sawadogo), harald.kunstmann@kit.edu (H. Kunstmann).

<https://doi.org/10.1016/j.renene.2023.119066>

Received 27 June 2022; Received in revised form 16 July 2023; Accepted 18 July 2023

Available online 19 July 2023

0960-1481/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

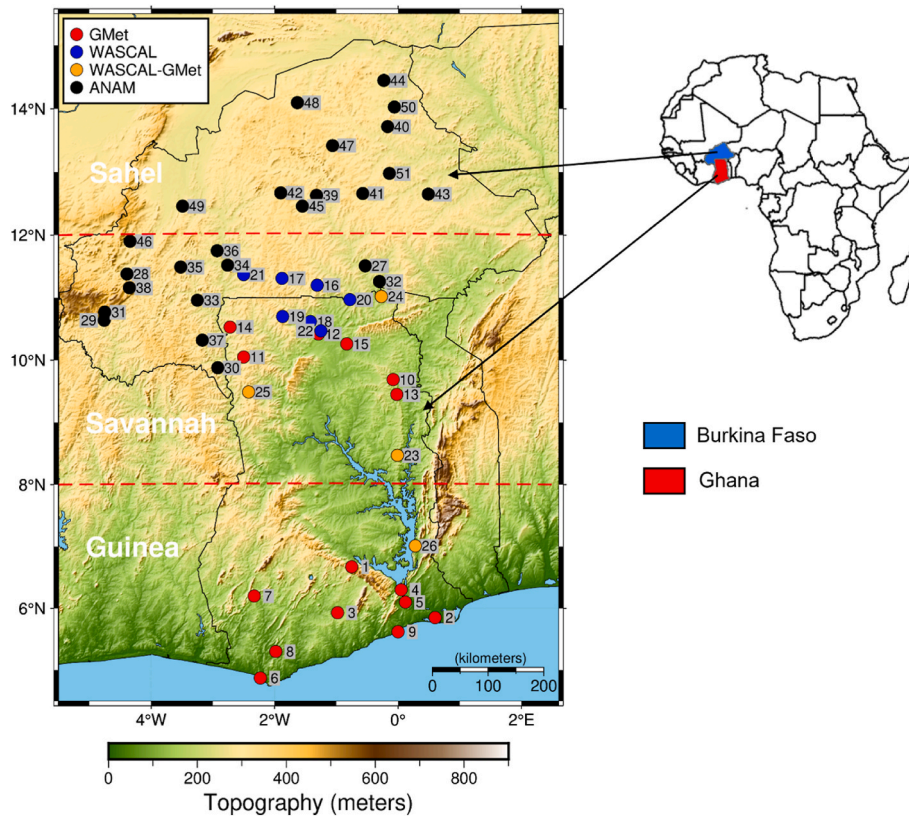


Fig. 1. Study area showing the topography of the region. The different dots are the location of the automatic weather stations (AWSs). The AWSs in red and black dots are owned by GMet and ANAM, respectively. The blue dots indicate WASCAL's AWSs and the orange locations are jointly operated by WASCAL and GMet. The red dashed lines delineate the different climatic zones. Each number corresponds to the station in [Table 1](#).

and other reasons [5–7]. In addition, station maintenance remains a challenge due to high costs, while support from local governments has declined [8]. This had a strong negative impact on data quality [9] and continuity in Africa [10]. Therefore, obtaining reliable long-term GHI observations and related information from weather stations across the region is a fundamental problem for recent and past periods. This strongly affects reliable GHI information for solar energy projects planning, operation, and quality assessment. Recently, a number of different initiatives such as WASCAL (West African Science Service Centre on Climate Change and Adapted Land Use; [6]), SASSCAL (Southern African Science Service Centre for Climate Change and Adaptive Land Management; [11]) and TAHMO (Trans-African Hydro-Meteorological Observatory; [12,13]) established a relatively dense network of automatic weather stations providing ground-based meteorological measurements at high temporal resolution for many parts of the Africa continent for the first time.

GHI satellite and reanalysis data are essential in supplementing ground-based measurements, particularly in data-scarce regions such as Africa. These datasets provide long-term GHI time series for recent periods in a relatively high spatio-temporal resolution [14,15] in uniform gridded data formats where users can retrieve the nearest grid point for their region of interest. Taking advantage of this, many investigations rely on GHI satellite-based or reanalysis data for the assessment of solar energy potential or climate impact studies [2,16–19].

However, to recommend the use of GHI satellite-based data or reanalysis data in the absence of ground-based measurements for these studies, a detailed inter-comparison and validation of these datasets for the region of interest are required. From this point of view, several studies have already carried out an inter-comparison between GHI observational, satellite, and reanalysis data. Most of them suggest that the accuracy of GHI from satellite-based and reanalysis data is lower than ground-based measurements [20]. For example, Yang and Bright.

[21] evaluated hourly GHI from 57 radiometric stations of the Baseline Surface Radiation Network (BSRN) distributed across the world with six satellite-based and two reanalysis data in a period of 27 years. They concluded that the satellite-derived hourly GHI performed better than the reanalysis data; and also, cloudy days have a higher bias than clear-sky days. Another study was carried out in the Netherlands by Marchand et al. [22]; where they used a dense 32 observational networks to assess the accuracy of hourly GHI using the Copernicus Atmosphere Monitoring Service version 3.2 (CAMS) and HelioClim-3 version 5 with correlation between 0.94 and 0.98. They showed that both satellite-based data showed a relatively good correlation with the 32 radiometric stations and satisfactorily reproduced the hourly variations of GHI. Another study conducted in Brazil showed that GHI derived from 3 satellite-based datasets could be used as an additional source for solar energy assessment in this region [23] where the relative mean bias of CAMS is about 7%. A recent study by Du et al. [24] evaluated the hourly GHI performance of the second version of the MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications Version 2) reanalysis data compared to 37 in-situ measurements over China under different sky conditions in 2018. In general, MERRA-2 overestimates the hourly GHI over China with a mean bias error of 69.35 W/m². Their results are consistent with [21] where high deviations occur under cloudy conditions.

For sub-Saharan Africa, [4] recently performed an inter-comparison of five datasets (CAMS, ERA5, SARA-2, MERRA-2 SOLCAST) for hourly GHI, with 13 ground-based data in South Africa, in which the MERRA-2 reanalysis exhibits the weakest performance with a relative mean bias error (rMBE) of 11%. The authors recommended the use of the CAMS (rMBE = 2.14%) and SARA-2 (Surface Solar Radiation Data Set – Heliosat; rMBE = 2.13%) datasets for solar energy applications in the country. In West Africa, [25] showed that ERA5 provided a good representation of daily GHI compared to ERA-Interim datasets at four

Table 1

The 51 AWSs used in this study with their basic measurement characteristics and pyranometer features.

ID	Station	Institution	Temporal Resolution	Climatic Zone	Pyranometer model	Maximum range (W/m ²)	Spectral range (µm)	Sensitivity (µV/W/m ²)			
1	Abetifi	GMet	10-min	Guinea	SMP12 Class A	2000	0.285–2.75	5–20			
2	Ada										
3	Akim_Oda										
4	Akosombo										
5	Akuse										
6	Axim										
7	Sefwi_Bekwai										
8	Tarkwa										
9	Tema										
10	Nakpaboni										
11	Wa_varenpera										
12	Fumbisi										
13	Yendi	WASCAL	5-min	Savannah	CMP3		0.3–2.8				
14	Jirapa				SMP12 Class A		0.285–2.75				
15	Loagri										
16	Oualem								First class Global Solar Radiation Sensor (RSG1)	0.3–3	~10
17	Nebou										
18	Doninga										
19	Aniabisi										
20	Bongo_Soe										
21	Tabou										
22	Gwosi										
23	Kpandai										
24	Manga										
25	Tuna										
26	Kpando	ANAM	15-min	Guinea	SP-Lite		0.4–1.1	~75			
27	Bagre										
28	Bama										
29	Banfora										
30	Batie										
31	Beregadougou										
32	Bitou										
33	Diebougou										
34	Fara										
35	Hounde										
36	Boromo										
37	Gaoua										
38	Bobo_Dioulasso			Sahel							
39	Guilougou										
40	Bani										
41	Boulsa										
42	Bousse										
43	Gayeri										
44	Gorom										
45	Kamboince										
46	Kouka										
47	Barsalougho										
48	Djibo										

(continued on next page)

Table 1 (continued)

ID	Station	Institution	Temporal Resolution	Climatic Zone	Pyranometer model	Maximum range (W/m ²)	Spectral range (μm)	Sensitivity (μV/W/m ²)
49	Dedougou							
50	Dori							
51	Bogande							

Table 2

Characteristics of different satellite and reanalysis datasets used in this study.

Data	SARAH-2	CAMS	MERRA-2	ERA5
Date type	satellite	Satellite	reanalysis	reanalysis
Spatial resolution	0.05 x 0.05 (~5 km)	Interpolation to the point of interest	50 km	31 km
Temporal resolution	30 min, day, month	1 min, 15 min, 1 h, day, month	1 h	1h
Radiative transfer model	LibRadTran [37]	LibRadtran [37]	Community Radiative Transfer Model [51]	RTTOVv11
AOD source	ECMWF-MACC	CAMS global services	Advanced Very High-Resolution Radiometer (AVHRR)	Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model
Spatial and temporal of AOD	120 km; monthly	40 km; 3-hourly	1.1 km	prescribed monthly climatology.
Time period	1983 to present	2004 to present (2 days delay)	1980 to present (2 months delay)	1979 to present
Area of coverage	Europe, Africa, Atlantic Ocean	Europe, Africa, Middle East, Eastern of South America, Atlantic Ocean	Global	Global
Data policy	Free	Free	free	Free

weather stations in Burkina Faso for the year 2017. Later, [26] used three radiometric observations from the African Monsoon Multidisciplinary Analysis program (AMMA) to validate the daily and monthly GHI against the SARAH-2 dataset. On both temporal scales, the SARAH-2 performed relatively well but with notable biases. However, GHI was evaluated on a daily and monthly basis with a limited number of stations in these studies, while hourly GHI data are essential for accurate solar power plant design and planning. Moreover, knowledge of hourly GHI is useful for GHI forecasting [27]. A detailed validation process with high-quality data is needed to substitute GHI from ground-based measurements to GHI satellite-based or reanalysis data. To our knowledge, no such study has used hourly GHI from dense

observation networks to validate GHI derived from satellite and reanalysis data over West Africa.

Therefore, this study aims to evaluate the performance of hourly GHI derived from MERRA-2, ERA5, SARAH-2 and CAMS data with ground-based data for the year 2020 for solar energy monitoring. For the first time in Africa, 51 automatic weather stations (AWS) are used for hourly GHI assessment. The AWS belongs to four different transboundary and national networks recently established by WASCAL, the Ghana Meteorological Agency (GMet) and the Burkina Faso National Meteorological Agency (ANAM) and partner institutions covering the most critical climate zones (Guinea, Savannah, and the Sahel) in West Africa. The focus of this study is on the evaluation of the different satellite and reanalysis datasets based on observations under different atmospheric conditions: (i) cloudy-sky, (ii) clear-sky and (iii) all-sky. This is realized by using a wide range of performance measures and methods and introducing a novel multi-objective performance measure to select the best performance among the datasets for the region. In addition, the effect of aerosols on the hourly GHI during the Harmattan period over the area is investigated.

The paper is structured as follows. The following section presents the study area, the detailed information on the different datasets, and the methodology used. Section 3 presents the outcomes of the study and highlights the discussion of the various findings of the study. The study ends with conclusions and general recommendations regarding satellite and reanalysis based on GHI information.

2. Materials and methodology

2.1. Study area

The study focuses on the West African region, particularly Burkina Faso and Ghana (Fig. 1). The region is governed by the West African Monsoon (WAM) which modulates atmospheric processes and triggers most of the rainfall in the region [28]. West Africa is characterized by a long dry season and a rainy season (during the summer months) with annual rainfall ranging between 150 and 2500 mm [29]. The Harmattan period lasts from late November to mid-March and transports dust from the Sahara Desert across the region [30]. The strong environmental transitions from the Guinean forests in the south to the hyper-arid

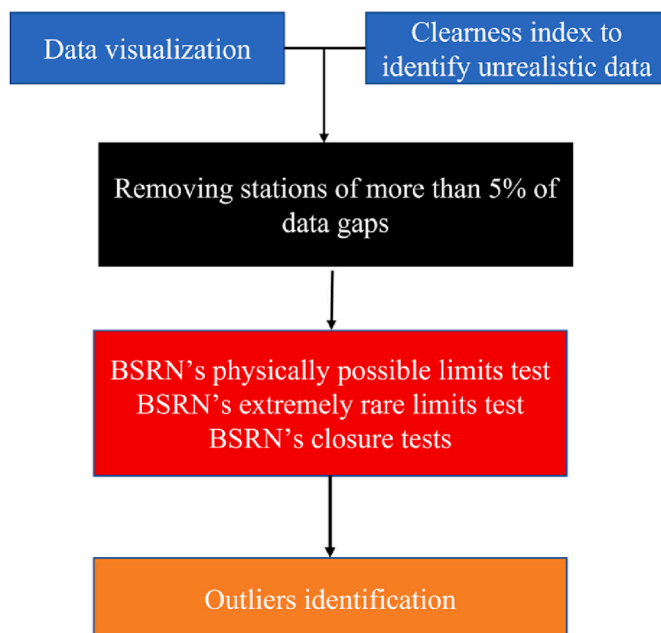


Fig. 2. Flowchart of the quality control of the ground-based measurement used in this study.

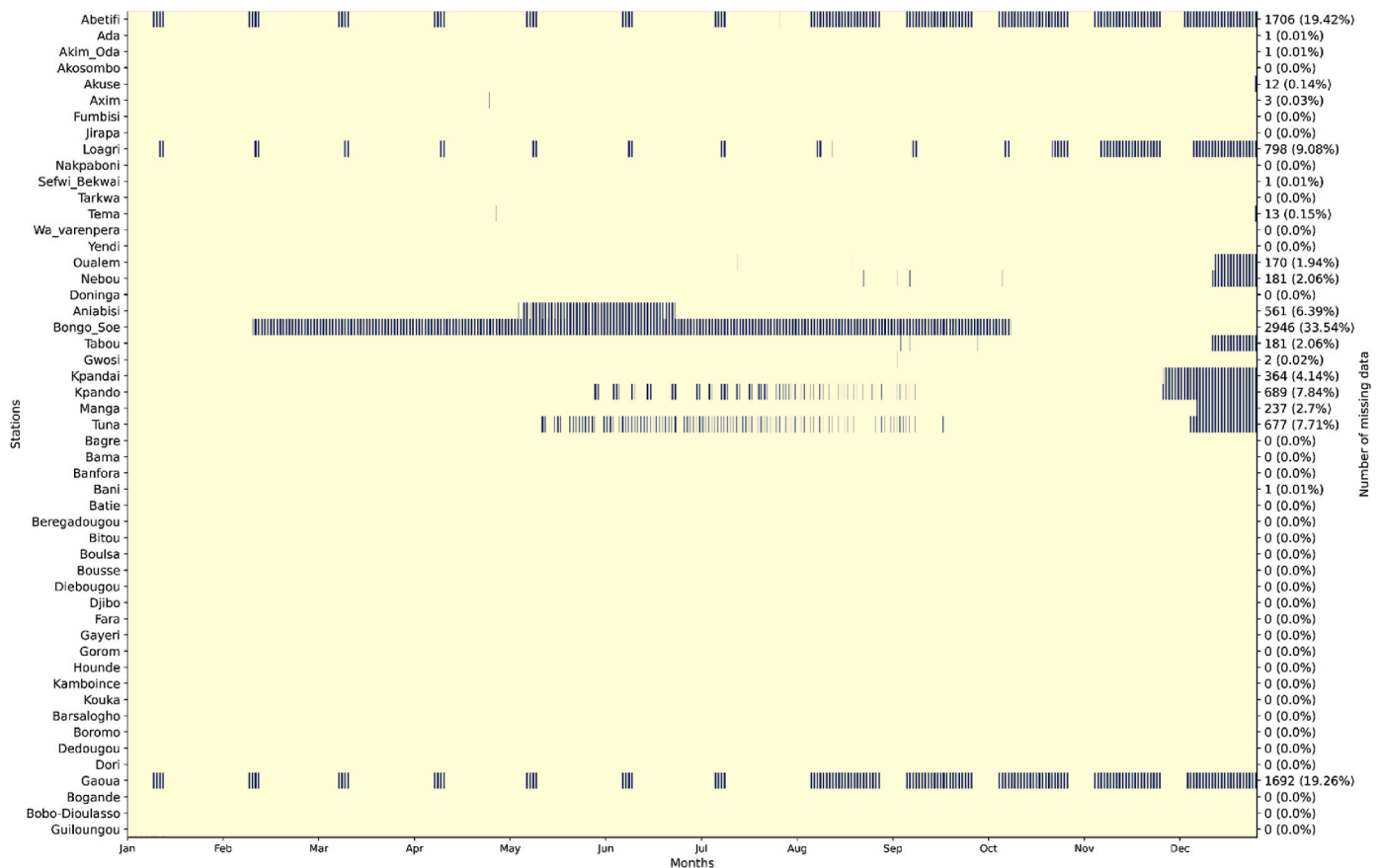


Fig. 3. Heatmap showing the missing values spread over the whole year for all the radiometric stations. The vertical black indicates a missing hour value. The total number of missing hours and the percentage is given on the right side.

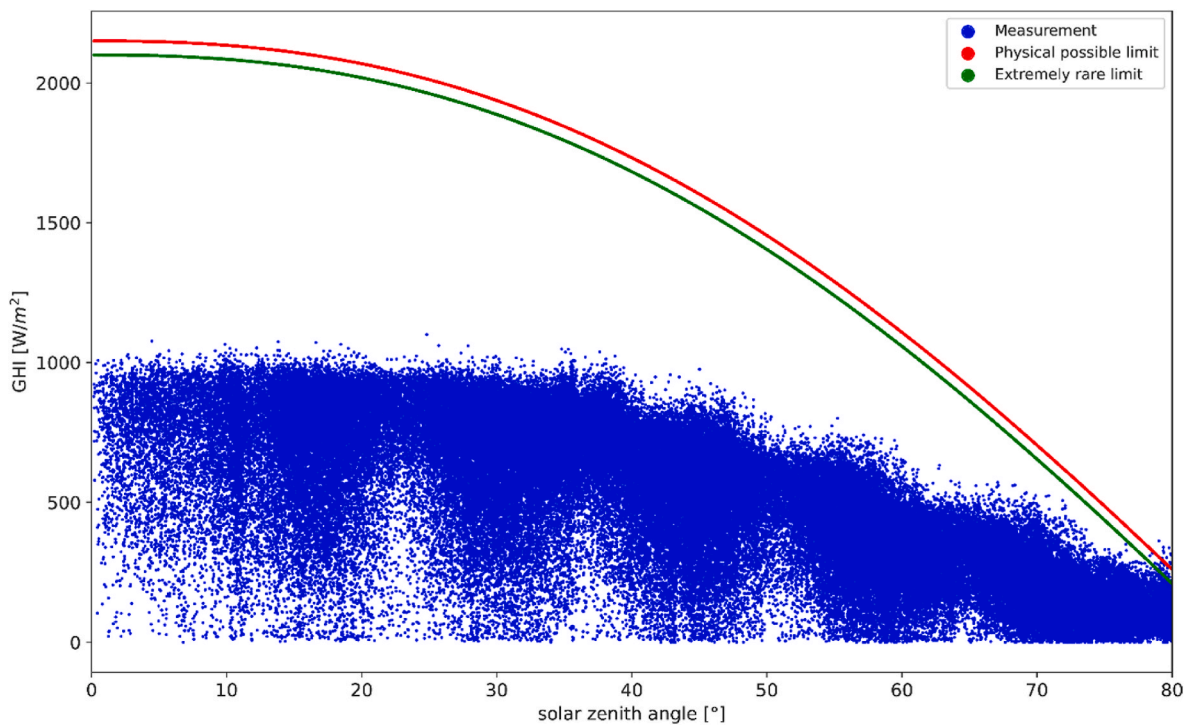


Fig. 4. Quality control of the 38 weather stations based on the Baseline Surface Radiation Network (BSRN). The measured hourly GHI are represented in blue dots. The red dots indicate the physically possible limit, while the extremely rare limit is in green dots.

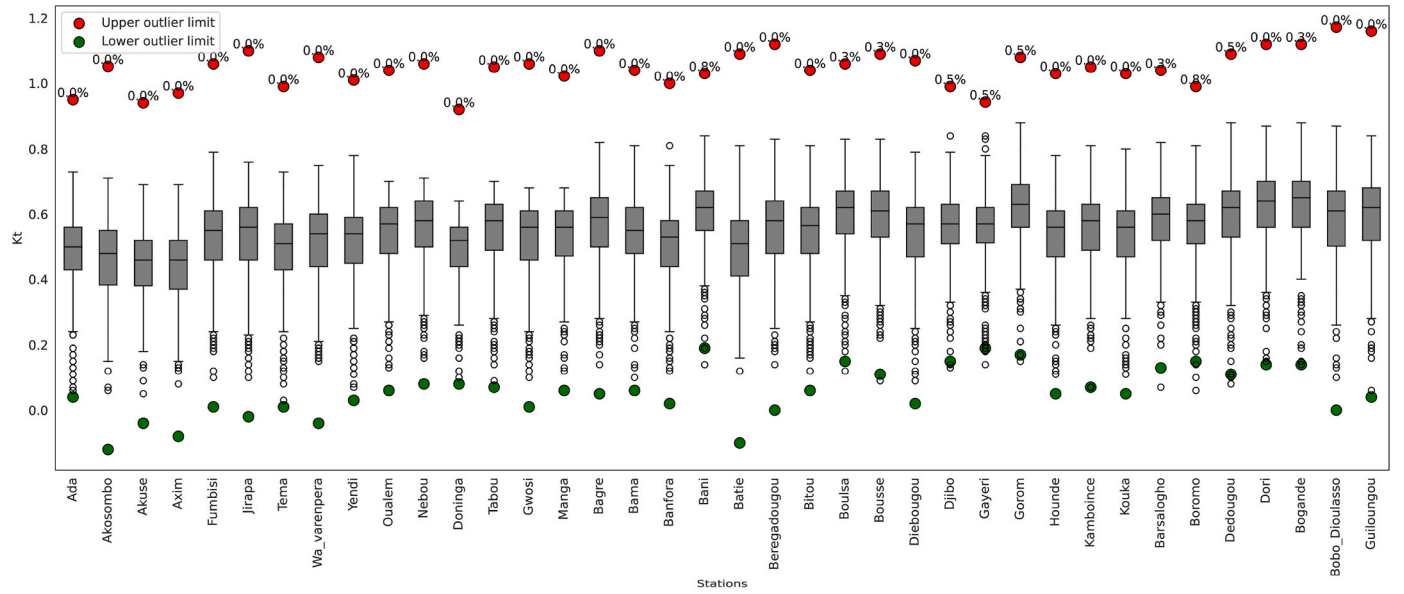


Fig. 5. Boxplot of the daily clearness index (Kt) of the different AWSs for the year 2020. The red dots indicate the upper outlier limit, while the green dots indicate the lower outlier limit of the individual stations. The number indicates the percentage of data points that fall outside the upper and lower outlier limits.

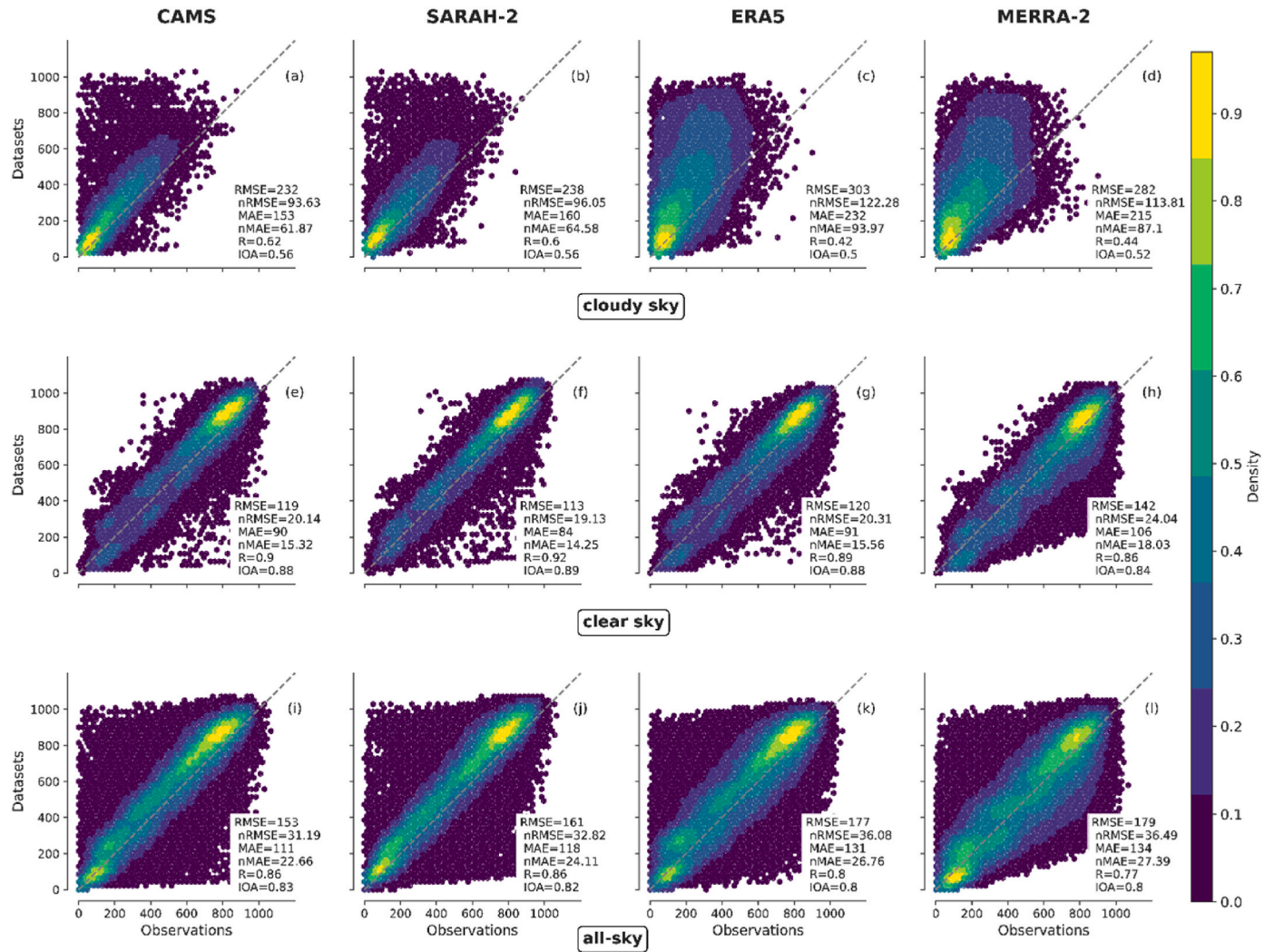


Fig. 6. Density plot of hourly GHI values from different datasets (CAMS, ERA5, SARAH-2, and MERRA-2) against observation for 37 stations using Gaussian kernels with normalized values of 0–1 for different sky conditions. The RMSE, R, IOA, and MAE denote the root-mean-square error, the Pearson correlation, the index of agreement, and the mean absolute error, respectively, while nRMSE and nMAE denote the normalized RMSE and normalized MAE, respectively.

Sahara Desert in the north, the region can be divided into three distinct climatic zones: Guinea (4°N–8°N), Savannah (8°N–12°N) and Sahel (12°N–16°N) [31] as shown in Fig. 1. The Guinea region is categorized as having a tropical monsoon climate near the coast and a tropical wet and dry climate in other areas. The zone is characterized by a humid climate and has an annual rainfall of 1250–1500 mm with a bimodal rainfall distribution. The intense presence of low clouds is common, while deep convective clouds are rare in this zone [32,33]. In addition, mid-level dust layers in the troposphere can occur in this area during the Harmattan period. The Savannah (tropical wet and dry climate) and Sahel (hot semi-arid climates) zones are semi-arid areas with average annual rainfall of 750–1250 mm and below 750 mm, respectively. Both zones have a unimodal rainfall distribution and are places where deep convective cloud activity is often associated with mesoscale convective

systems and heavy rainfall during the summer monsoon (from June to September), which peaks in August. The Sahel zone is known as a predominantly cloud-free zone and is an important source of mineral dust.

2.2. Data

2.2.1. Ground-based measurements

Fig. 1 shows the spatial distribution of the 51 AWSs used in this study. The different AWSs measure in most cases several meteorological variables such as relative humidity, wind speed and direction, precipitation, air temperature at 2 m height and GHI. Of the 51 AWSs, 7 are owned by WASCAL, 15 are the property of GMet, 4 belong to WASCAL-GMet and 25 to ANAM.

The AWSs of the WASCAL network are part of a mesoscale research observation network established by WASCAL and partner institutions in the Sudan Savannah in Ghana and Burkina Faso in 2012 and 2013 [6, 34]. Measurements from this network are made at a temporal resolution (average over each 5 min) and standard equipment maintenance such as cleaning radiation sensors is carried out regularly (e.g., twice a month).

GMet operates a surface observation network of 120 weather stations in Ghana, which are well distributed across the country. In late 2018 and early 2019, 22 novel AWS were installed by GMet and radiation measurements of which 15 AWS were are in the current study. The temporal

Table 3

Error metrics of different datasets and atmospheric conditions on GHI of the aggregated 37 stations. The bold number shows the best metric values.

Sky condition	Metric	CAMS	SARAH-2	ERA5	MERRA-2
Cloudy	RMSE (W/m2)	232	238	303	282
	nRMSE (%)	93.63	96.05	122.28	113.81
	R	0.62	0.60	0.42	0.44
	MAE (W/m2)	153	160	232	215
	nMAE (%)	61.87	64.58	93.97	87.10
	IOA	0.56	0.56	0.50	0.52
Clear	RMSE(W/m2)	119	113	120	142
	nRMSE (%)	20.14	19.13	20.31	24.04
	R	0.90	0.92	0.89	0.86
	MAE (W/m2)	90	84	91	106
	nMAE (%)	15.32	14.25	15.56	18.03
	IOA	0.88	0.89	0.88	0.84
All	RMSE (W/m2)	153	161	177	179
	nRMSE (%)	31.19	32.82	36.08	36.49
	R	0.86	0.86	0.80	0.77
	MAE (W/m2)	111	118	131	134
	nMAE (%)	22.66	24.11	26.76	27.39
	IOA	0.83	0.82	0.80	0.80

Table 4

Kolmogorov-Smirnov (KS) metric values for CAMS, SARAH-2, ERA5, and MERRA-2 datasets under different atmospheric conditions and datasets.

Sky condition		CAMS	SARAH-2	ERA5	MERRA-2
Cloudy	KS	0.224	0.215	0.331	0.294
	pvalue	P < 0.05	P < 0.05	P < 0.05	P < 0.05
Clear	KS	0.090	0.110	0.042	0.070
	pvalue	P < 0.05	P < 0.05	P < 0.05	P < 0.05
All	KS	0.104	0.142	0.088	0.036
	pvalue	P < 0.05	P < 0.05	P < 0.05	P < 0.05

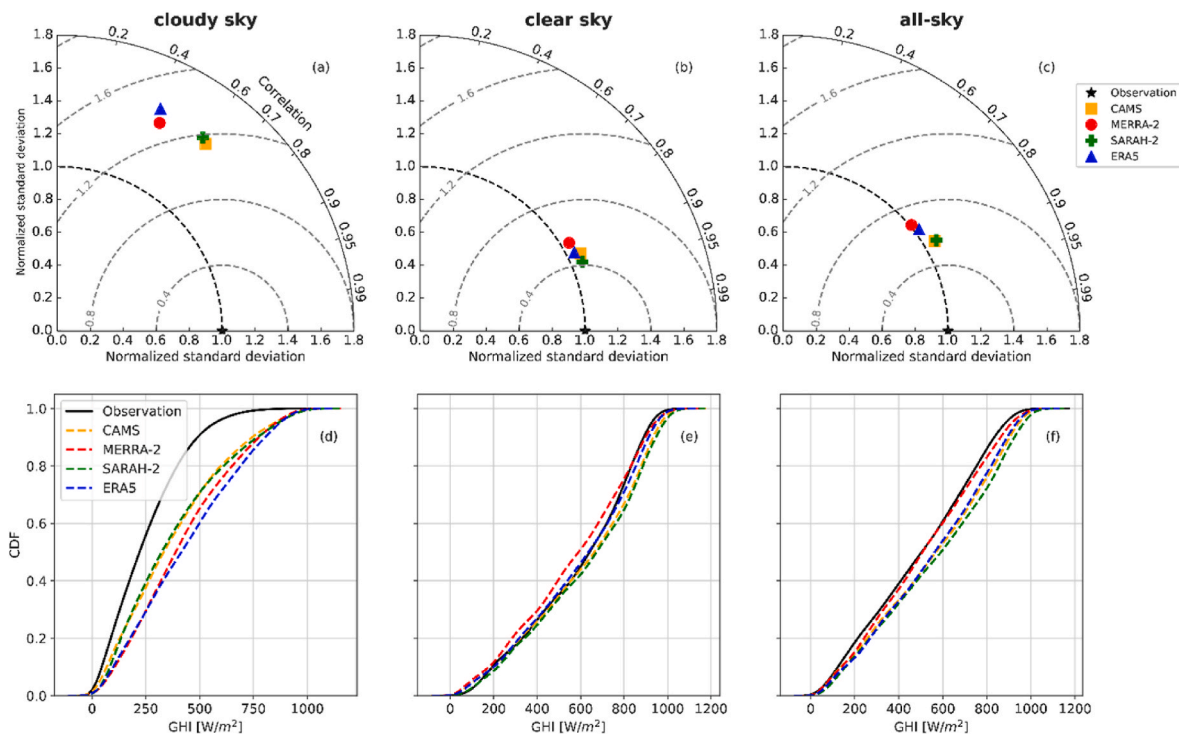


Fig. 7. The panel (a–c) shows the Taylor diagram of different datasets under clear, cloudy and all-sky conditions for the 37 stations. The dashed grey circle indicates the centered mean-square-error. The panel (d–f) presents the cumulative distribution function for the 37 stations under different atmospheric conditions.

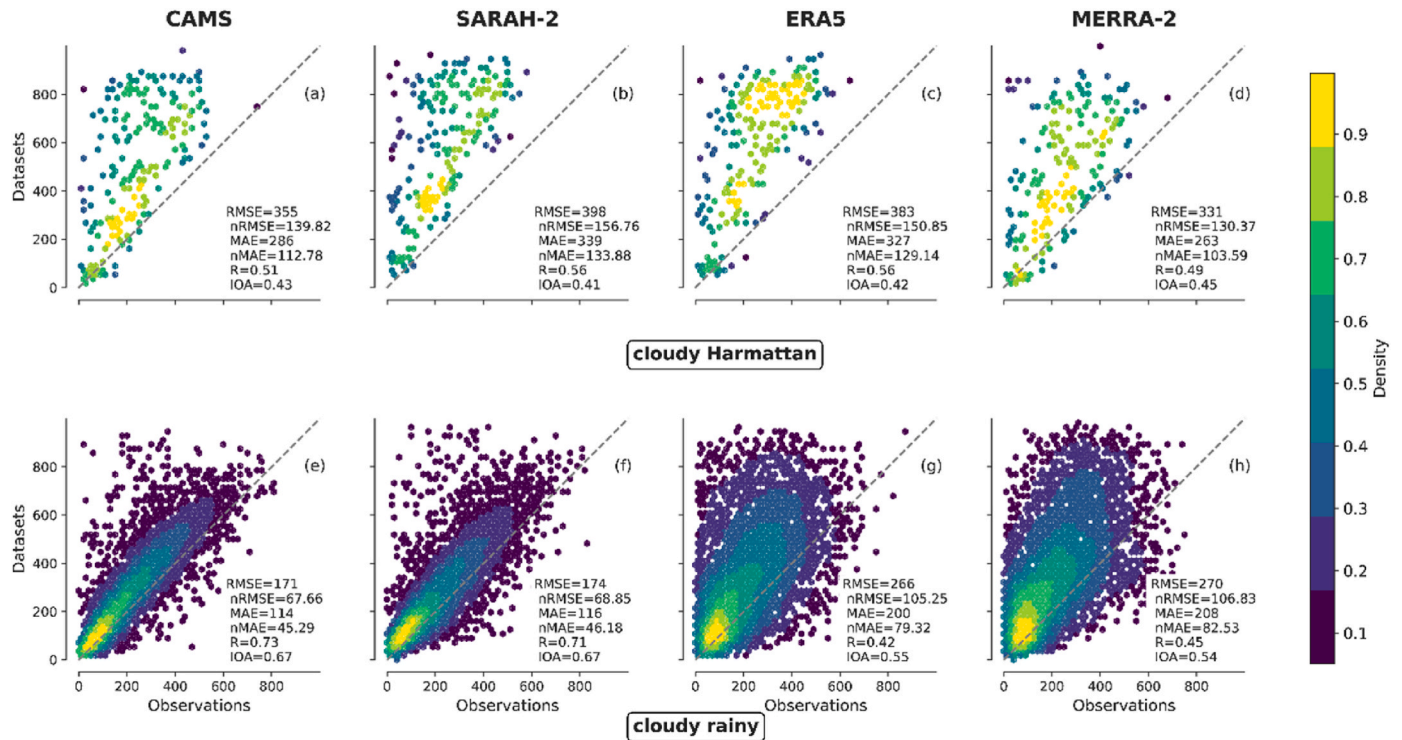


Fig. 8. Similar to Fig.6 but for cloudy days occurring during the Harmattan period (DJF) and the rainy season (JJA).

resolution of the radiation measurements is an average of 15 min. Due to the number of weather stations across the country, maintenance is done twice a year.

The WASCAL-GMet stations belong to a transboundary climate observation network established under the WASCAL programme for different West African countries [6]. 6 AWSs were donated by WASCAL through a funding from the German Federal Ministry of Education and Research (BMBF) and were installed by a joint team from both WASCAL and GMet in December 2017 after the signature of Memorandum of Understanding (MoU) on data sharing and services development. The implemented stations were handed over to GMet, which manages their maintenance. Measurements are being recorded on an average of every 10 min.

ANAM has a total of 270 weather stations across the country whereof which 22 were selected in this study, as outlined in Table 1. New AWSs were installed in 2017 in cooperation between the Burkina Faso government and its technical and financial partners. The maintenance schedule of these stations is similar to GMet, and data is recorded at 15-min intervals on an average basis. Note that all data recorded in the different AWS are subject to basic quality control (e.g., data format, measurement interval, and data consistency) by different institutions. Accordingly, to the data availability, we have collected raw data for the year 2020 to validate GHI with the datasets from ERA5, MERRA-2, SARAH-2, and CAMS datasets.

2.2.2. SARAH-2 dataset

The satellite dataset used in this study is the second edition of the Surface Solar Radiation Data Set – Heliosat Edition 2 (SARAH-2) from the Satellite Application Facility on Climate Monitoring [35]. The SARAH-2 covers the region of $\pm 65^\circ$ longitude and $\pm 65^\circ$ latitude (Europe, Africa, and the Atlantic Ocean) with a spatial resolution of 0.05° by 0.05° (~ 5 km). The dataset has a temporal resolution of 30 min (instantaneous values) and is available from 1983 to the present. The SARAH-2 products are based on the Heliosat algorithm, which incorporates the LibRadTran radiative transfer model and the MAGIC SOL clear sky model to estimate GHI under cloud-free conditions [36,37].

The GHI data used in the SARAH-2 (referred to as surface incoming shortwave radiation) product are calculated using a radiative transfer model from water vapor, surface albedo, a cloud index (from satellite observations), aerosols and ozone. SARAH-2 uses the monthly aerosol climatology from the Monitoring Atmospheric Composition and Climate (MACC) project, which has a spatial resolution of 120 km and is interpolated on the SARAH-2 grid [38]. The 30-min instantaneous values of GHI were downloaded from the SARAH-2 database (<https://wui.cmsaf.eu/>) for the year 2020. The hourly GHI is the average of two 30-min periods within 1 h.

2.2.3. CAMS dataset

The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service provides solar energy radiation products. Its algorithm for calculating these products is based on the Heliosat-4 approach [39]. The method uses the McClear algorithm to estimate GHI under clear-sky conditions [40] and the McCloud model to estimate the attenuation of solar irradiance caused by clouds. The McClear and McCloud models are implemented using the libRadtran radiative transfer model developed by Mayer and Kylling. [37]. The radiative transfer model calculates GHI under all-sky conditions by the product of GHI under clear-sky conditions with a clear-sky index, also called the cloud modification factor [39,41]. The aerosol optical depth (AOD) inputs are from the CAMS service with a spatial resolution of 40 km and are updated every 3-h. CAMS covers Africa, Europe, the Eastern part of South America, the Middle East and the Atlantic Ocean and has been available from 2004 to the present with a delay of 2 days. The data are accessible in high-temporal resolution and different resolutions (e.g., 1 min, 15 min, hourly, daily, and monthly); users can access the data up to the point of interest. In this study, we used the latest version of CAMS radiation service (version 4.5), which uses a second APOLLO NG production chain to improve cloud redundancy. We downloaded the 1-min GHI for the 51 AWS sites for the year 2020 and then computed the average hourly GHI values.

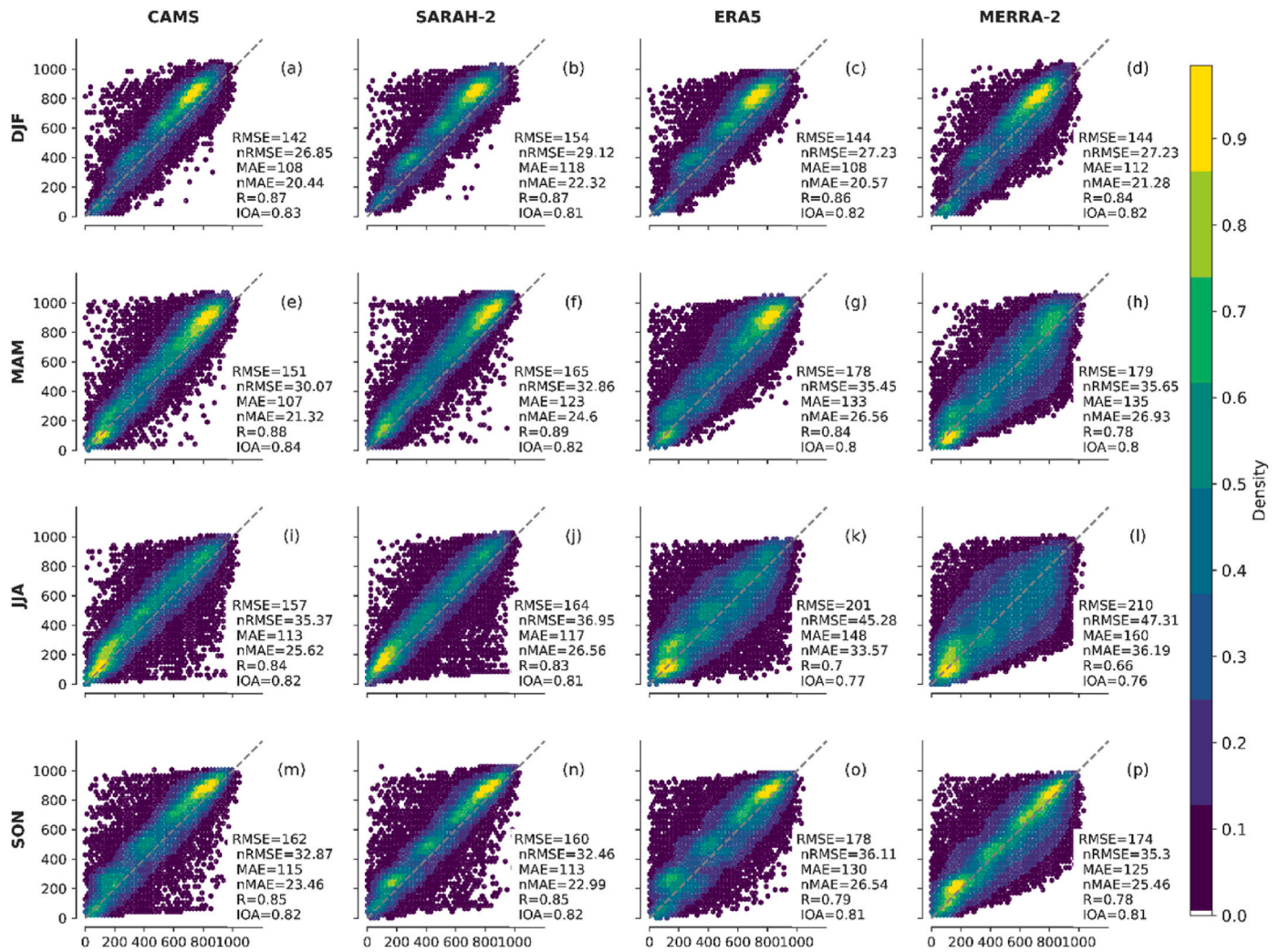


Fig. 9. Similar to Fig.6 but for all-sky conditions for different seasons.

2.2.4. MERRA-2 dataset

The Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) is a product of the NASA atmospheric reanalysis [42]. MERRA-2 replaces the original MERRA with an improved data assimilation system of the Goddard Earth Observing System Model version 5 (GEOS-5). The GEOS-5 model is coupled with the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model and simulates five types of aerosols: sulfate, dust, sea salt, and black and organic carbon [43,44]. The system includes a large-scale prognostic cloud in the moist physics scheme and uses a shortwave and longwave radiation scheme from Chou and Suarez [45] and Chou et al. [46] respectively. MERRA-2 uses real-time bias-corrected AOD inputs from the Advanced Very High Resolution Radiometer (AVHRR) instruments with a spatial resolution of 1.1 km [47]. It has a spatial resolution of 0.5° by 0.625° (~50 km) with an output of 72 model levels and 42 pressure levels from the surface to 0.01 hPa and a temporal resolution of 1h. The data cover the period from 1980 to present with a lag of 2 months. GHI hourly data were downloaded from the MERRA-2 server for the year 2020. Hourly data in MERRA-2 are averaged over the specified hour and stamped at the central hour, i.e., 00:30 GMT, 01:30 GMT, etc.

2.2.5. ERA5 dataset

ERA5 is the fifth-generation of atmospheric reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF; [48]). ERA5 has a spatial resolution of 0.25° by 0.25° (~31 km) and a temporal

resolution of 1 h. It includes 137 model levels, and 37 pressure levels and covers the entire globe. ERA5 uses the RTTOVv11 model as the radiative transfer model and “McRad” as the radiation scheme, which includes the shortwave and longwave Rapid Radiative Transfer Model for GCM (RRTMG) schemes. ERA5 uses a prescribed monthly climatological aerosol information from the Global Ozone Chemistry Aerosol Radiation and Transport (GOCART) model with a horizontal resolution of 2.5° longitude by 2° latitude which includes stratospheric sulfate aerosols [49,50]. Over West Africa, the GOCART shows a discrepancy with the observed AOD from AERONET data which is attributed to the strong perturbation of local dust source [44]. The ERA5 data are available from 1979 to the present. From the ECMWF platform, we retrieved the hourly GHI, which refers to surface solar radiation for the year 2020. The ERA5 GHI values are hourly expressed in J/m^2 . We divided the accumulated values by 3600 s to get the average GHI values in W/m^2 . The hourly data of GHI in ERA5 are computed as the mean rate of the previous hour. For example, the GHI value at 12:00:00 UTC corresponds to the average GHI from 11:00:00 UTC to 11:59:59 UTC. To ensure consistency with the observation data and other datasets where the hourly averaged is computed on the current time, we adjusted the time to a 1 h shift.

Table 2 shows the different datasets used with their characteristics. We used a linear interpolation technique to determine the radiation information from the ERA5, MERRA-2, and SARAH-2 datasets for corresponding sites of the in-situ measurements.

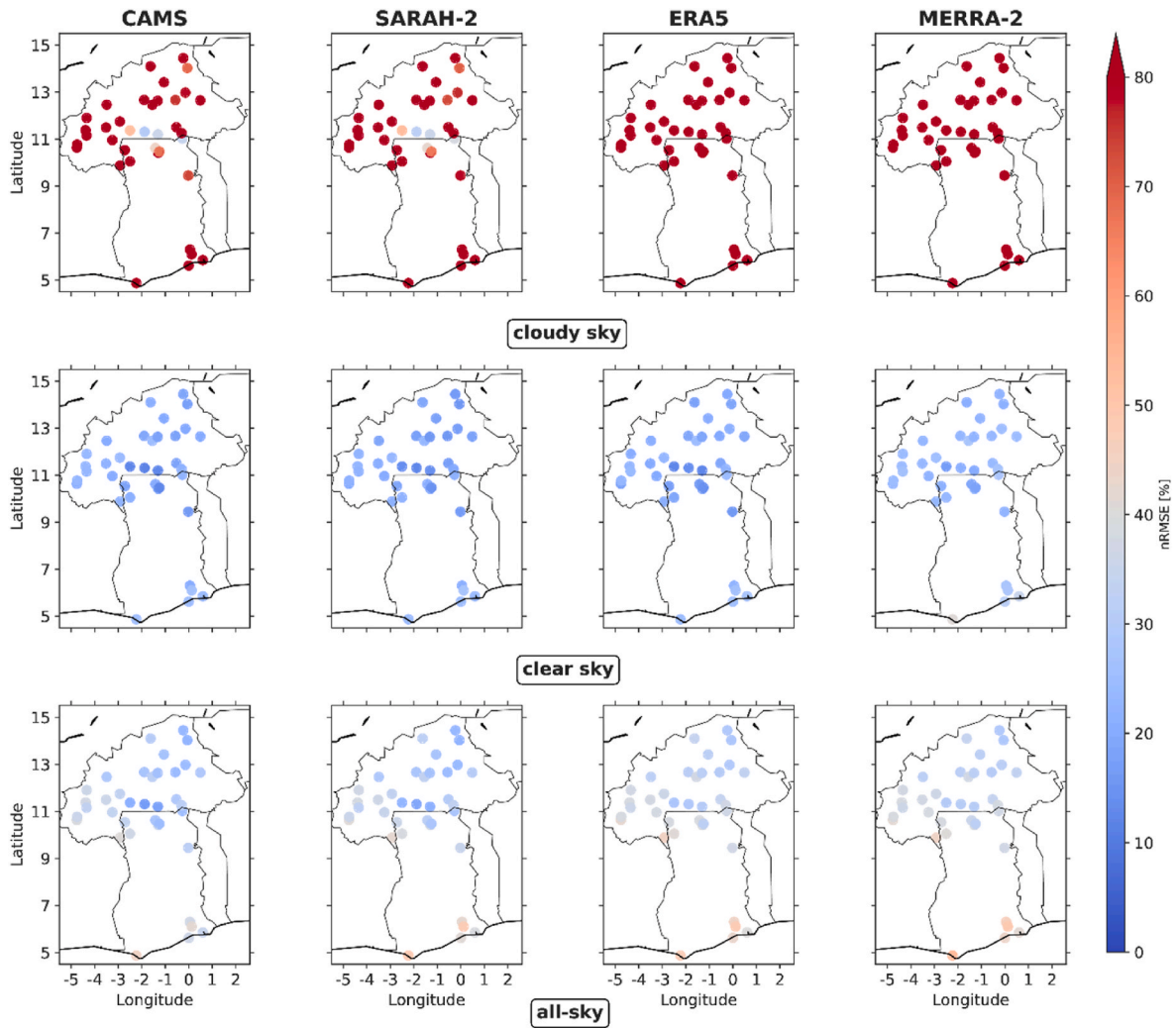


Fig. 10. Normalized root-mean-square error (nRMSE) for hourly GHI at each AWS for cloudy-sky, clear-sky, and all-sky conditions and different datasets. Each color point indicates the value of nRMSE represented by the color bar.

2.3. Methodology

2.3.1. Quality control

Fig. 2 outlines a comprehensive process for quality control of the individual weather stations. This process includes visualization of the data, various tests and techniques, identification of unrealistic values and removal of outliers to ensure data quality. These steps help improve the integrity and reliability of the station data used in this study.

The observational data used in the study have different temporal resolutions (5 min, 10 min and 15 min, see Table 1). To compute the hourly data, the sub-hourly data were averaged using the following steps:

1. If there is a missing date in the time series, the date is added, and the value for GHI is marked as missing.
2. All GHI values during nighttime are set to 0, even if there are missing values.
3. For the 5-min data, the values of GHI are averaged to an hourly value if 95% of the measurements are available within the specific hour. Otherwise, the value is set to a missing value. For the 10- and 15-min data, 100% of the measures must be available to calculate hourly GHI values.

To validate the accuracy of the hourly GHI satellite and reanalysis data, reliable ground-based GHI measurements are essential. To ensure

the quality of the different AWSs, we applied the techniques shown in Fig. 2. Our first step was to exclude stations with large missing data. Fig. 3 shows the periods with missing hourly GHI data for the different AWS in 2020. The vertical bars indicate missing periods, while the sum of the missing hours and their percentage (in parentheses) can be seen on the right ordinate. Overall, 44 out of 51 stations have no data gaps or only a few missing measurements. However, stations such as Abefiti, Loagri, Aniabisi, Bango Soe, Kpando, Tuna and Gaoua have a much higher percentage of missing data between 5% and 33.5%. For the data quality assessment, these stations were excluded. No gap-filling techniques were applied to stations with less than 5% missing values. All missing data were removed from the station in question and the extracted coordinate of this station were subjected to the same exclusion process in the corresponding satellite and reanalysis datasets.

The second step was to categorize the different AWS based on their respective climate zones (see Appendix Figs. 17–20). We then excluded stations that differed from their counterparts. Such discrepancies could be caused by shadows, faulty sensors, or calibration problems. We also combined this analysis with the clearness index (K_t) to identify suspicious AWS. For this purpose, we calculated the daily average (K_t) for all AWSs. The K_t is defined as the ratio of surface solar irradiance to extraterrestrial solar irradiance G_0 and is expressed as follows:

$$K_t = \frac{GHI}{G_0} \quad (1)$$

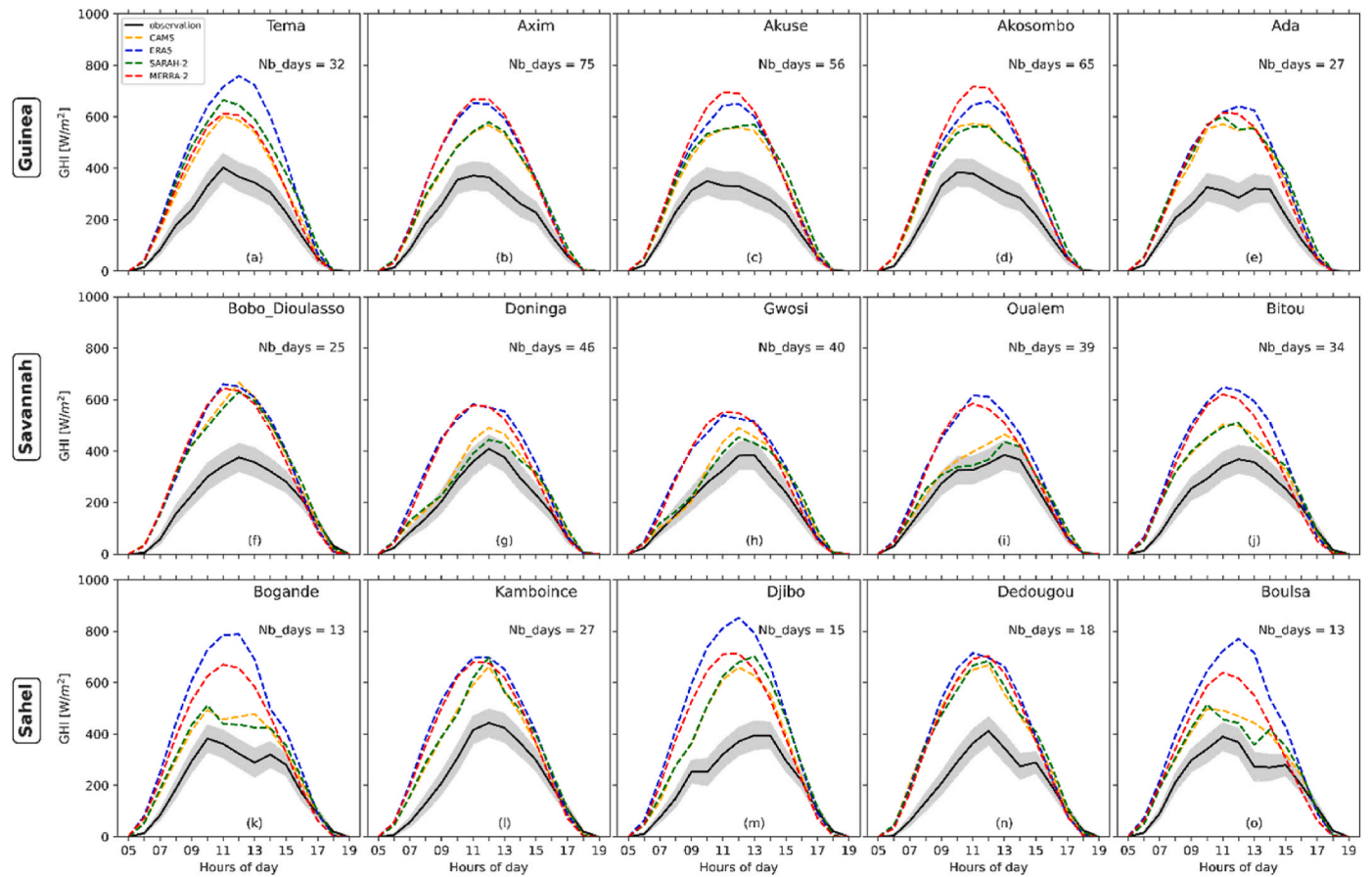


Fig. 11. Average diurnal variation of the observed GHI compared with the CAMS, ERA5, SARAH-2, MERRA-2 dataset for the selected stations under cloudy skies. Nb_days means the number of days that fall in clear skies conditions. The grey shaded curve indicates the 95% interval confidence of the measurement.

The daily GHI is determined from the hourly GHI if there is no single missing value.

The clearness index has been used in previous studies to identify sky conditions. For instance, Du et al. [24] classified the sky conditions using K_t to validate the MERRA-2 hourly dataset for clear-sky and cloudy conditions over China. However, the values of K_t used to define cloudy and clear skies vary by location. [52] have used the modified clearness, K_t' introduced by Perez et al. [53]; for clear skies they used $0.65 < K_t' \leq 1$. On the other hand, [54] have used the diffuse fraction, K_d , and established the range of $0 \leq K_d \leq 0.26$ to correspond to clear skies worldwide. This study describes clear-sky when $K_t \geq 0.6$ and cloudy-sky when $0.12 \leq K_t < 0.35$. These values were adopted from previous studies on West Africa [55–57]. Based on this information, the number of clear-sky days and cloudy days was calculated for each station, and those stations with no realistic clear-sky days throughout the year were removed (see Fig. 21 in the Appendix). After the first and second steps, only 38 stations passed these tests and were used for other quality checks.

The third step was to identify GHI values that are outside the normal range of the 38 AWSs, we therefore used the extremely rare limit (Eq. (2)) and the physically possible limit (Eq. (3)) of GHI measurements from the BSRN guidelines [58].

$$-2 \text{ W/m}^2 < GHI < I_0 \cdot 1.5 \cdot \cos(SZA)^{1.2} + 50 \text{ W/m}^2 \quad (2)$$

$$-4 \text{ W/m}^2 < GHI < I_0 \cdot 1.5 \cdot \cos(SZA)^{1.2} + 100 \text{ W/m}^2 \quad (3)$$

where I_0 the solar constant (1367 W m^{-2} ; [59]) and SZA is the solar zenith angle. For the BSRN's closure tests, the analyses were done when $SZA < 80^\circ$ to account for the seasonality of sunrise and sunset over the

region.

Fig. 4 illustrates the quality control of the hourly GHI aggregated data for all stations based on the Eqs. (2) and (3). The physically possible limit is drawn in red, the extremely rare limit in green. The blue dots indicate the individual hourly GHI measurements for all 38 weather stations. Most data points that fall outside the BSRN interval are for $75^\circ < SZA < 80^\circ$. These intervals correspond to early morning and late afternoon measurements, i.e., 7am–8am and 5pm–6pm, respectively according to the region. At some stations such as Oualem, Nebou and Mange (see Tab.5 in the Appendix), there are some data points that show a high value of GHI under conditions of low irradiance and high zenith angle. These deviations could be due to interfering reflections from the roof edge in the early morning and late afternoon hours [60]. These data points have GHI values that are above the physically possible and extremely rare limits GHI. About 649 (0.44%) such data points were flagged and removed from the analysis.

In the last step, we employed outliers to identify erroneous GHI from the different AWSs. In this study, we analyzed a far outlier for observation, which is calculated as follows [61]:

$$\text{Upper outlier limit} = 3^{\text{rd}} \text{ quartile} + 3 \times (3^{\text{rd}} \text{ quartile} - 1^{\text{st}} \text{ quartile}) \quad (4)$$

$$\text{lower outlier limit} = 1^{\text{st}} \text{ quartile} - 3 \times (3^{\text{rd}} \text{ quartile} - 1^{\text{st}} \text{ quartile}) \quad (5)$$

The outlier analysis was based on daily K_t , and we removed from the analysis data that fall outside the upper and lower limit. Fig. 5 shows the interquartile range (in grey) and the upper outlier limit (red dot) and the lower limit (green dot) of the different AWSs. There are some stations where some data points are beyond the designed bound. Consequently, with combination of other AWSs from the same area, Bani was removed from the analysis. After performing all the steps outlined in this study,

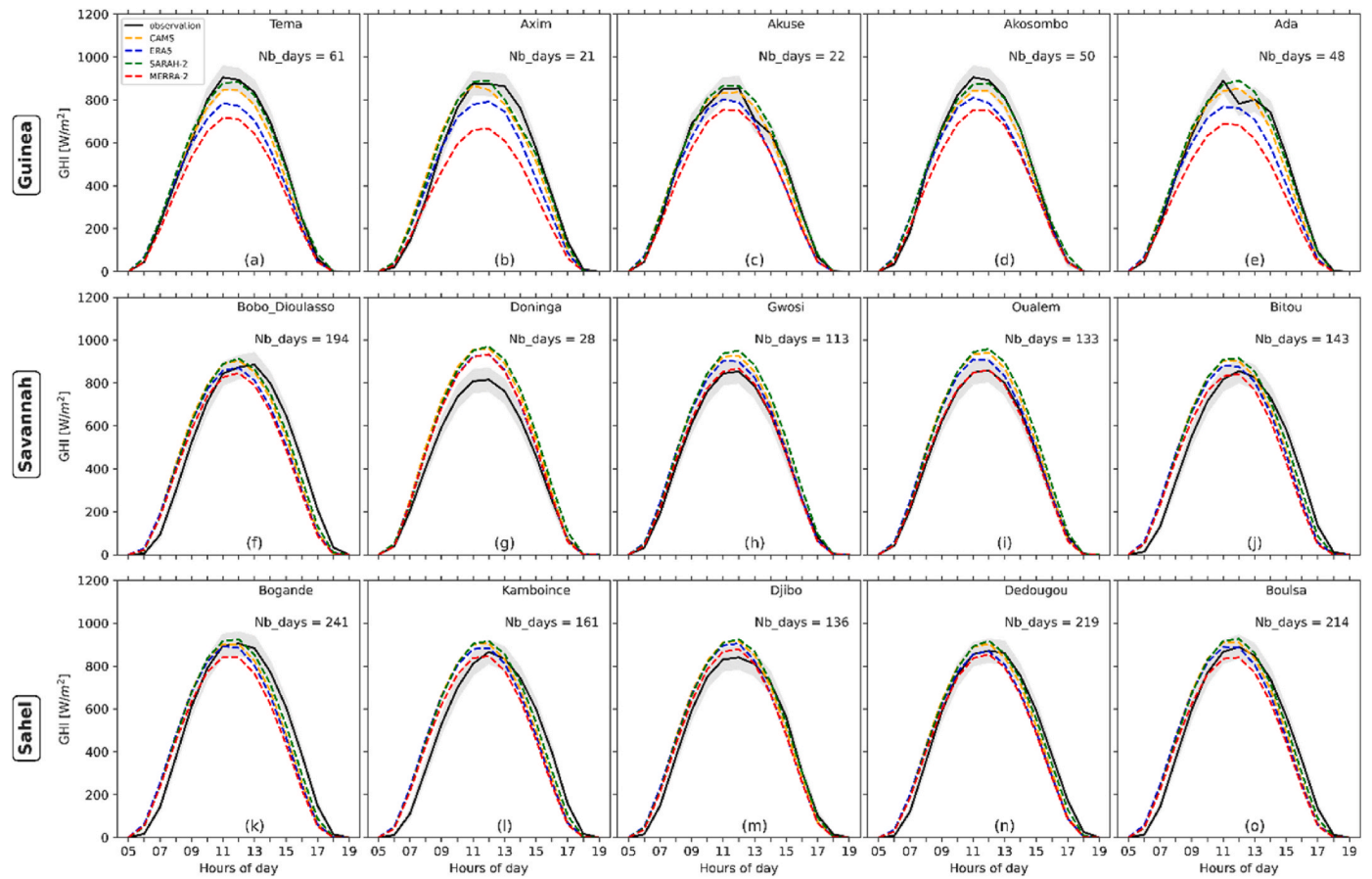


Fig. 12. Similar to Fig. 11, but for clear-sky conditions.

only 37 AWSs were used to evaluate the performance of GHI, based on satellite and reanalysis data, for the year 2020.

2.3.2. Performance metrics

The performance of the different datasets against the AWSs was assessed using several statistical metrics. We used the mean absolute error (MAE), the root mean square error (RMSE) and their normalized versions (nRMSE and nMAE) as important accuracy measures. In addition, the Pearson's correlation (R) was used to include a skill score in the current analysis. A statistical metric that is sensitive to extreme values is important for evaluating GHI. For that we applied the index of agreement (IOA), which represents the ratio between the mean square error and the potential error. The value of IOA ranges from 0 to 1; 1 means perfect agreement while 0 means no agreement [62]. The different statistical metrics are expressed as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n (|P_i - O_i|) \quad (6)$$

$$nMAE = \left[\frac{MAE}{\bar{O}} \right] * 100 \quad (7)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (8)$$

$$nRMSE = \left[\frac{RMSE}{\bar{O}} \right] * 100 \quad (9)$$

$$R = \frac{\sum_{i=1}^n (O_i - \bar{O})(P_i - \bar{P})}{\sqrt{\sum_{i=1}^n (O_i - \bar{O})^2 \sum_{i=1}^n (P_i - \bar{P})^2}} \quad (10)$$

$$IOA = 1 - \frac{\sum_{i=1}^n (P_i - O_i)}{\sum_{i=1}^n (|P_i - \bar{O}| |O_i - \bar{O}|)^2} \quad (11)$$

where P is the reanalysis or satellite data value, O the observation data at timestep i and n the number of data points used for comparison. \bar{O} and \bar{P} are the mean values of the observation and reanalysis or satellite data, respectively.

Comparing observations and different datasets using the above statistical metrics can sometimes be challenging to select the best dataset. For example, some datasets may have low RMSE, high correlation, and high IOA, while other datasets may have a low RMSE, low correlation, and low or high IOA compared to their subjects. We included therefore an additional performance measure based on the nRMSE, R and IOA to better determine the overall performance for the different datasets. Based on these metrics, a satellite or reanalysis dataset perfectly fits to the ground-based observations, if the nRMSE = 0, the IOA = 1, and the R = 1. The new overall performance measure (OP) can be expressed as follows:

$$OP = 1 - \left[\frac{nRMSE}{100} + (1 - R) + (1 - IOA) \right] \quad (12)$$

This new coefficient is dimensionless. +1 means that the dataset is perfectly close to the observation, while a negative value means that the

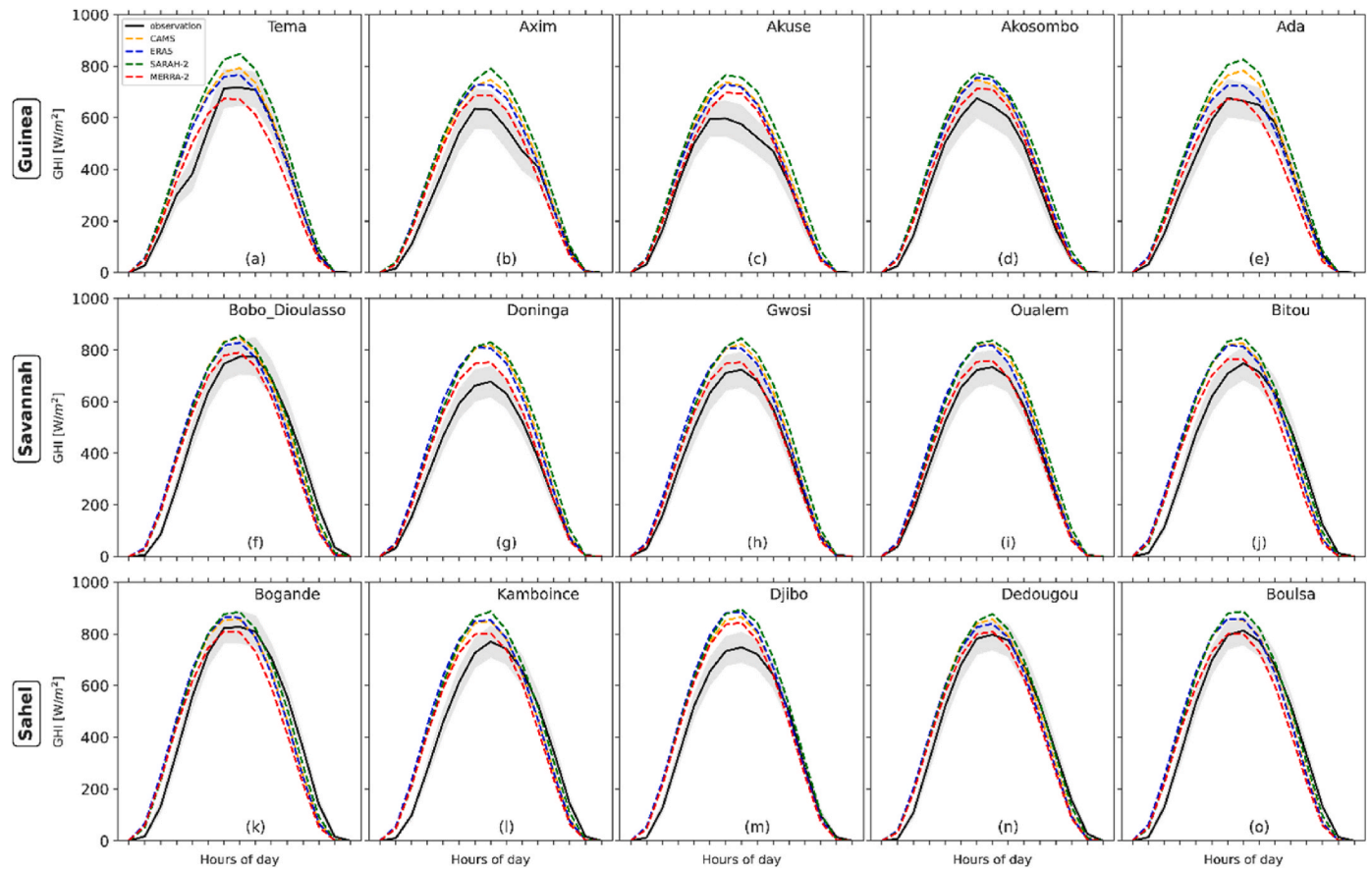


Fig. 13. Similar to Fig.11, but for all-sky conditions.

dataset is far from the observation. Moreover, the OP provides a unified grade that considers a range of statistical metrics to assess the overall performance of the dataset. It allows a more comprehensive assessment of a dataset's agreement with ground-based observations and gives valuable insight into the performance of a dataset and its suitability for a particular application or assessment.

2.3.3. Evaluation of GHI

The analysis was based on "clear-sky", "cloudy-sky" and "all-sky" conditions. The atmospheric sky condition depends on the observations. An algorithm was developed to identify the days that meet the criteria for average cloudy and clear sky days for different AWSs. Based on the day found in the observation for the sky condition classification, the same day was used as cloudy-sky or clear-sky for the different datasets. Nevertheless, the criteria used may consider the day with aerosol particles present in the atmosphere as a cloudy day. Dust aerosols and carbonaceous aerosols from biomass burning are the main aerosol types over the region. The latter aerosol type is the most important during the winter season (Harmattan period), while dust aerosol dominates in the rest of the year [44]. Therefore, we analyzed conditions on cloudy days during the Harmattan period (December-January-February) and on cloudy days during the rainy season (June-July-August). We selected 15 stations to analyze the diurnal variation of GHI. The selection was based on the representativeness of the stations in their respective climatic zones, i.e., we have taken the minimum, maximum, median, 25th percentile and 75th percentile based on the annual mean of GHI. We also used the Taylor diagram [63] and the cumulative distribution function (CDF) to evaluate the different datasets. Finally, we analyzed the performance of the different datasets under different atmospheric conditions at the seasonal level for individual stations and also for the different climate zones.

3. Results and discussions

3.1. Performance of reanalysis and satellite-based hourly GHI

The performance of the different datasets varies according to the sky conditions for the 37 AWSs (Fig. 6. a-d). High performance occurs in clear skies, while low performance occurs in cloudy skies for CAMS, ERA5, SARA-2, and MERRA-2. This performance also differs from dataset to dataset. Under cloudy skies, most data points are on the left side of the 1:1 line, i.e., all datasets overestimate the hourly GHI. The RMSE ranges from 232 to 303 W/m^2 and the MAE varies from 153 to 232 W/m^2 . CAMS shows the lowest RMSE and MAE, while ERA5 gives the highest values. In general, both satellites (CAMS and SARA-2) show good performance compared to the reanalysis data (ERA5 and MERRA-2). The biases in the reanalyses are higher than those in the satellite data. For example, the MAE in ERA5 is 303 W/m^2 (122.28%) and the SARA-2 has a value of 238 W/m^2 (96.07%). This discrepancy between the satellite and reanalysis data could be explained by the methodology used to calculate the cloud contents and their optical properties in the radiative transfer model. The cloud contents and their optical properties used in CAMS and SARA-2 come from satellite observations, while the cloud contents in the reanalysis (ERA5 and MERRA-2) are prognostic clouds [64,65]. In addition, the misinterpretation of cloudy skies as clear skies could also be a factor in the poor performance of the reanalysis (Fig. 21 a-d in the Appendix). The reanalysis data show a poor correlation (ERA5 = 0.04; MERRA-2 = 0.08) on cloudy-sky days, while the satellite data indicate a moderate correlation (CAMS = 0.22; SARA-2 = 0.23). However, all datasets show high MAE and RMSE under cloudy skies.

Under clear skies, the performance of the different datasets improved significantly compared to that under cloudy skies, with a difference of

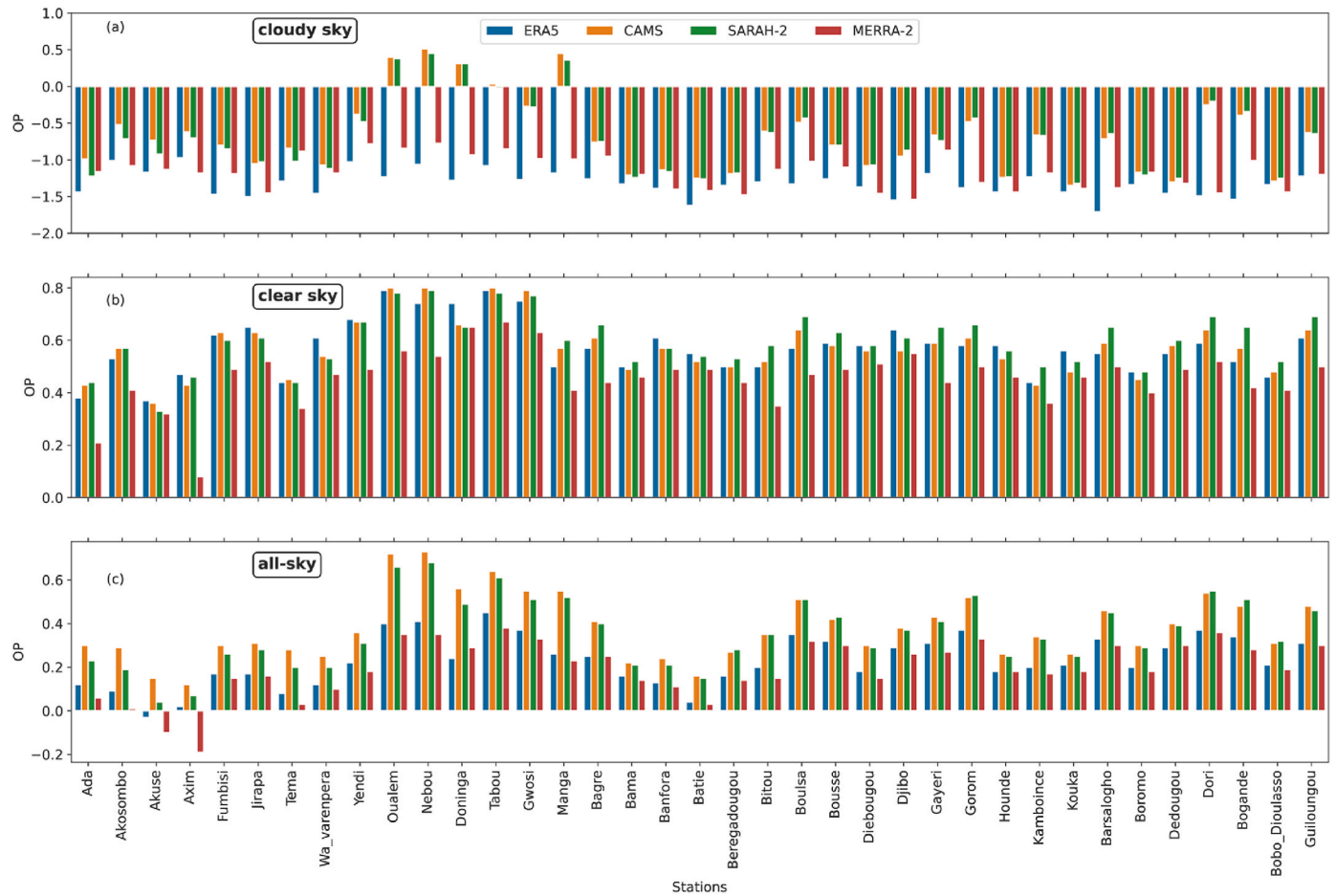


Fig. 14. Overall performance of hourly GHI for different AWS under cloud (a), clear (b) and all (c) sky conditions.

more than 150 W/m^2 in terms of RMSE (Fig. 6. e-h). This shows how difficult it is for reanalysis and satellite data to reproduce the hourly GHI under cloudy skies. The RMSE, R, and IOA of ERA5 (120 W/m^2 ; 0.89; 0.88), CAMS (119 W/m^2 ; 0.90; 0.88) are comparable, but MERRA-2 (142 W/m^2 ; 0.86; 0.84) shows poor performance under clear-sky conditions. There is good agreement between SARAH-2 and observations. The values of RMSE, MAE, R, and IOA for SARAH-2 are 113 W/m^2 , 84 W/m^2 , 0.92, and 0.89, respectively, indicating that the MAGIC SOL clear sky model used in SARAH-2 to derive GHI under cloud-free conditions performs well over the area compared to the other clear sky models used in ERA5, MERRA-2 and CAMS.

For all-sky conditions, CAMS outperforms the datasets from ERA5, MERRA-2, and SARAH-2 in the hourly estimates of GHI (Fig. 6. i-l). MERRA-2 shows poor performance with an RMSE value of 179 W/m^2 (36.49%) and a MAE value of 134 W/m^2 (27.39%). The unsatisfactory performance of MERRA-2 is the result of poor performance under a clear sky. A similar result of poor performance of MERRA-2 in hourly GHI estimation was highlighted in South Africa [4]. Moreover, our results are comparable with different sites around the world under all-sky conditions. For example, the study by Yang and Bright. [21] found that the nRMSE values for the hourly GHI of MERRA-2, ERA5, CAMS and SARAH-2 ranged from 8% to 127% under all-sky conditions. Our results are consistent with previous studies that found satellite data to perform better than reanalysis data in estimating GHI [4,21,66,67]. The statistical metrics of the datasets under different atmospheric conditions are summarized in Table 3.

Fig. 7 shows the Taylor diagram and the cumulative distribution of the hourly GHIs under different sky conditions. The Taylor diagram displays the correlation coefficient, the centralized RMSE and the

normalized standard deviation of each dataset relative to observations. A dataset performs well when it is closer to the observation, while a dataset with large differences is far from the observation. From the Taylor diagram, it is clear that the SARAH-2 and CAMS exhibits the best performance in estimating the hourly GHIs under different atmospheric conditions over the area (Fig. 7 a-c). However, the satellite and reanalysis data exhibit poor performance and each source is clustered under cloudy-sky conditions. Moreover, both satellite and reanalysis data miss the shape of the observation and overestimate the hourly values (Fig. 7 d). This shows how difficult it is to mimic the spatio-temporal variation of cloud properties with reanalysis and satellite data. In clear skies, the ERA5, MERRA-2 and CAMS are clustered with a slightly high value of the centered root-mean-square (0.6 W/m^2) from MERRA-2 compared to the SARAH-2 dataset where the value is about 0.4 W/m^2 . All datasets are able to capture the pattern of the observation, but the MERRA-2 shows a slight underestimation for values of $400\text{--}800 \text{ W/m}^2$ but agrees under all-sky conditions (Fig. 7 e-f). Under all-sky conditions, the ERA5, CAMS and SARAH-2 slightly overestimate the observed values of $400\text{--}800 \text{ W/m}^2$.

To assess how well the different datasets capture the maximum observed GHIs, we used the Kolmogorov-Smirnov (KS) Integral metric. This metric measures the maximum vertical distance between two CDFs. The KS metric ranges between 0 and 1, where 0 indicates that the CDFs are identical. Table 4 displays the significant KS values at a 95% confidence level for different datasets under various sky conditions. When compared to the satellite data, the reanalysis data demonstrate high KS values under cloudy conditions. In other words, the satellite demonstrates the capability of capturing the maximum observed GHIs with low bias compared to reanalysis. Conversely, the reanalysis data exhibit a

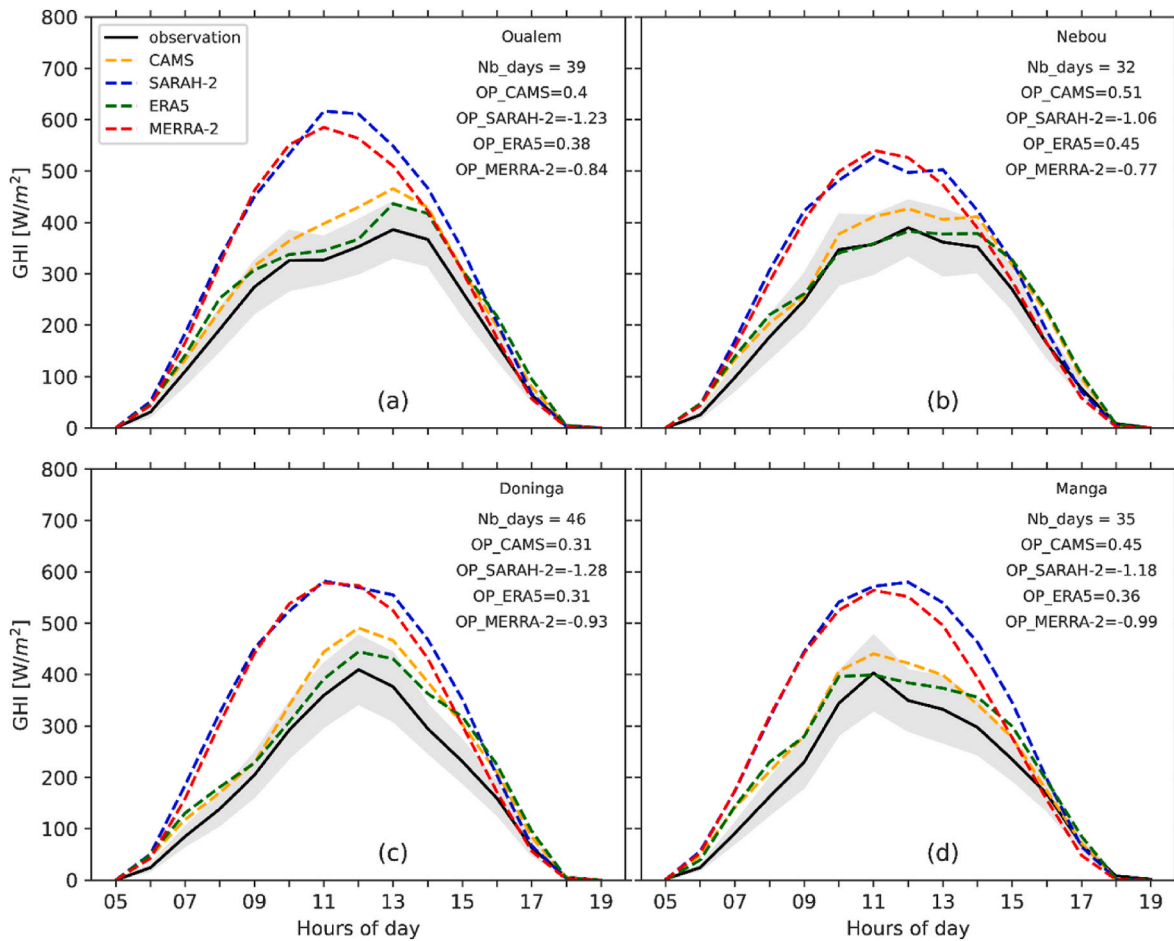


Fig. 15. Average diurnal variation of the observed GHI compared with the CAMS, ERA5, SARAH-2, MERRA-2 datasets with high positive overall performance (OP) under cloudy skies. Nb_days means the number of days that fall in clear skies conditions. The grey shaded curve indicates the 95% confidence interval of the average diurnal cycle.

low bias in capturing the maximum observed GHIs compared to the satellite data under clear skies. Overall, our analysis revealed that the ERA5 (KS = 0.088) and MERRA-2 (KS = 0.036) demonstrate a low bias in capturing the maximum observed GHIs, whereas the SARAH-2 (KS = 0.142) and CAMS (KS = 0.104) exhibit a higher bias under all-sky conditions.

To better understand the poor performance of the different datasets under cloudy skies, Fig. 8 shows a density plot of GHI for cloudy skies during the Harmattan period (DJF) and the rainy season (JJA) over the region. In general, all datasets perform better in the rainy season than in the Harmattan period. In the Harmattan period, the nRMSE value reaches 20–50% of the RMSE values in the rainy season. During the Harmattan period, trade winds transport large amounts of mineral dust from the Chad Basin to the Sahel and the Guinean coast [68]. The effect of aerosol could explain the large RMSE, and MAE found over the region under cloudy skies. The effects of aerosols as a source of large uncertainties in the estimation of GHI are well known in the literature [60, 69,70]. Among the datasets, the MERRA-2 shows the lowest RMSE (331 W/m²), MAE (263 W/m²) during the Harmattan period. The relatively better performance of MERRA-2 in DJF (Harmattan period) is also seen under all skies (Fig. 8). The AOD inputs to MERRA-2 have a spatial resolution of 1.1 km and a temporal resolution of 1h. This suggests that high spatial and temporal resolution of the AOD could improve the estimated hourly GHI over the region. However, the observed large deviation suggests that the reanalysis and satellite data did not correctly estimate the hourly GHI during the dust period. This result is consistent with [24,71]. During the rainy season under cloudy-sky (Fig. 7 e–f), the

CAMS shows the lowest RMSE (171 W/m²), while the MERRA-2 gives the highest value (270 W/m²). The good performance of SARAH-2 and CAMS under cloudy-sky could be a consequence of their performance during the rainy season. This can be confirmed in Fig. 9(i–l) where both datasets show good performance under all skies compared to that for MERRA-2 and ERA5. In the seasons of MAM (Fig. 9 e–h) and SON (Fig. 9 m–p), the satellite data also outperform the reanalysis data.

3.2. Spatial distribution of the nRMSE

Fig. 10 depicts the spatial distribution of the nRMSE over the area for different sky conditions. For a given sky condition, the nRMSE decreases from south to north, i.e., high nRMSEs are in the Guinea zone and low nRMSEs in the Sahel zone. The Sahel zone is known as a zone with low cloud cover, while the Guinea zone is a place with frequent occurrence of clouds and higher humidity throughout the year. This result leads to a similar conclusion where the reanalysis and satellite data show a large bias in the GHI estimate for cloudy regions [21,72]. Under cloudy skies, most stations have a high nRMSE in the range of 80–120%. This large bias in cloudy regions could be due to the 3D effect of clouds leading to overshoots – a feature that becomes important in the case of patchy cumulus clouds, especially if the clouds have a large height. In particular, the angle of view in each pixel by the satellite could be a relevant factor in this respect. Clouds are 3D structures, and the way they reflect, absorb and scatter light can affect the angle from which the satellite observes them [73]. On the other hand, most AWSs show low nRMSE values under clear-sky and all-sky conditions. The nRMSE values under

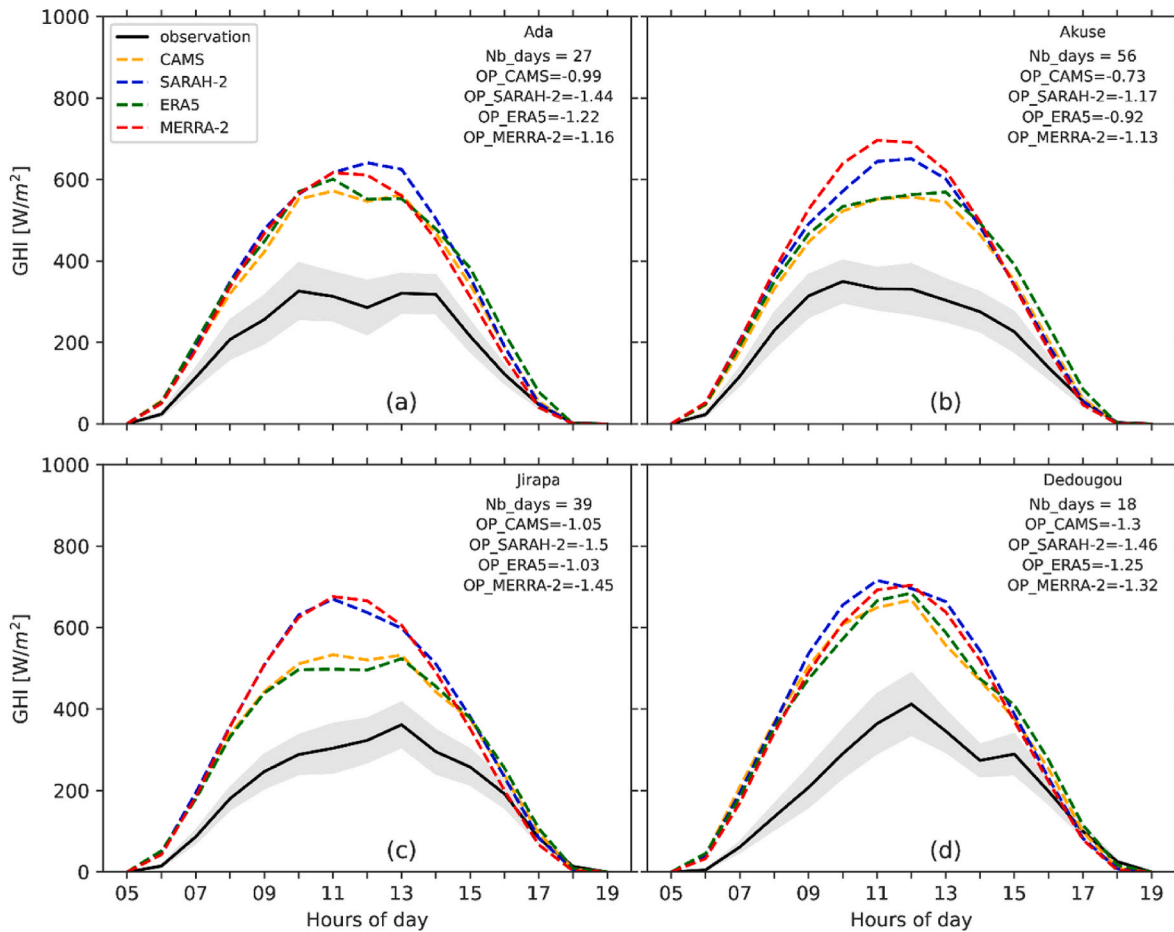


Fig. 16. Similar to Fig. 10, but for a high negative overall performance (OP).

clear-sky are better than those under all-sky conditions. The majority of the stations indicate good coherence with the datasets of the SARAH-2 and CAMS, while the ERA5 and MERRA-2 show relatively poor performance under different atmospheric conditions. The ERA5 has the highest nRMSE in most of the stations under cloudy conditions. The high biases in the ERA5 dataset could be due to overestimation or underestimation of cloud properties as reported in some studies [4,72]. However, the good performance of ERA5 has been demonstrated in some regions [66,74,75]. The discrepancy of the ERA5 performance in the studied area under cloudy conditions could be due to the low number of weather stations in the region for the ERA5 reanalysis assimilation and/or the representation of cloud properties in the dataset, as the region is located within the Intertropical Convergence Zone (ITCZ). In the region, low-level clouds are common, and it is well known that reanalysis and climate models poorly represent them [76].

3.3. Average diurnal cycle of GHI

3.3.1. Cloudy-sky conditions

The average diurnal variation between the measured and estimated values of GHI for 15 selected stations within the three climate zones under cloudy skies is shown in Fig. 11. It can be observed that the Guinea zone experiences a greater number of cloudy days compared to the Sahel zone. All datasets are able to reproduce the pattern of observed GHI but overestimate the average diurnal variation. The overestimation occurs mainly at midday for all datasets and also in the early morning and late afternoon for some of them. The overestimation in the early morning could be related to cloud cover, as there is stratus in the morning especially on the Guinea coast [77]. A minimum of convective activity

occurs over the climate zones around noon and the maximum occurs in the late afternoon (~17:00 local time) mainly at latitudes below 9° N (Guinea zone and some parts of the Savannah zone) and also above 9° N (some parts of the Savannah zone and the Sahel zone) around 20:00 [77]. In the Savannah and Sahel zones, all datasets are able to mimic the late afternoon observation well. In addition, these overestimates of the diurnal GHI pattern could also be due to the suspension of dust particles, especially during the DJF season when the reanalysis and satellite data are challenging to estimate GHI (see Fig. 8 a–d). However, the satellite data show less bias compared to that of the reanalysis data in estimating the maximum observed GHI. This is consistent with the results of Table 3. Overall, the reanalysis and satellite data show how difficult it is to reproduce the average daily variations of the selected stations under cloudy skies.

3.3.2. Clear-sky and all-sky conditions

Figs. 12 and 13 display the aggregate diurnal variations of GHI from the observation and the datasets under clear-sky and all-sky conditions, respectively. Unlike cloudy skies, most of the datasets show a good pattern of the measured GHI in most stations under clear and all skies. The number of clear sky days increases towards the north. In the Guinea zone, the ERA5 and MERRA-2 generally underestimate the maximum of the observation, while the SARAH-2 and CAMS are able to record the maximum under clear skies. In the Savannah and Sahel, most datasets also capture the maximum GHI, whereas the SARAH-2 and CAMS slightly overestimates the maximum. Similarly, in all skies, the SARAH-2 and CAMS slightly overestimates the maximum GHI. This agrees with the KS values previously mentioned (see Table 3) for both clear and all-sky conditions. In general, most datasets overestimate the maximum

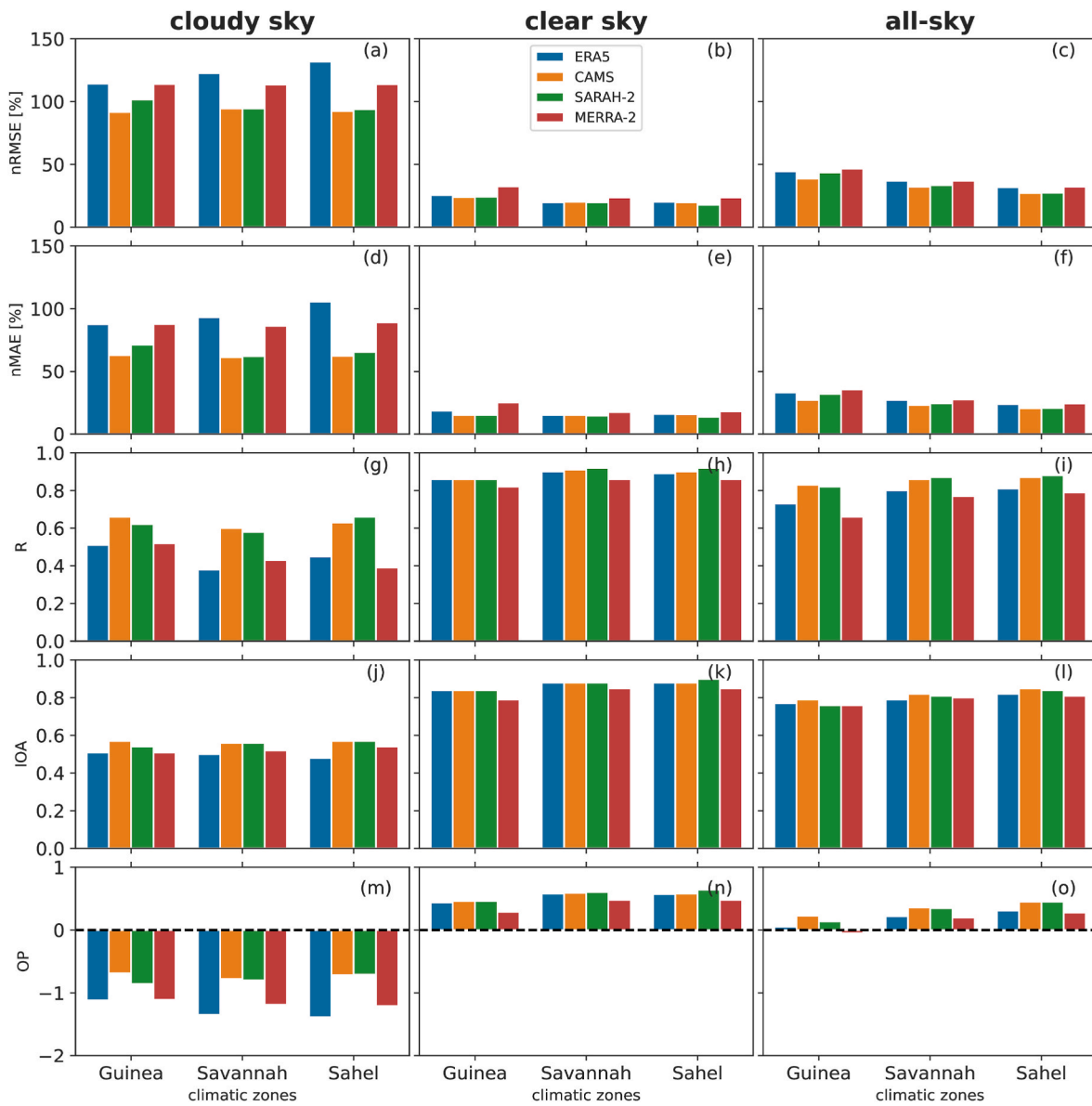


Fig. 17. Performance metrics showing the normalized root-mean-square-error (nRMSE), normalized mean absolute error (nMAE), correlation (R), index of agreement (IOA) and the overall performance (OP) for the hourly GHI in different climate zones and various sky conditions. Panels (a, d, g, j, m) show the performance of different datasets under cloudy skies, while panels (b, e, h, k, n) indicate that for clear skies. The performance under all-skies is depicted in panels (c, f, i, l, o).

GHI under all-sky conditions in all climate zones, especially in the Guinea zone. This could be the result of an overestimation of the average diurnal variation of GHI under cloudy and/or overcast sky ($K_t < 0.2$, which is not shown in this study).

3.4. Overall performance over different stations

The use of GHI, derived from reanalysis and satellite data, to assess and monitor solar energy is widespread. However, selecting the best product can be a difficult task. Here we present a new overall performance based on the nRMSE, correlation, and IOA (see Eq. (12)) to select the best product for the area. The corresponding statistical metrics (nRMSE, nMAE, R, IOA) for each station are given in the Appendix (see Figs. 22–24). Fig. 14 shows the OP of the different AWSs under various sky conditions. Under cloudy-sky conditions, all the datasets show a negative value with a maximum of -1.5 at some stations. This means that the datasets are significantly far from observations. However, the SARA-2 and CAMS show the lowest OP values compared to that for the ERA5 and MERRA-2 at most stations. Some stations like Oualem, Nebou,

Doninga, and Manga show good OP for the CAMS and SARA-2 datasets with a high positive value especially in Nebou. The OP value is about 0.5, which means that CAMS and SARA-2 are consistent with the observations. To verify this, Fig. 15 shows the average diurnal variation of four stations under cloudy conditions. We can clearly see that the stations of Nebou, Oualem, Doninga and Manga, which show a high OP value for SARA-2 and CAMS, are closer to the average diurnal variation of measured GHI in comparison with ERA5 and MERRA-2. We also plotted the average diurnal variation of GHI with stations showing a high negative OP (Ada, Akue, Jirapa, and Dedougou), as shown in Fig. 16. The average diurnal variations of all datasets are far from the observations. The results confirm that it is a good choice to use an overall performance indicator for the selection of datasets for the estimation of GHI. The satellite data, however, show the best performance at most stations under cloudy conditions.

In both clear-sky and all-sky conditions, all stations show a positive value of OP. The OP value of SARA-2 and CAMS are higher than that of the MERRA-2 and ERA5 datasets, especially in stations that belong to the Guinea and the Savannah zones. In the Sahel region, the OP values are

comparable between the ERA5, CAMS and SARA-2 under clear skies at some stations. The OP value reaches about 0.7 under clear skies in Oualem, Nebou and Manga for the SARA-2 dataset. In summary, it can be deduced from this analysis that the satellite data are better than the reanalysis data over the entire area.

We also examined the performance of different datasets at different stations and different seasons, considering different atmospheric conditions. A more detailed analysis can be found in the Figs. 25–27 in the Appendix. During the DJF season, when the sky is cloudy, we observed the highest uncertainties at each station. Most datasets showed similar values, but the MERRA-2 dataset showed relatively better results. In contrast, the satellite data performed better than the reanalysis data during the rainy season, which is consistent with the results shown in Fig. 9. Under clear skies, the datasets showed relatively low nRMSE values at each station throughout the year. However, during the JJA season we noted high nRMSE values at some stations, reaching up to 45%. This indicates larger uncertainties during this period. These results are consistent under all-sky conditions. Both the satellite and reanalysis data showed higher nRMSE values during the JJA season than in other seasons. Nevertheless, the satellite data outperformed the reanalysis data at each station overall.

3.5. Overall performance over the climate zones

Fig. 17 shows the performance metrics of different datasets in different climate zones for hourly GHI. The values were obtained by aggregating the stations in each climate zone. The Guinea zone and the different sky conditions have high values for nRMSE and nMAE with low correlation and IOA. In Guinea and Savannah, the nRMSE and nMAE values are comparable under cloudy skies. The satellite-derived data outperform the reanalysis data in the Sahel with low nRMSE (~25%) and nMAE (~20%) under cloudy skies. Under cloudy skies, all the zones show a negative OP value; the CAMS and SARA-2 datasets show the lowest value compared to that of the two-reanalysis datasets. All climate zones exhibit a positive value for clear skies and all skies, with SARA-2 and CAMS showing a higher value. The ERA5 also performs well for clear skies in all climate zones. When estimating the hourly GHIs, the satellite data outperform the reanalysis data under all-sky conditions in all climatic zones.

4. Conclusion

The aim of this study was to validate four state-of-the-art satellite and reanalysis (CAMS, SARA-2, ERA5, and MERRA-2) data using hourly GHI data from ANAM, WASCAL and GMet for the year 2020. To ensure the accuracy of the data, the ground-based measured data were subjected to strict quality controls; only 37 out of 51 stations were finally used as reference stations for analysis. The evaluation was conducted under different weather conditions, including cloudy skies, clear skies and all skies, using a new overall measure to identify the best product for the region, along with other criteria. In addition, the study examined the relationship between aerosol, clouds, and radiation during the Harmattan period and the rainy season. The results of the study can be summarized as follows:

- For the combined 37 stations, the hourly GHI values derived from satellite and reanalysis data perform better in an area with cloud-free conditions than in a cloudy region in terms of the RMSE and MAE metrics.
- Both satellite-based hourly GHI estimates perform well in cloudy conditions compared to the reanalysis data.
- MERRA-2 outperforms SARA-2, ERA5 and CAMS in estimating hourly GHI during the Harmattan period (DJF season), while SARA-2 performs best during the rainy season (JJA) under cloudy skies.

- Most datasets capture the average diurnal variation in measured GHI under cloudy and all skies, while overestimating it under cloudy skies.
- ERA5 reanalysis also shows a good performance in estimating hourly GHI under clear-sky conditions.
- The overall performance measure shows that the SARA-2 and CAMS data outperforms the ERA5 and MERRA-2 ones in all climate zones of the region and under different atmospheric conditions.

The results of this study showed that the satellite data from SARA-2 and CAMS perform well in estimating hourly GHI data over the study area and may serve as viable alternative to ground-based measurements for assessing solar energy in West Africa. However, the data showed significant biases, especially during the Harmattan period when dust is more prevalent in the region. Future research should focus on exploring the spatial and temporal resolution of the AOD data from SARA-2 and CAMS. On the other hand, the atmospheric reanalysis datasets used in this study performed poorly under cloudy conditions compared to the satellite data. It is important to note that the use of a one-year dataset could limit the generality of conclusions between reanalysis and satellite data in the region. For the poor performance of the reanalysis data, we hypothesize that the parameterization of the convective scheme and the interaction between radiation and aerosols in global circulation models needs to be improved to better capture the specific features of the monsoon, such as squall lines in this challenging region [78]. In addition to the evaluation of the GHI products, the novel AWS networks with the sub hourly GHI measurements enables many other important applications such as the evaluation of regional climate models, as shown for the Weather Research and Forecasting (WRF) model [79–81]. The data can also be used for statistical refinement of the satellite and reanalysis products to remove biases and perform spatio-temporal disaggregation of the satellite products to better meet the needs of local applications. In addition, the high-resolution measurements of the novel networks could also improve the reconstruction of weather conditions on the ground and lead to better GHI estimates over West Africa, if this information is directly incorporated into the atmospheric models that to produce reanalysis products. Thus, there are many opportunities to further improve GHI data products for solar energy applications that need to be explored in future studies for West Africa. This will enable better planning and design of PV systems and directly contribute to better meeting the rapidly increasing demand for sustainable electricity in Africa.

Funding

This research is part of the project EnerSHelf, which is funded by the German Federal Ministry of Education and Research as part of the CLIENT II program. Funding reference number: 03SF0567D. It is additionally funded partly by the WASCAL CONCERT project (funding reference number 01LG2089A, BMBF).

Data availability

The automatic weather station data collected by ANAM, WASCAL and GMet are not publicly available. Data requests can be sent to ANAM, WASCAL and GMet via the following email addresses: info@mereoburkina.bf, secretariat_cc@wascal.org and client@meteo.gov.gh. The satellite and reanalysis data can be accessed via their respective platforms.

CRedit authorship contribution statement

Windmanagda Sawadogo: Conceptualization, Methodology, Software, Figures, Formal analysis. **Jan Bliefernicht:** Writing – review & editing, Conceptualization, Methodology. **Benjamin Fersch:** Writing – review & editing, Conceptualization, Methodology. **Seyni Salack:** Writing – review & editing. **Samuel Guug:** Writing – review & editing. **Belko Diallo:** Writing – review & editing. **Kehinde.O. Ogunjobi:**

Writing – review & editing. **Guillaume Nacoulma**: Writing – review & editing. **Michael Tanu**: Writing – review & editing. **Stefanie Meilinger**: Writing – review & editing. **Harald Kunstmann**: Supervision, All authors have read and agreed to the published version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have influenced the work in this article.

Acknowledgement

We thank the European Centre for Medium-Range Weather Forecasts (ECMWF) for the ERA5 reanalysis and the Copernicus Atmosphere Monitoring Service (CAMS) datasets for producing the GHI data. We also thank the Satellite Application Facility on Climate Monitoring for providing the GHI data from SARA-2 and NASA for providing the MERRA-2 derived GHI. The authors thank the anonymous reviewers for their comments and suggestions.

Appendix

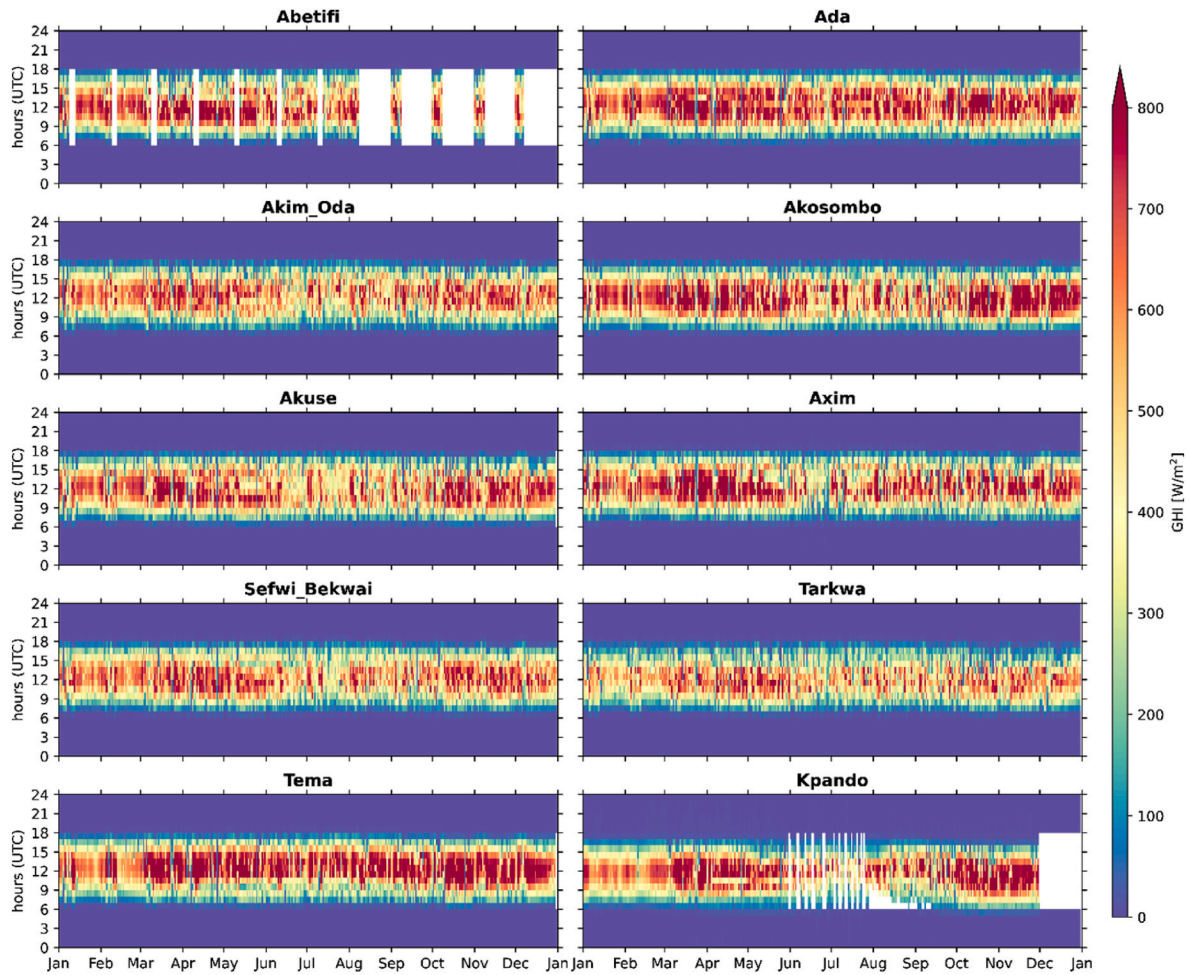


Fig. 17. Hourly GHI plot for different AWS within the Guinea zones for the year 2020.

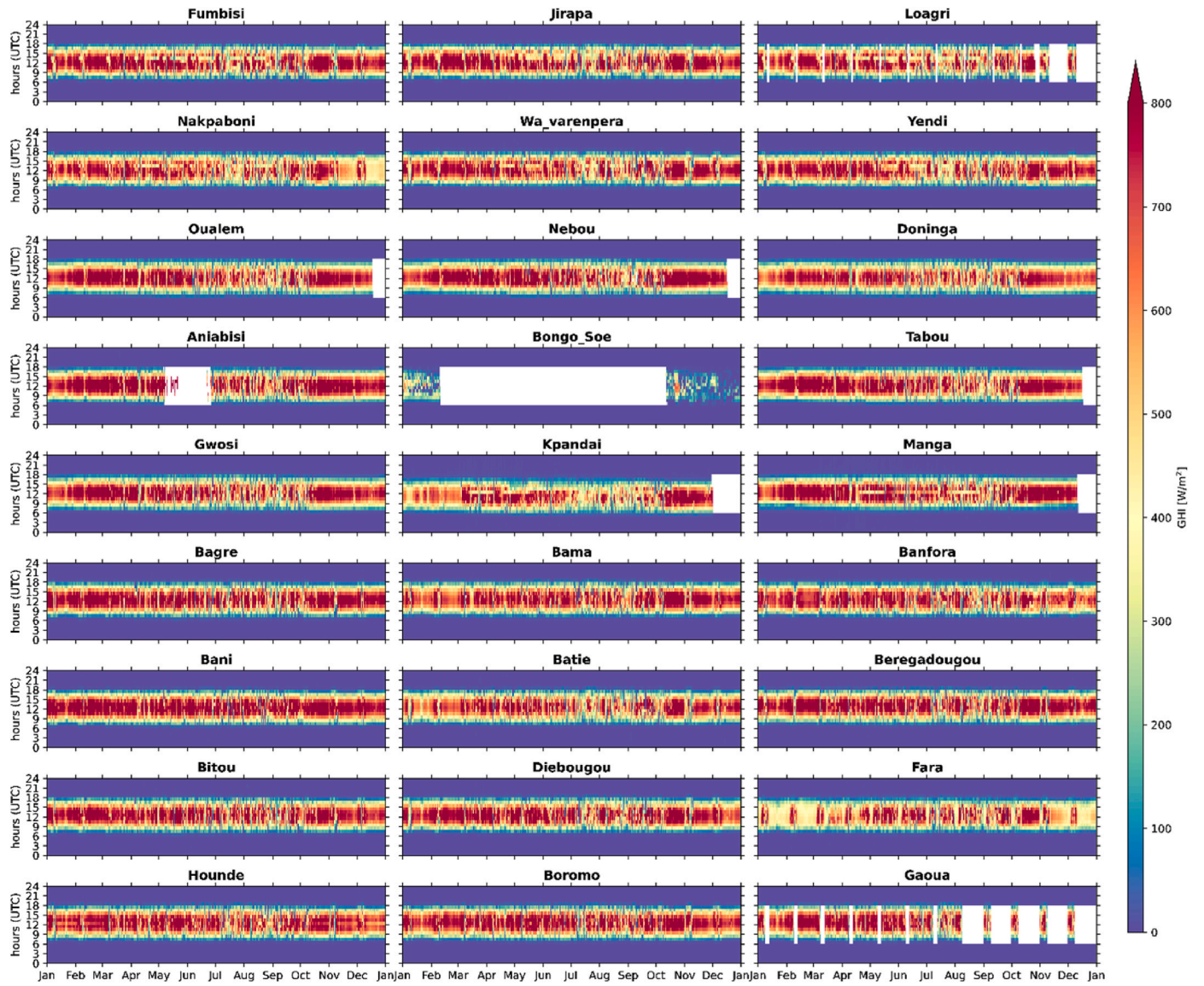


Fig. 18. Same ad Fig.17, but within the Savannah zone.

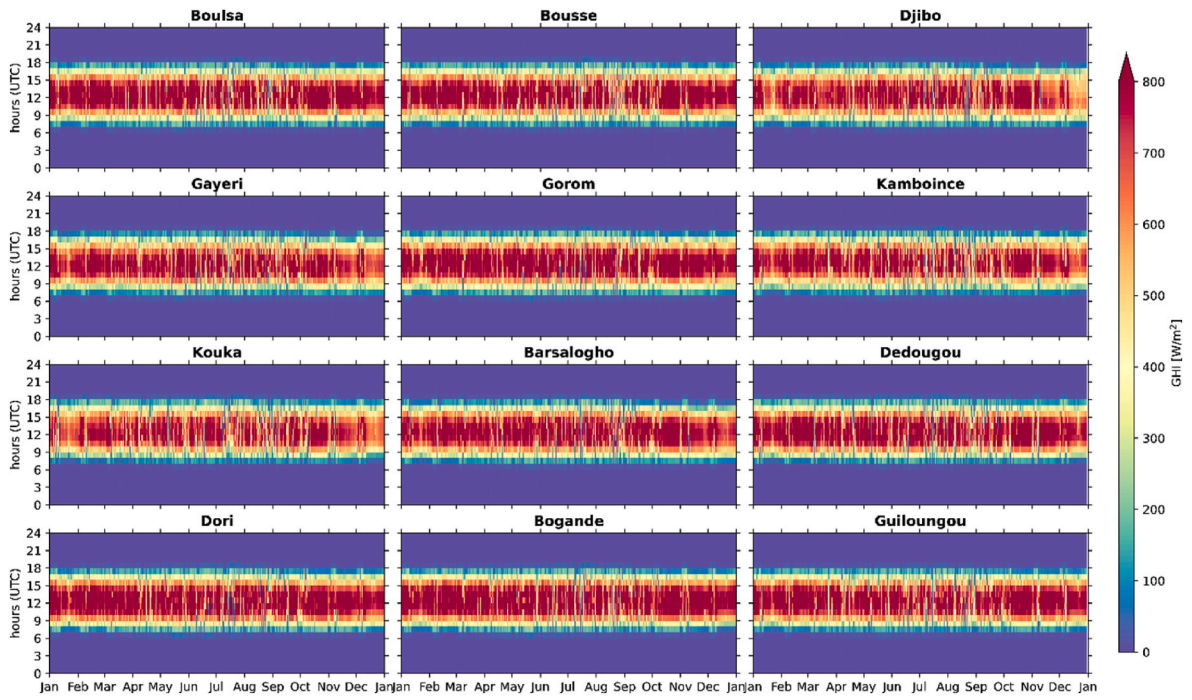


Fig. 19. Same as Fig.17, but within the Sahel zone.

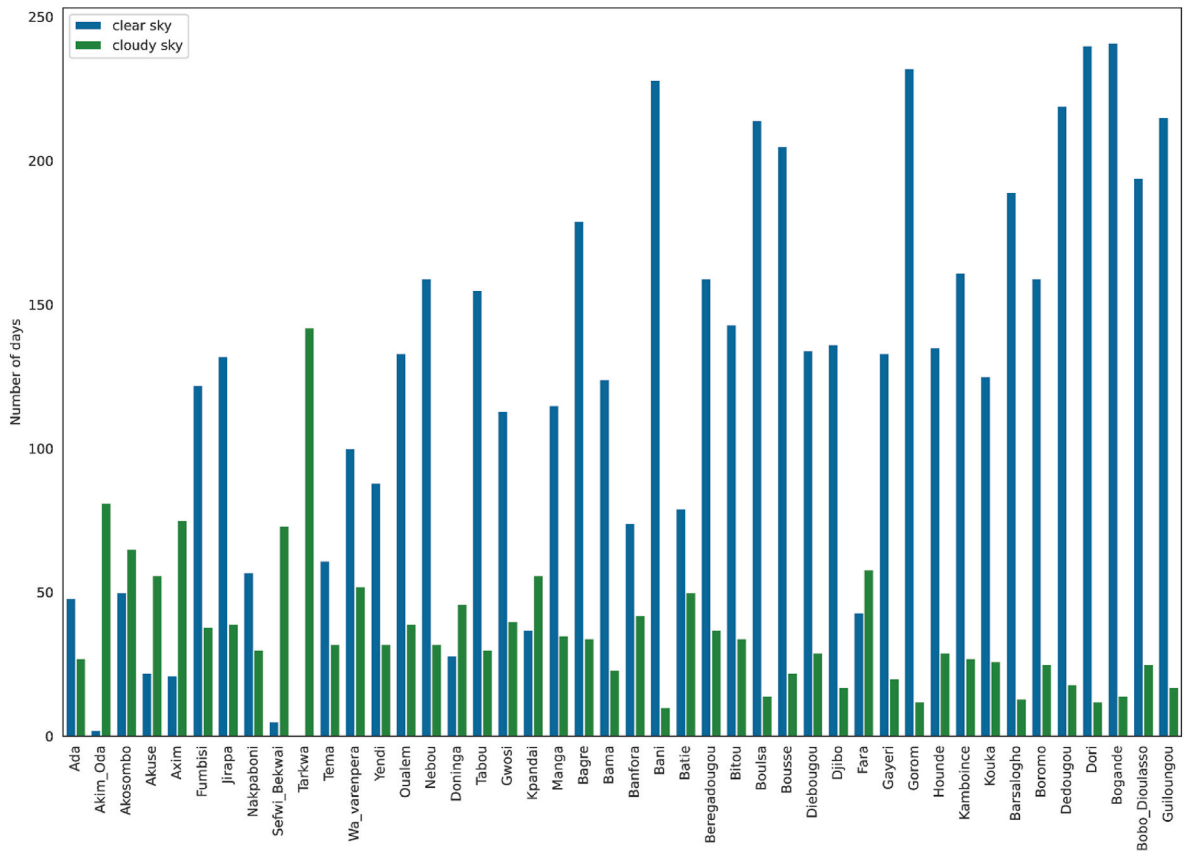


Fig. 20. Bar plot showing the number of clear-sky and cloudy-sky days for different stations.

Table.5

Number of data point that are above the physically possible limit and the extremely rare limit for different stations.

Station	Number of data outside from the BSRN range	Number of data outside from the BSRN range in percentage
Ada	2	0.05%
Akim_Oda	1	0.03%
Akosombo	1	0.03%
Akuse	11	0.29%
Axim	3	0.08%
Fumbisi	1	0.03%
Jirapa	0	0.00%
Nakpaboni	2	0.05%
Tema	10	0.26%
Wa_varenpera	0	0.00%
Yendi	2	0.05%
Oualem	134	3.47%
Nebou	146	3.79%
Doninga	0	0.00%
Tabou	151	3.91%
Gwosi	2	0.05%
Manga	179	4.64%
Bagre	1	0.03%
Bama	1	0.03%
Banfora	0	0.00%
Bani	1	0.03%
Batie	0	0.00%
Beregadougou	1	0.03%
Bitou	0	0.00%
Boulsa	0	0.00%
Bousse	0	0.00%
Diebougou	0	0.00%
Djibo	0	0.00%
Gayeri	0	0.00%
Gorom	0	0.00%
Hounde	0	0.00%
Kamboince	0	0.00%
Kouka	1	0.03%
Barsalogho	0	0.00%
Boromo	0	0.00%
Dedougou	0	0.00%
Dori	0	0.00%
Bogande	0	0.00%
Bobo_Dioulasso	2	0.05%
Guiloungou	0	0.00%
all stations	649	0.44%

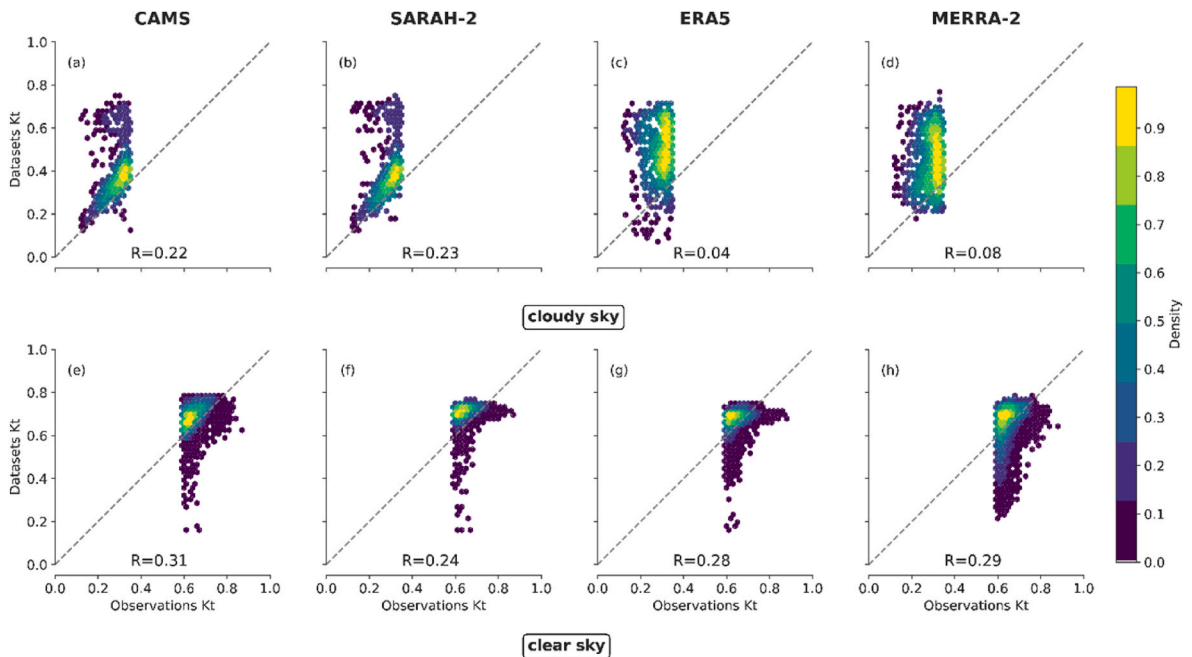


Fig. 21. Density plot of the daily clearness index (Kt) from different datasets (CAMS, ERA5, SARAH-2, and MERRA-2) against observation for 37 stations using Gaussian kernels with normalized values of 0–1 for different clear and cloudy skies. The dashed grey line shows the line: 1:1 line. R indicates the Pearson correlation.

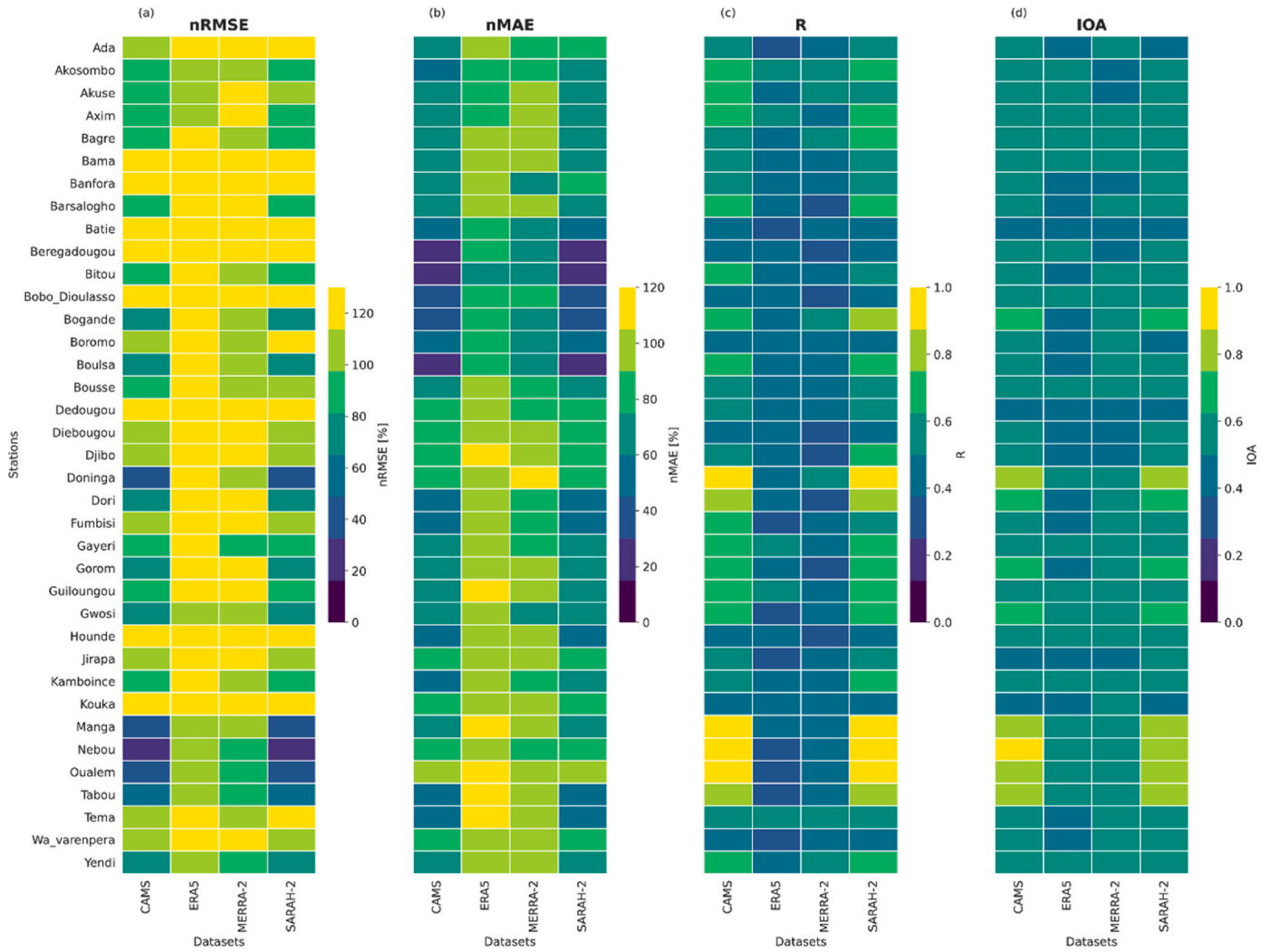


Fig. 22. Performance metrics of different datasets at different weather stations under cloudy skies. Panel (a) shows the normalized root-mean-square-error (nRMSE); panel (b) indicates the normalized mean absolute error (nMAE); panel (c) shows the correlation, and panel (d) displays the Index of Agreement (IOA).

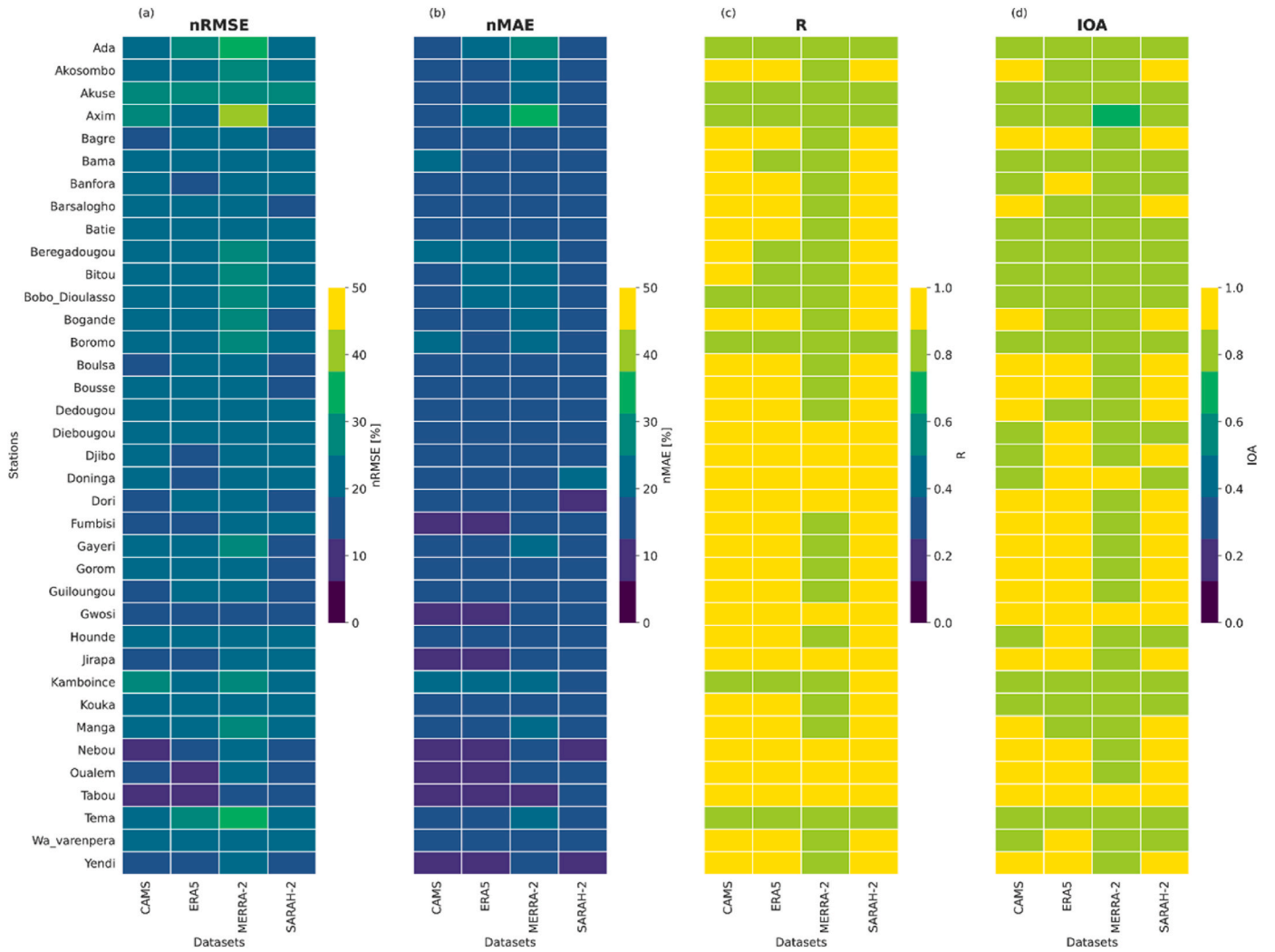


Fig. 23. Similar to Fig.22, but for clear skies.

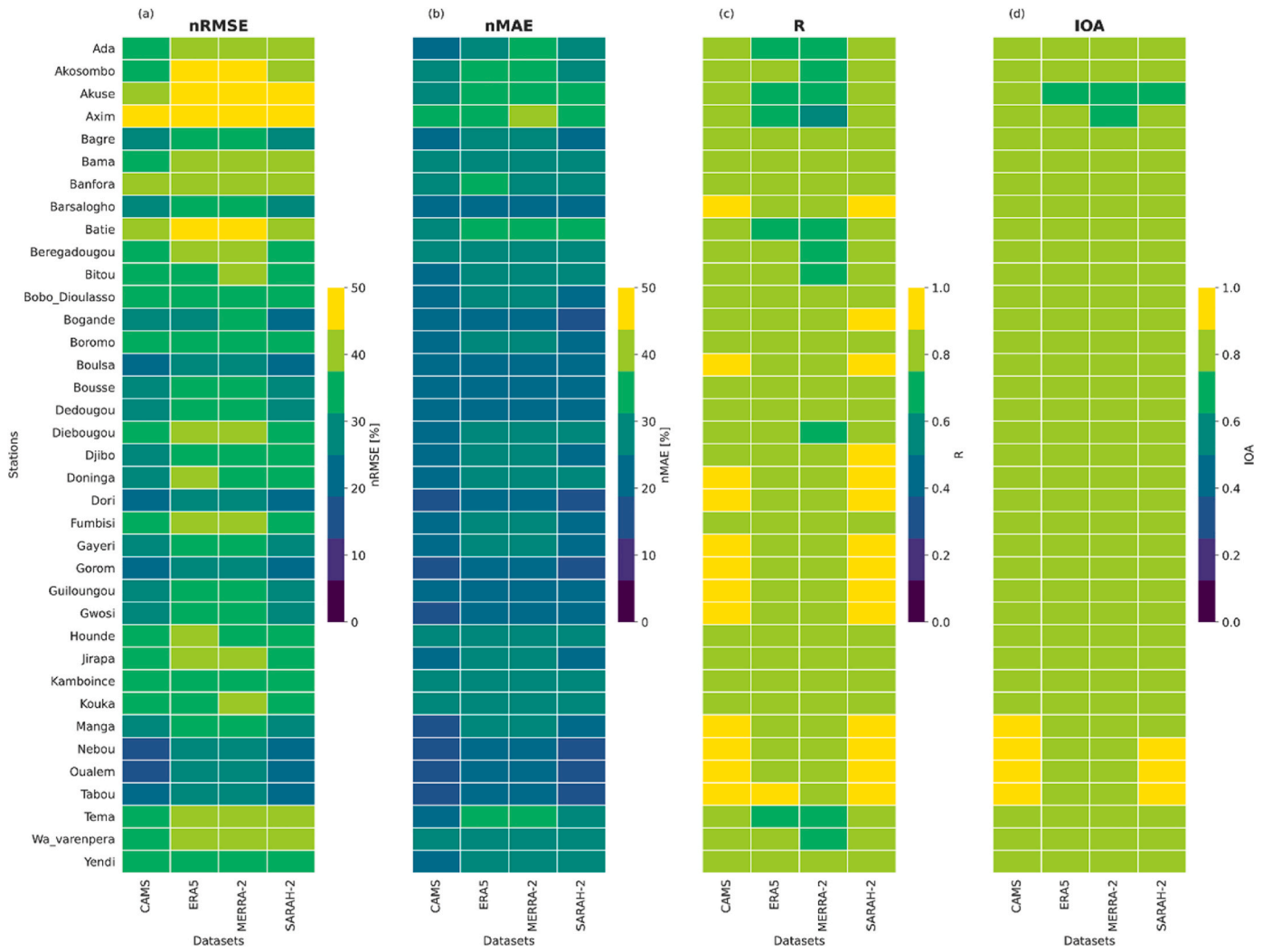


Fig. 24. Similar to Fig.22, but for all skies.

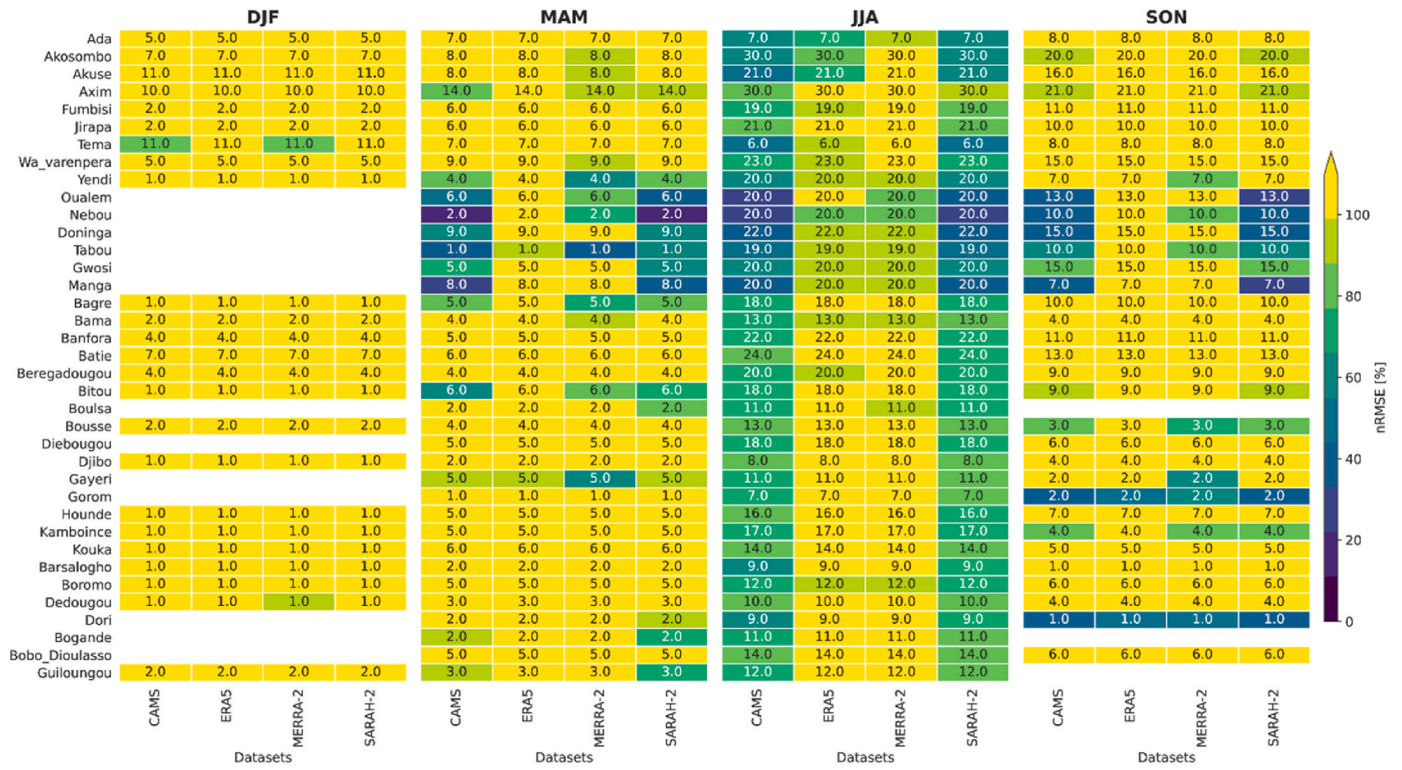


Fig. 25. Performance metrics of different datasets at different season under cloudy skies. The number in the heatmap shows the number of cloudy days that occur at a given season and station. The empty areas indicate absence of cloudy days.

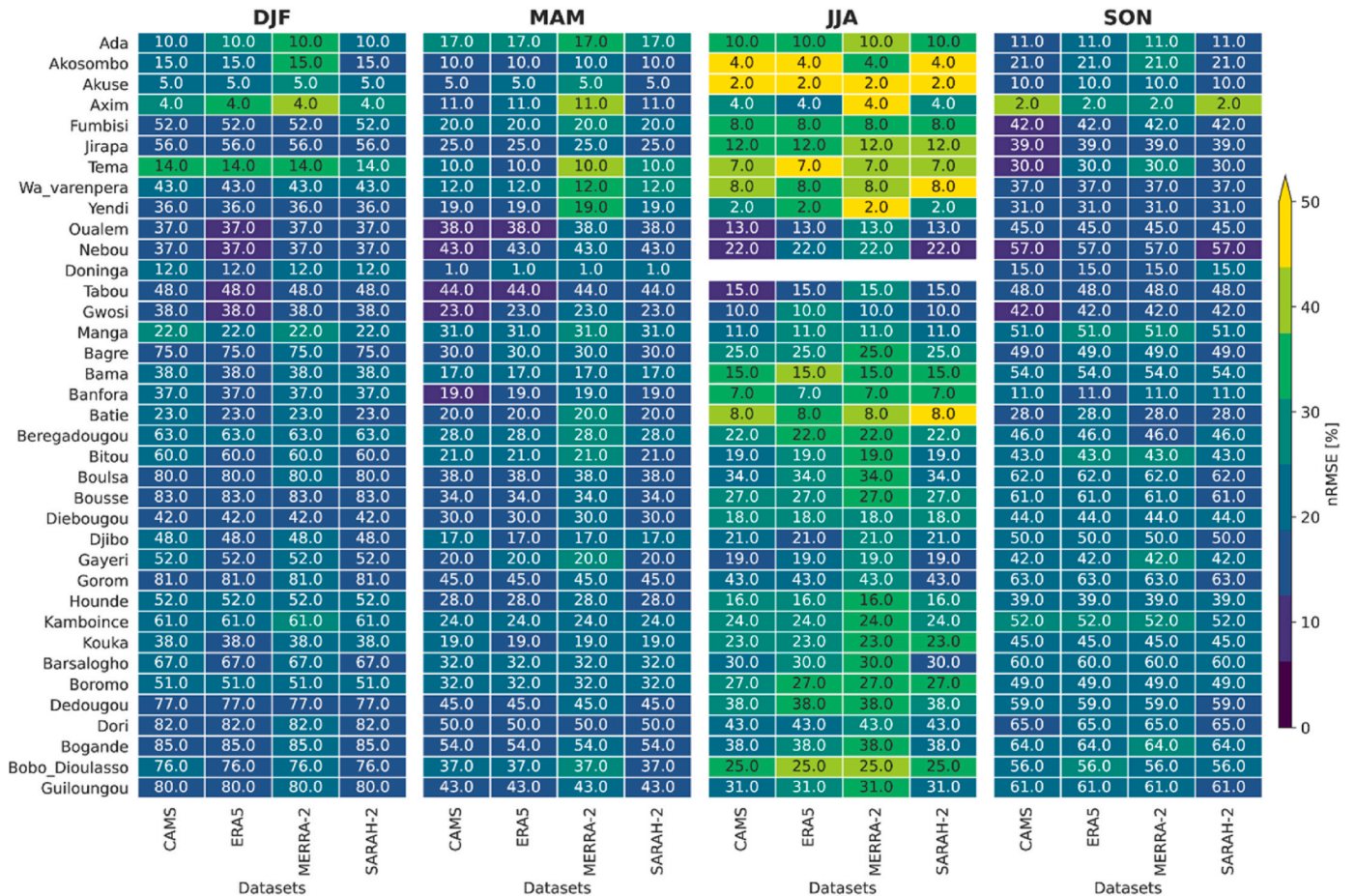


Fig. 26. Similar to Fig.25, but for clear skies.

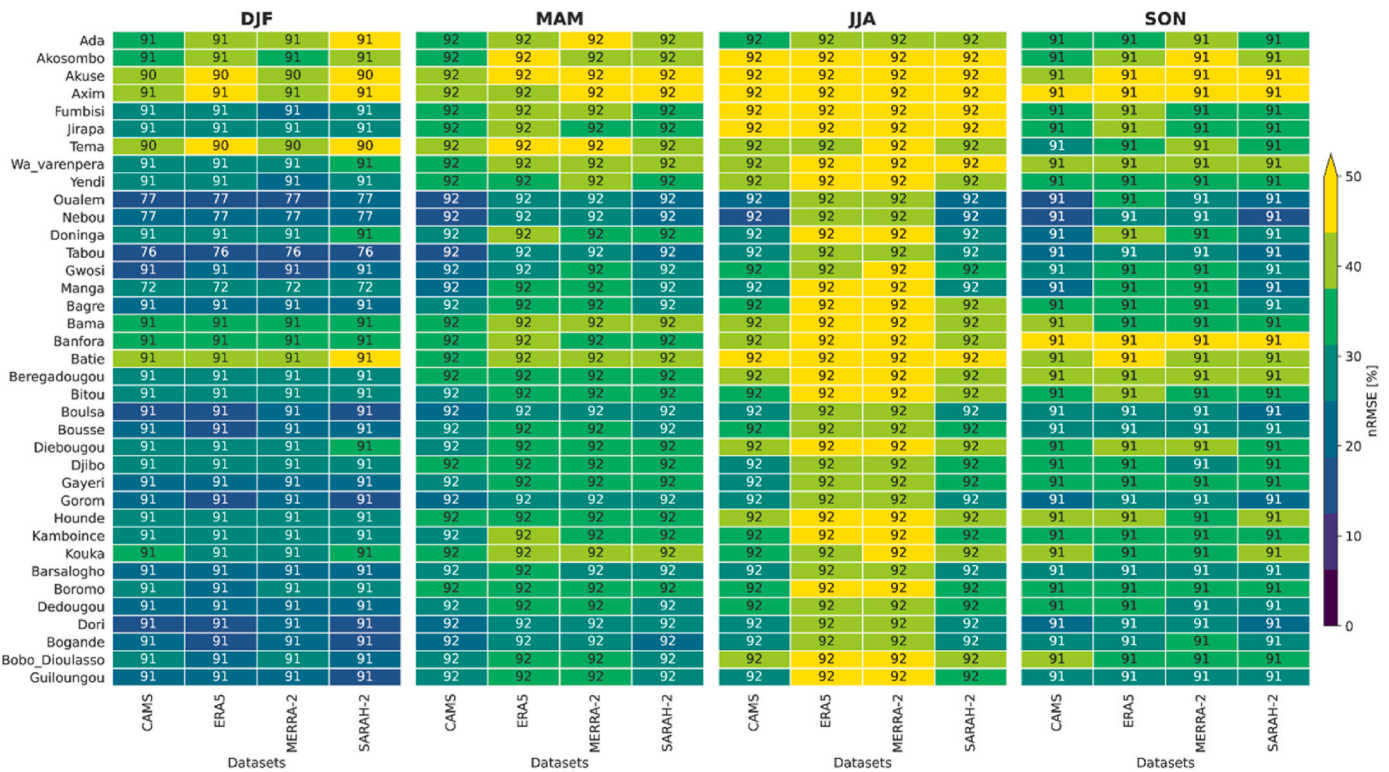


Fig. 27. Similar to Fig.25, but for all skies.

References

[1] G. Huang, Z. Li, X. Li, S. Liang, K. Yang, D. Wang, Y. Zhang, Estimating surface solar irradiance from satellites: past, present, and future perspectives, *Remote Sens. Environ.* 233 (2019), 111371, <https://doi.org/10.1016/j.rse.2019.111371>.

[2] W. Sawadogo, B.J. Abiodun, E.C. Okogbue, Impacts of global warming on photovoltaic power generation over West Africa, *Renew. Energy* 151 (2020) 263–277, <https://doi.org/10.1016/j.renene.2019.11.032>.

[3] ECREEE, Regional Progress Report on Renewable Energy, Energy Efficiency and Energy Access in ECOWAS Region Monitoring Year, 2020, 2018.

[4] B. Mabasa, M.D. Lysko, S.J. Moloi, Validating hourly satellite based and reanalysis based global horizontal irradiance datasets over South Africa, *Geomatics* 1 (4) (2021) 429–449, <https://doi.org/10.3390/geomatics1040025>.

[5] J. Bliefernicht, S. Salack, M. Waongo, T. Annor, P. Laux, H. Kunstmann, Towards a historical precipitation database for West Africa: overview, quality control and harmonization, *Int. J. Climatol.* (2021).

[6] S. Salack, A. Bossa, J. Bliefernicht, S. Berger, Y. Yira, K.A. Sanoussi, S. Guug, D. Heinzeller, A.S. Avocanh, B. Hamadou, others, Designing Transnational Hydroclimatological Observation Networks and Data Sharing Policies in West Africa, 2019.

[7] T. Dinku, Challenges with availability and quality of climate data in Africa, in: *Extreme Hydrology and Climate Variability*, Elsevier, 2019, pp. 71–80, <https://doi.org/10.1016/B978-0-12-815998-9.00007-5>.

[8] V.N. Dike, M. Addi, H.A. Andang’o, B.F. Attig, R. Barimalala, U.J. Diasso, M. Du Plessis, S. Lamine, P.N. Mongwe, M. Zaroug, others, Obstacles facing Africa’s young climate scientists, *Nat. Clim. Change* 8 (6) (2018) 447–449.

[9] United Nations Economic Commission for Africa, African climate policy Centre, United Nations. Economic Commission for Africa. (2011). An assessment of Africa’s climate observing networks and data including strategies for rescuing of climatic data, <https://hdl.handle.net/10855/21110>.

[10] C. Lorenz, H. Kunstmann, The hydrological cycle in three state-of-the-art reanalyses: intercomparison and performance analysis, *J. Hydrometeorol.* 13 (5) (2012) 1397–1420.

[11] F. Kaspar, J. Helmschrot, A. Mhanda, M. Butale, W. de Clercq, J.K. Kanyanga, F.O. S. Neto, S. Kruger, M. Castro Matsheka, G. Muche, T. Hillmann, K. Josenhans, R. Posada, J. Riede, M. Seely, C. Ribeiro, P. Kenabatho, R. Vogt, N. Jürgens, The SASSCAL contribution to climate observation, climate data management and data rescue in Southern Africa, *Adv. Sci. Res.* 12 (1) (2015) 171–177, <https://doi.org/10.5194/asr-12-171-2015>.

[12] N. van de Giesen, R. Hut, J. Selker, The trans-African hydro-meteorological observatory (TAHMO), *Wiley Interdisciplinary Reviews: Water* 1 (4) (2014) 341–348.

[13] J. Schunke, P. Laux, J. Bliefernicht, M. Waongo, W. Sawadogo, H. Kunstmann, Exploring the potential of the cost-efficient TAHMO observation data for hydro-meteorological applications in sub-saharan Africa, *Water* 13 (22) (2021) 3308, <https://doi.org/10.3390/w13223308>.

[14] J. Polo, S. Wilbert, J.A. Ruiz-Arias, R. Meyer, C. Gueymard, M. Suri, L. Mart`in, T. Mieslinger, P. Blanc, I. Grant, others, Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets, *Sol. Energy* 132 (2016) 25–37.

[15] C.A. Gueymard, S.M. Wilcox, Assessment of spatial and temporal variability in the US solar resource from radiometric measurements and predictions from models using ground-based or satellite data, *Sol. Energy* 85 (5) (2011) 1068–1084, <https://doi.org/10.1016/j.solener.2011.02.030>.

[16] W. Sawadogo, M.S. Reboita, A. Faye, R.P. da Rocha, R.C. Odoulami, C.F. Olusegun, M.O. Adeniyi, B.J. Abiodun, M.B. Sylla, I. Diallo, E. Coppola, F. Giorgi, Current and future potential of solar and wind energy over Africa using the RegCM4 CORDEX-CORE ensemble, *Clim. Dynam.* (2020), <https://doi.org/10.1007/s00382-020-05377-1>.

[17] C. Tang, B. Morel, M. Wild, B. Pohl, B. Abiodun, M. Bessafi, Numerical simulation of surface solar radiation over Southern Africa. Part 1: evaluation of regional and global climate models, *Clim. Dynam.* (2018), <https://doi.org/10.1007/s00382-018-4143-1>.

[18] C. Fant, C.A. Schlosser, K. Strzepek, The impact of climate change on wind and solar resources in southern Africa, *Appl. Energy* 161 (2016) 556–564.

[19] R. Mahtra, P.K. Joshi, A.K. Jindal, Solar power potential mapping in India using remote sensing inputs and environmental parameters, *Renew. Energy* 71 (2014) 255–262, <https://doi.org/10.1016/j.renene.2014.05.037>.

[20] D. Yang, A correct validation of the national solar radiation data base (NSRDB), *Renew. Sustain. Energy Rev.* 97 (2018) 152–155, <https://doi.org/10.1016/j.rser.2018.08.023>.

[21] D. Yang, J.M. Bright, 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* 210 (2020) 3–19, <https://doi.org/10.1016/j.solener.2020.04.016>.

[22] M. Marchand, M. Lefevre, L. Saboret, E. Wey, L. Wald, 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.* 16 (2019) 103–111, <https://doi.org/10.5194/asr-16-103-2019>.

[23] C. Thomas, E. Wey, P. Blanc, L. Wald, Validation of three satellite-derived databases of surface solar radiation using measurements performed at 42 stations in Brazil, *Adv. Sci. Res.* 13 (2016) 81–86, <https://doi.org/10.5194/asr-13-81-2016>.

- [24] Y. Du, H. Shi, J. Zhang, X. Xia, Z. Yao, D. Fu, B. Hu, C. Huang, Evaluation of MERRA-2 hourly surface solar radiation across China, *Sol. Energy* 234 (2022) 103–110, <https://doi.org/10.1016/j.solener.2022.01.066>.
- [25] M. Tall, C. Albergel, B. Bonan, Y. Zheng, F. Guichard, M. Dramé, A. Gaye, L. Sintondji, F. Hountondji, P. Nikiema, J.-C. Calvet, Towards a long-term reanalysis of land surface variables over Western Africa: LDAS-monde applied over Burkina Faso from 2001 to 2018, *Rem. Sens.* 11 (6) (2019) 735, <https://doi.org/10.3390/rs11060735>.
- [26] I. Neher, S. Crewell, S. Meilinger, U. Pfeifroth, J. Trentmann, Photovoltaic power potential in West Africa using long-term satellite data, *Atmos. Chem. Phys.* 20 (21) (2020) 12871–12888, <https://doi.org/10.5194/acp-20-12871-2020>.
- [27] T. Khatib, W. Elmenreich, A model for hourly solar radiation data generation from daily solar radiation data using a generalized regression artificial neural network, *Int. J. Photoenergy* (2015) 1–13, <https://doi.org/10.1155/2015/968024>, 2015.
- [28] S.E. Nicholson, A.H. Fink, C. Funk, Assessing recovery and change in West Africa's rainfall regime from a 161-year record, *Int. J. Climatol.* 38 (10) (2018) 3770–3786.
- [29] J. Raj, H.K. Bangalath, G. Stenichkov, West African Monsoon: current state and future projections in a high-resolution AGCM, *Clim. Dynam.* 52 (11) (2019) 6441–6461, <https://doi.org/10.1007/s00382-018-4522-7>.
- [30] A. Sunnu, G. Afeti, F. Resch, A long-term experimental study of the Saharan dust presence in West Africa, *Atmos. Res.* 87 (1) (2008) 13–26, <https://doi.org/10.1016/j.atmosres.2007.07.004>.
- [31] B.J. Abiodun, Z.D. Adeyewa, P.G. Oguntunde, A.T. Salami, V.O. Ajayi, Modeling the impacts of reforestation on future climate in West Africa, *Theor. Appl. Climatol.* 110 (1–2) (2012) 77–96.
- [32] D.J. Parker, M. Diop-Kane, *Meteorology of Tropical West Africa: the Forecasters' Handbook*, John Wiley & Sons, 2017.
- [33] R. Schuster, A.H. Fink, P. Knippertz, Formation and maintenance of nocturnal low-level stratus over the southern West African monsoon region during AMMA 2006, *J. Atmos. Sci.* 70 (8) (2013) 2337–2355.
- [34] J. Bliefernicht, S. Berger, S. Salack, S. Guug, L. Hingerl, D. Heinzeller, M. Mauder, R. Steinbrecher, G. Steup, A.Y. Bossa, others, The WASCAL hydrometeorological observatory in the Sudan Savanna of Burkina Faso and Ghana, *Vadose Zone J.* 17 (1) (2018).
- [35] U. Pfeifroth, A. Sanchez-Lorenzo, V. Manara, J. Trentmann, R. Hollmann, Trends and variability of surface solar radiation in Europe based on surface- and satellite-based data records, *J. Geophys. Res. Atmos.* 123 (3) (2018) 1735–1754.
- [36] R. Posselt, R.W. Mueller, R. Stöckli, J. Trentmann, Remote sensing of solar surface radiation for climate monitoring — the CM-SAF retrieval in international comparison, *Remote Sens. Environ.* 118 (2012) 186–198, <https://doi.org/10.1016/j.rse.2011.11.016>.
- [37] B. Mayer, A. Kylling, Technical note: the libRadtran software package for radiative transfer calculations - description and examples of use, *Atmos. Chem. Phys.* 5 (7) (2005) 1855–1877, <https://doi.org/10.5194/acp-5-1855-2005>.
- [38] A. Amillo, T. Huld, R. Müller, A new database of global and direct solar radiation using the eastern meteosat satellite, models and validation, *Rem. Sens.* 6 (9) (2014) 8165–8189, <https://doi.org/10.3390/rs6098165>.
- [39] Z. Qu, A. Oumbe, P. Blanc, B. Espinar, G. Gesell, B. Gschwind, L. Klüser, M. Lefèvre, L. Saboret, M. Schroedter-Homscheidt, L. Wald, Fast radiative transfer parameterisation for assessing the surface solar irradiance: the Heliosat-4 method, *Meteorol. Z.* 26 (1) (2017) 33–57, <https://doi.org/10.1127/metz/2016/0781>.
- [40] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J.W. Kaiser, J.-J. Morcrette, McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions, *Atmos. Meas. Tech.* 6 (9) (2013) 2403–2418, <https://doi.org/10.5194/amt-6-2403-2013>.
- [41] A. Oumbe, Z. Qu, P. Blanc, M. Lefèvre, L. Wald, S. Cros, Decoupling the effects of clear atmosphere and clouds to simplify calculations of the broadband solar irradiance at ground level, *Geosci. Model Dev. (GMD)* 7 (4) (2014) 1661–1669, <https://doi.org/10.5194/gmd-7-1661-2014>.
- [42] V. Buchard, C.A. Randles, A.M. da Silva, A. Darmanov, P.R. Colarco, R. Govindaraju, R. Ferrare, J. Hair, A.J. Beyersdorf, L.D. Ziemba, H. Yu, The MERRA-2 aerosol reanalysis, 1980 onward. Part II: evaluation and case studies, *J. Clim.* (2017), <https://doi.org/10.1175/JCLI-D-16-0613.1>.
- [43] P. Colarco, A. Da Silva, M. Chin, T. Diehl, Online simulations of global aerosol distributions in the NASA GEOS-4 model and comparisons to satellite and ground-based aerosol optical depth, *J. Geophys. Res. Atmos.* (2010), <https://doi.org/10.1029/2009JD012820>.
- [44] M. Chin, P. Ginoux, S. Kinne, O. Torres, B.N. Holben, B.N. Duncan, R.V. Martin, J. A. Logan, A. Higurashi, T. Nakajima, Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and sun photometer measurements, *J. Atmos. Sci.* (2002), [https://doi.org/10.1175/1520-0469\(2002\)059<0461:taotft>2.0.co;2](https://doi.org/10.1175/1520-0469(2002)059<0461:taotft>2.0.co;2).
- [45] M.-D. Chou, M.J. Suarez, A solar radiation parameterization (CLIRAD-SW) for atmospheric studies, *NASA Tech. Memo.* 10460 (1999) 48.
- [46] M.-D. Chou, M.J. Suarez, X.-Z. Liang, M.M.-H. Yan, C. Cote, A Thermal Infrared Radiation Parameterization for Atmospheric Studies, 2001.
- [47] A.K. Heidinger, M.J. Foster, A. Walther, X. Zhao, Tom), The pathfinder atmospheres-extended AVHRR climate dataset, *Bull. Am. Meteorol. Soc.* 95 (6) (2014) 909–922, <https://doi.org/10.1175/BAMS-D-12-00246.1>.
- [48] H. Hersbach, B. Bell, P. Berrisford, A. Horányi, J.M. Sabater, J. Nicolas, R. Radu, D. Schepers, A. Simmons, C. Soci, D. Dee, Global Reanalysis: Goodbye ERA-Interim, Hello ERA5, *ECMWF Newsletter*, 2019.
- [49] H. Hersbach, C. Peubey, A. Simmons, P. Berrisford, P. Poli, D. Dee, ERA-20CM: a twentieth-century atmospheric model ensemble, *Q. J. R. Meteorol. Soc.* 141 (691) (2015) 2350–2375, <https://doi.org/10.1002/qj.2528>.
- [50] H. Liu, A global view of aerosols from merged transport models, satellite, and ground observations, *J. Geophys. Res.* 110 (D10) (2005) D10S15, <https://doi.org/10.1029/2004JD004695>.
- [51] Y. Chen, F. Weng, Y. Han, Q. Liu, Validation of the community radiative transfer model by using CloudSat data, *J. Geophys. Res.* 113 (2008) D00A03, <https://doi.org/10.1029/2007JD009561>.
- [52] H.D. Kambezidis, B.E. Psiloglou, Climatology of the linke and unsworth-monteith turbidity parameters for Greece: introduction to the notion of a typical atmospheric turbidity year, *Appl. Sci.* 10 (11) (2020) 4043, <https://doi.org/10.3390/app10114043>.
- [53] R. Perez, P. Ineichen, R. Seals, A. Zelenka, Making full use of the clearness index for parameterizing hourly insolation conditions, *Sol. Energy* 45 (2) (1990) 111–114, [https://doi.org/10.1016/0038-092X\(90\)90036-C](https://doi.org/10.1016/0038-092X(90)90036-C).
- [54] H.D. Kambezidis, S.I. Kampeziou, D. Kampeziou, Mathematical determination of the upper and lower limits of the diffuse fraction at any site, *Appl. Sci.* 11 (18) (2021) 8654, <https://doi.org/10.3390/app11188654>.
- [55] O.O. Soneye, Evaluation of Clearness Index and Cloudiness Index Using Measured Global Solar Radiation Data: A Case Study for a Tropical Climatic Region of Nigeria, 2020, <https://doi.org/10.20937/ATM.52796>. *Atmósfera*.
- [56] E.C. Okogbue, J.A. Adedokun, B. Holmgren, Hourly and daily clearness index and diffuse fraction at a tropical station, Ile-Ife, Nigeria, *Int. J. Climatol.: A Journal of the Royal Meteorological Society* 29 (8) (2009) 1035–1047.
- [57] A. Kuye, S.S. Jagtap, Analysis of solar radiation data for Port Harcourt, Nigeria, *Sol. Energy* 49 (2) (1992) 139–145.
- [58] BSRN, Baseline Surface Radiation Network, 2021. <https://bsrn.awi.de/>.
- [59] H. Li, Y. Lian, X. Wang, W. Ma, L. Zhao, Solar constant values for estimating solar radiation, *Energy* 36 (3) (2011) 1785–1789, <https://doi.org/10.1016/j.energy.2010.12.050>.
- [60] I. Neher, T. Buchmann, S. Crewell, B. Evers-Dietze, K. Pfeilsticker, B. Pospichal, C. Schirmer, S. Meilinger, Impact of atmospheric aerosols on photovoltaic energy production Scenario for the Sahel zone, *Energy Proc.* 125 (2017) 170–179, <https://doi.org/10.1016/j.egypro.2017.08.168>.
- [61] S. Younes, R. Claywell, T. Muneer, Quality control of solar radiation data: present status and proposed new approaches, *Energy* 30 (9) (2005) 1533–1549.
- [62] C.J. Willmott, On the validation of models, *Phys. Geogr.* 2 (2) (1981) 184–194.
- [63] K.E. Taylor, Summarizing multiple aspects of model performance in a single diagram, *J. Geophys. Res. Atmos.* 106 (D7) (2001) 7183–7192, <https://doi.org/10.1029/2000JD900719>.
- [64] L.M. Hinkelman, The global radiative energy budget in MERRA and MERRA-2: evaluation with respect to CERES EBAF data, *J. Clim.* 32 (6) (2019) 1973–1994, <https://doi.org/10.1175/JCLI-D-18-0445.1>.
- [65] J.-J. Morcrette, H.W. Barker, J.N.S. Cole, M.J. Iacono, R. Pincus, Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system, *Mon. Weather Rev.* 136 (12) (2008) 4773–4798, <https://doi.org/10.1175/2008MWR2363.1>.
- [66] G. Salazar, C. Gueymard, J.B. Galdino, O. de Castro Vilela, N. Fraidenraich, Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil, *Renew. Sustain. Energy Rev.* 117 (2020), 109478, <https://doi.org/10.1016/j.rser.2019.109478>.
- [67] B. Babar, R. Graversen, T. Boström, Solar radiation estimation at high latitudes: assessment of the CMSAF databases, ASR and ERA5, *Sol. Energy* 182 (2019) 397–411, <https://doi.org/10.1016/j.solener.2019.02.058>.
- [68] W. Schwanhart, B. Schütt, Meteorological causes of harmattan dust in West Africa, *Geomorphology* 95 (3–4) (2008) 412–428, <https://doi.org/10.1016/j.geomorph.2007.07.002>.
- [69] S. Chander, A. Purohit, A. Sharma, Arvind, S.P. Nehra, M.S. Dhaka, A study on photovoltaic parameters of mono-crystalline silicon solar cell with cell temperature, *Energy Rep.* 1 (2015) 104–109, <https://doi.org/10.1016/J.EGYR.2015.03.004>.
- [70] V. Ramanathan, P.J. Crutzen, J.T. Kiehl, D. Rosenfeld, Aerosols, climate, and the hydrological cycle, *Science* 294 (5549) (2001) 2119–2124, <https://doi.org/10.1126/science.1064034>.
- [71] P.G. Kosmopoulos, S. Kazadzis, M. Taylor, E. Athanasopoulou, O. Speyer, P. I. Raptis, E. Marinou, E. Proestakis, S. Solomos, E. Gerasopoulos, V. Amiridis, A. Bais, C. Kontoes, Dust impact on surface solar irradiance assessed with model simulations, satellite observations and ground-based measurements, *Atmos. Meas. Tech.* 10 (7) (2017) 2435–2453, <https://doi.org/10.5194/amt-10-2435-2017>.
- [72] R. Urraca, T. Huld, A. Gracia-Amillo, F.J. Martinez-de-Pison, F. Kaspar, A. Sanz-García, Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data, *Sol. Energy* 164 (2018) 339–354, <https://doi.org/10.1016/j.solener.2018.02.059>.
- [73] O. Dubovik, G.L. Schuster, F. Xu, Y. Hu, H. Bösch, J. Landgraf, Z. Li, Grand challenges in satellite remote sensing, *Frontiers in Remote Sensing* 2 (2021), <https://doi.org/10.3389/frsen.2021.619818>.
- [74] X. Zhang, N. Lu, H. Jiang, L. Yao, Evaluation of reanalysis surface incident solar radiation data in China, *Sci. Rep.* 10 (1) (2020) 3494, <https://doi.org/10.1038/s41598-020-60460-1>.
- [75] Y. Sianturi, Marjuki, K. Sartika, Evaluation of ERA5 and MERRA2 Reanalyses to Estimate Solar Irradiance Using Ground Observations over Indonesia Region, 2020, 020002, <https://doi.org/10.1063/5.0000854>.
- [76] L. Hannak, P. Knippertz, A.H. Fink, A. Kniffka, G. Pante, Why do global climate models struggle to represent low-level clouds in the West African summer monsoon? *J. Clim.* 30 (5) (2017) 1665–1687.
- [77] P. Knippertz, A.H. Fink, R. Schuster, J. Trentmann, J.M. Schrage, C. Yorke, Ultra-low clouds over the southern West African monsoon region, *Geophys. Res. Lett.* 38 (21) (2011).

- [78] K. Deetz, H. Vogel, P. Knippertz, B. Adler, J. Taylor, H. Coe, K. Bower, S. Haslett, M. Flynn, J. Dorsey, others, Numerical simulations of aerosol radiative effects and their impact on clouds and atmospheric dynamics over southern West Africa, *Atmos. Chem. Phys.* 18 (13) (2018) 9767–9788.
- [79] P.A. Jiménez, J. Yang, J.-H. Kim, M. Sengupta, J. Dudhia, Assessing the WRF-Solar model performance using satellite-derived irradiance from the National Solar Radiation Database, *J. Appl. Meteorol. Climatol.* 61 (2) (2022) 129–142.
- [80] S. Incecik, S. Sakarya, S. Tilev, A. Kahraman, B. Aksoy, E. Caliskan, S. Topcu, C. Kahya, M.T. Odman, Evaluation of WRF parameterizations for global horizontal irradiation forecasts: a study for Turkey, *Atmósfera* 32 (2) (2019) 143–158.
- [81] M.-M. Zempila, T.M. Giannaros, A. Bais, D. Melas, A. Kazantzidis, Evaluation of WRF shortwave radiation parameterizations in predicting Global Horizontal Irradiance in Greece, *Renew. Energy* 86 (2015) 831–840.