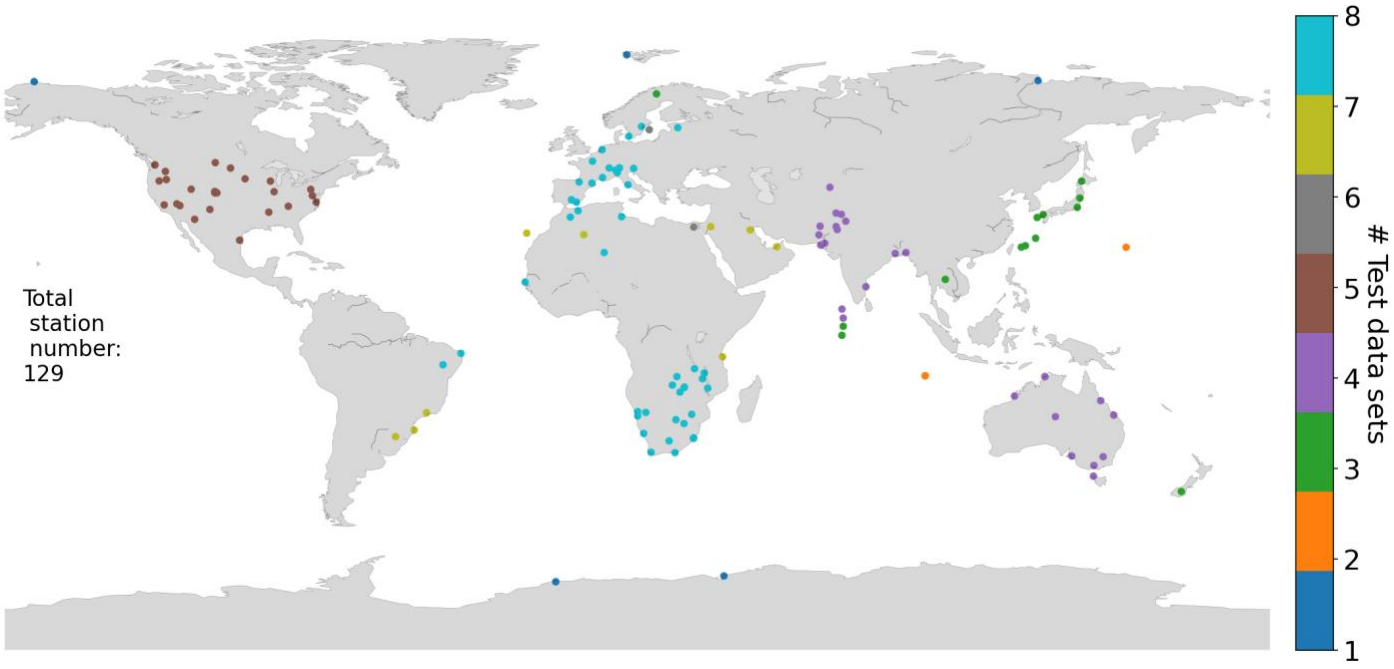
International Energy Agency
Photovoltaic Power Systems Programme



Task 16 Solar Resource for High Penetration and Large Scale Applications

PVPS

Worldwide Benchmark of Modelled Solar Irradiance Data

2023

Report IEA-PVPS T16-05: 2023

What is IEA PVPS TCP?

The International Energy Agency (IEA), founded in 1974, is an autonomous body within the framework of the Organization for Economic Cooperation and Development (OECD). The Technology Collaboration Programme (TCP) was created with a belief that the future of energy security and sustainability starts with global collaboration. The programme is made up of 6000 experts across government, academia, and industry dedicated to advancing common research and the application of specific energy technologies.

The IEA Photovoltaic Power Systems Programme (IEA PVPS) is one of the TCPs within the IEA and was established in 1993. The mission of the programme is to "enhance the international collaborative efforts which facilitate the role of photovoltaic solar energy as a cornerstone in the transition to sustainable energy systems." In order to achieve this, the Programme's participants have undertaken a variety of joint research projects in PV power systems applications. The overall programme is headed by an Executive Committee, comprised of one delegate from each country or organisation member, which designates distinct 'Tasks,' that may be research projects or activity areas.

The IEA PVPS participating countries are Australia, Austria, Belgium, Canada, Chile, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Malaysia, Mexico, Morocco, the Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, and the United States of America. The European Commission, Solar Power Europe, the Smart Electric Power Alliance (SEPA), the Solar Energy Industries Association and the Copper Alliance are also members.

Visit us at: www.iea-pvps.org

What is IEA PVPS Task 16?

The objective of Task 16 of the IEA Photovoltaic Power Systems Programme is to lower barriers and costs of grid integration of PV and lowering planning and investment costs for PV by enhancing the quality of the resource assessments and solar forecasts.

Authors

- **Main Content:** Anne Forstinger (CSPS), Stefan Wilbert (DLR), Adam R. Jensen (DTU), Birk Kraas (CSPS), Carlos Fernández Peruchena (CENER), Christian A. Gueymard (Solar Consulting Services), Dario Ronzio (RSE), Dazhi Yang (Harbin Institute of Technology), Elena Collino (RSE), Jesús Polo Martinez (CIEMAT), Jose A. Ruiz-Arias (Uni Malaga), Natalie Hanrieder (DLR), Philippe Blanc (MINES ParisTech), Yves-Marie Saint-Drenan (MINES ParisTech)

- **Editor:** Anne Forstinger

COVER PICTURE

Position of reference radiometric station and number of test data sets per station. © CSP Services.

ISBN 978-3-907281-44-4: Solar Resource for High Penetration and Large-Scale Applications 2023

INTERNATIONAL ENERGY AGENCY

PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

Worldwide Benchmark of Modelled Solar Irradiance Data

IEA PVPS
Task 16
Solar Resource for High Penetration and Large-Scale Applications

Report IEA-PVPS T16-05:2023
June - 2023

ISBN 978-3-907281-44-44

TABLE OF CONTENTS

Acknowledgements 6
Executive summary 7
1 Introduction 9
2 Test and reference datasets 10
2.1 Reference database 10
2.1 Test data sets 12
3 Evaluation method 17
4 Quality control and data selection 19
4.1 QC methodology for 1-minute data 19
4.2 Data selection 26
5 Benchmark Results 27
5.1 Scatter density plots 27
5.2 World maps 29
5.3 Results overview per continent 32
6 Conclusions and summary 35
Description of the data file annex 36
References 37

ACKNOWLEDGEMENTS

The authors would like to thank NamPower in Namibia, Yuldash Solirov from the Institute of Material Science in Uzbekistan, Frank Vignola from the University of Oregon, David Pozo from the University of Jaén, Dietmar Baumgartner from the University of Graz, Julian Gröbner at PMOD, Nicolas Fernay from the University of Lille, Peter Armstrong at the Masdar Institute, Laurent Vuilleumier at MeteoSwiss, Irena Balog at ENEA, Sophie Pelland at CanmetÉNERGIE Varennes, Etienne Guillot at CNRS-PROMES Odeillo, and Majed Al-Rasheedi at the Kuwait Institute for Scientific Research for the provision of data for this study. Furthermore, we gratefully thank the Department of Civil and Mechanical Engineering at the Technical University of Denmark, the Swedish Meteorological and Hydrological Institute, the Australian Government Bureau of Meteorology, the INPE National Institute of Space Research, the CCST Center for Earth System Sciences, along with FINEP Financier of Studies and Projects Ministry of Science and Technology, PETROBRAS Petróleo Brasileiro, the SKYNET organization (in particular Hitoshi Irie and Tamio Takamura (CEReS/Chiba-U.), Chiba University, Tadahiro Hayasaka (Tohoku University), and Chulalongkorn University), as well as the ESMAP program of the World Bank Group (in particular Joana Zerbin, Clara Ivanescu, Branislav Schnierer, Roman Affolter, GeoSUN Africa, Rachel Fox, and Margot King), the NOAA Global Monitoring Laboratory and the BSRN (in particular for stations 1, 3, 4, 6, 7, 8, 9, 10, 11, 13, 17, 20, 21, 31, 32, 33, 34, 35, 36, 37, 40, 42, 45, 47, 48, 49, 53, 56, 57, 58, 59, 60, 61, 63, 65, 70, 71, 72, 74). We would also like to thank the German Federal Foreign Office for funding and coordinating the enerMENA project, and to express our deep gratitude to the other project partners for their efforts in measuring the meteo-solar data and for agreeing to share their data. These partners are the Cairo University in Egypt, the University of Oujda and Institute Research Solar Energy et Energies Nouvelles (IRESEN) in Morocco, the Research and Technology Centre of Energy (CRTEn) in Tunisia, the University of Jordan in Jordan, and the Centre de Developpement des Energies Renouvelables (CDER) in Algeria.

CSPS and DLR thank the German Ministry for Economic Affairs and Climate Action for funding their contribution to the study within the SOLREV project (contract number 03EE1010). Adam R. Jensen thanks the Danish Energy Agency for funding his participation (grant number: 64019-0512). Part of this work has been financed by Research Fund for the Italian Electrical System with the Decree of 16 April 2018.

The authors thank Solargis, Meteotest, DWD, NREL, KNMI, BoM, and CAMS for sharing their data for the evaluation within this benchmark as well as their comments and remarks regarding the evaluation.

EXECUTIVE SUMMARY

Modelled irradiance data based on satellite products and numerical weather prediction models are frequently used in solar energy applications and atmospheric sciences. Many such sources of data are now offered by many different institutional or commercial providers, but it is currently difficult for users to independently identify the best provider for their specific application and location. This work presents a benchmark of model-derived direct normal irradiance (DNI) as well as global horizontal irradiance (GHI) data at the sites of 129 globally distributed ground-based radiation measurement stations. DNI and GHI estimates from ten different solar radiation datasets, either public-domain or commercial, are compared against high-quality ground-based irradiance observations from these stations. The comparison of the modelled to observed data is conducted at hourly temporal resolution. The performance of the modelled data is analysed with respect to different regions and climate zones. This study is intended to help the solar industry make better informed decisions about solar resource assessments.

The reference observational database consisting of ground measurements, is collected from 25 different providers or radiometric networks. Most stations provide measurements of DNI, GHI, and diffuse horizontal irradiance (DIF) with thermopile radiometers and a solar tracker. A few stations provide measurements of only two independent components, with either two thermopile radiometers or a single rotating shadowband irradiometer (RSI). The used reference database is at high temporal resolution (1 min) from 129 stations during 2015–2020. Only quality-assured data have been considered in this benchmark through a comprehensive set of best practices and newly implemented quality-control procedures. These include automatic as well as manual data quality-control tests carried out by a team of experts for all stations and result in flags describing the quality for each time stamp. The 129 stations are spread out worldwide, including 31 stations in Africa, 31 in Asia, 27 in North America, 20 in Europe, 13 in Australia, 5 in South America, and 2 in Antarctica. The bulk of the quality-controlled data from the 129 stations has been published within this benchmark including the results of the quality control.

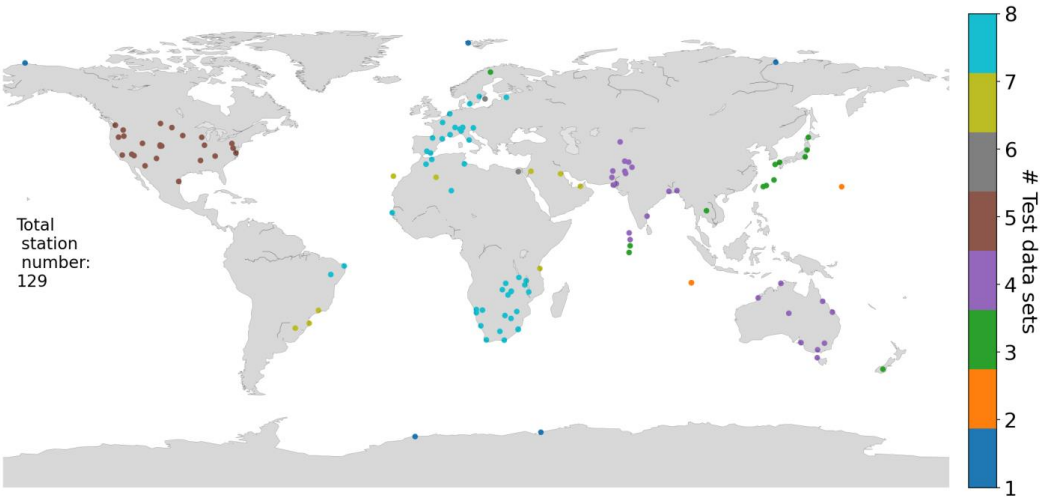


Figure 1: Station map of reference stations for the benchmark. Each point represents one station and its color corresponds to the number of modelled data sets that are tested at that site.

The modelled data sets, which are tested by comparing them to ground-based reference measurements in this benchmark, are called test data sets. They stem from ten different models from nine different providers. Not all models provide estimates for all stations, as Figure 1 shows.

Amongst other statistical performance parameters, the mean bias deviation, root mean square deviation, and standard deviation are calculated for each year and for all stations. The results for the relative mean bias deviation affecting GHI are shown in Figure 2.

Based on the results of the statistical analysis, the most appropriate data set might depend on site, climate, or continent of interest. The model errors and the differences between the various modelled data sets are much higher for DNI than for GHI.

Based on this work, analysts can make an informed decision about which surface radiation model(s) and data provider(s) are most suited for their location and application.

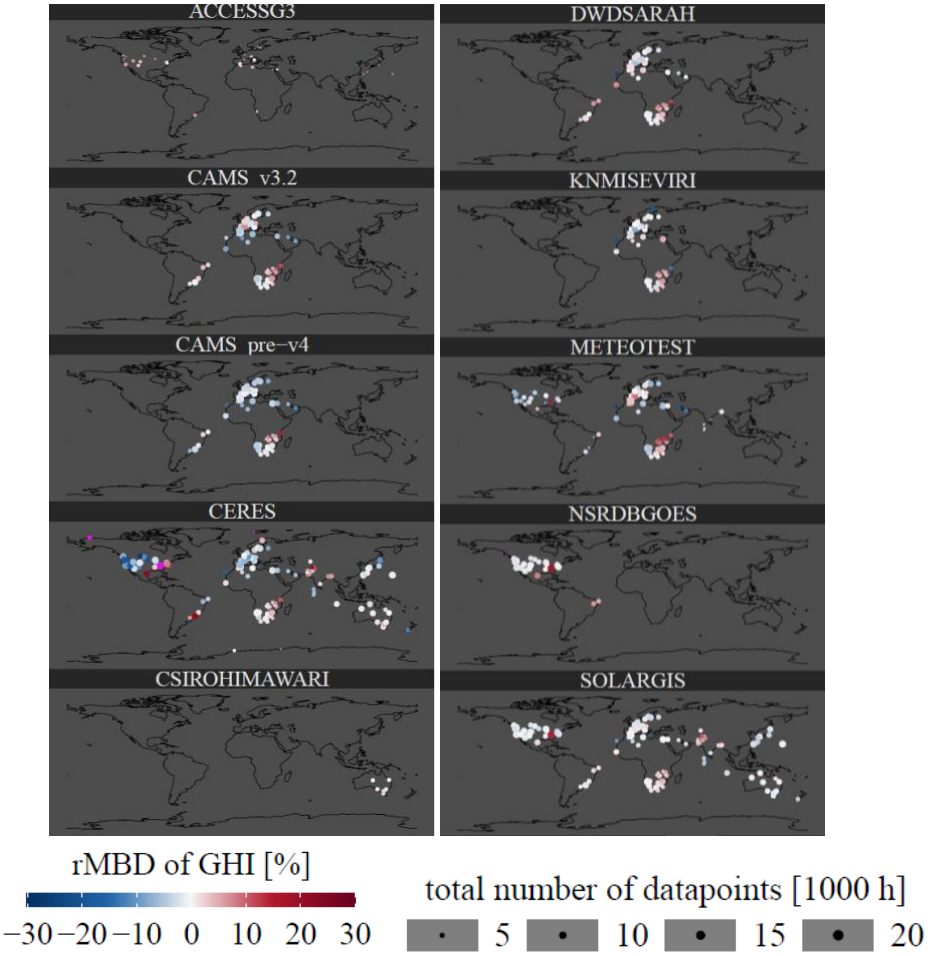


Figure 2: Relative mean bias deviation for GHI and all stations and years. Magenta color indicates results out of the color bar range. The point size corresponds to the total number of datapoints in the tested time series from 2015 to 2020.

1 INTRODUCTION

Modelled solar irradiance data, based on satellite products and numerical weather prediction (NWP) models, are frequently used in solar energy applications and atmospheric sciences. This kind of data is offered by several institutional or commercial providers, and currently it is not practically feasible for users to independently identify the best provider for their specific application and location. This work presents a benchmark of model-derived direct normal irradiance (DNI) as well as global horizontal irradiance (GHI) data at the sites of 129 globally distributed ground-based radiation measurement stations. DNI and GHI estimates from 10 different solar radiation databases, either commercial or public-domain, are compared against these stations' high-quality ground-based irradiance observations. The comparison of the original model data delivered by the data providers is conducted at hourly temporal resolution, even though some of these data sets are available at a finer temporal resolution. The performance of the data is analysed with respect to different regions and climate zones. This study is intended to help the solar industry make better informed decisions about solar resource assessments and solar potential studies.

The reference observational database, consisting of ground measurements, is collected from 25 different providers or radiometric networks. Most stations provide measurements of DNI, GHI, and diffuse horizontal irradiance (DIF) with thermopile radiometers and a solar tracker. A few stations provide measurements of only two independent components, either with two thermopile radiometers or a single rotating shadowband irradiometer (RSI). The reference database is at high temporal resolution (1 min) from 129 stations during 2015–2020. Only quality-assured data have been considered in this benchmark through a comprehensive set of best practices and newly implemented quality-control procedures (Forstinger et al., 2021). These include both automatic and manual data quality-control tests, as well as descriptive quality flagging, as carried out by a team of experts from this Task. The 129 stations are spread out worldwide with data from all continents. These stations were selected from an initial pool of 161 stations that were submitted to the initial quality-control process. The quality control process, as well as other practical or technical considerations, resulted in the elimination of 32 stations. The solar irradiance modelled datasets stem from ten models from nine different providers at the 129 stations considered in the final reference dataset. Not all models provide data for all stations, because of limitations in the geographical coverage of the satellite on which they depend. Both publicly available datasets and commercial data sets are included in the present benchmark. This multi-model multi-site study constitutes a considerably enhanced effort in comparison with the earlier benchmark that was conducted under the auspices of previous IEA Tasks (Ineichen 2014; Šúri et al. 2008), and various investigations of the literature (e.g., (Amillo et al. 2018; Marchand et al. 2018; Salazar et al. 2020)).

This report is structured as follows. Section 2 presents the test and reference data sets used for this benchmark. Section 3 describes the evaluation method and related performance metrics. The quality-control procedure and data selection are explained in Section 4. Finally, Section 5 presents the results of the benchmark, followed by a summary and outlook in Section 6. The report is accompanied by a data Annex with visualization and tables of the results and reference station information. These files are described in the Annex at the end of the report.

2 TEST AND REFERENCE DATASETS

The first part of this section describes the reference data acquired with ground-based radiometers from 129 stations. As mentioned above, the test data are compared to the reference data from the ground stations. The 10 test data sets of modelled irradiance time series that are analysed in the benchmark are introduced in Subsection 2.2.

2.1 Reference database

Figure 3: Example of a Tier-1 (left) and Tier-2 (right) station.

The reference data used in this benchmark originates from 129 ground stations distributed worldwide, as provided by 25 distinct sources. These include large national or international networks, as well as some private stations that have not been used yet to test any modelled data.

Most stations provide measurements of DNI, GHI, and DIF, obtained with three thermopile radiometers and a solar tracker (example shown in Figure 3, left). Such stations are called “Tier-1 stations” in this work. A few stations provide measurements of only two independent components, either with two thermopile radiometers or a single RSI (Figure 3, right). Such stations are called “Tier-2 stations”.

The reference database is at high temporal resolution (1 min) from 129 stations and spans the period 2015–2020. Initially, 161 ground stations were quality-controlled to determine their applicability for the benchmark (Figure 4). The quality control process is explained in Section 4. The final set of 129 stations was selected based on the quality and data availability in the evaluation period (2015–2020). The selection of the evaluation years 2015–2020 was done based on the data availability of both modelled and quality-checked reference data.

Figure 5 (left) shows the providers of the selected reference stations. The database is partly obtained from the Southern African Universities Radiometric Network (Brooks et al. 2015), the National Renewable Energy Laboratory (Andreas and Stoffel 1981; Andreas and Wilcox 2010; 2012; Andreas and Stoffel 2006; Vignola and Andreas 2013; Ramos and Andreas 2011), the Baseline Surface Radiation Network (BSRN; Driemel et al. 2018; Gueymard et al. 2022) and further sources. As often in similar radiation data benchmarks, BSRN contributes a significant part of the stations. However, nearly 75% of the final pool of stations is *not* part of BSRN, which increases the novelty and relevance of this benchmark. Another important contributor is the ESMAP network, which has rarely been used for this kind of validation studies. About 25% of the stations were so far not in the public domain, and some are still only available to CSPS and their clients. These new stations are thus of particular value for the benchmark because they were unseen to any developer of modelled irradiance data.

The bulk of the ground measurement data set (122 stations out of the 161 quality-controlled stations used for the analysis), including the quality-control flags derived per Section 4, has been made publicly available by the benchmark evaluator team (Forstinger, et al. 2021: link). The data providers of these 122 stations, which all provide all three components, are also shown in Figure 5. Note, that the selected stations for the benchmark (Figure 5 (left)) and the published stations (Figure 5 (right)) are chosen from the 161 quality-controlled stations. The number of stations therefore varies between the two groups. In some cases, stations that had not been used for the benchmark were made publicly available.

Figure 6 shows all the 129 measurement stations that have finally been selected as reference stations for the benchmark. The color scale indicates the number of test data sets per reference station. Stations that were used by one or two providers for any kind of post-processing prior to the benchmark are marked with crosses.

A list of all stations including their coordinates, climate zone, station code, continent, altitude above mean sea level (AMSL), data source, number of available test data sets, tier level, and availability of calibration records, is included in the data Annex to this report (StationList.xlsx).

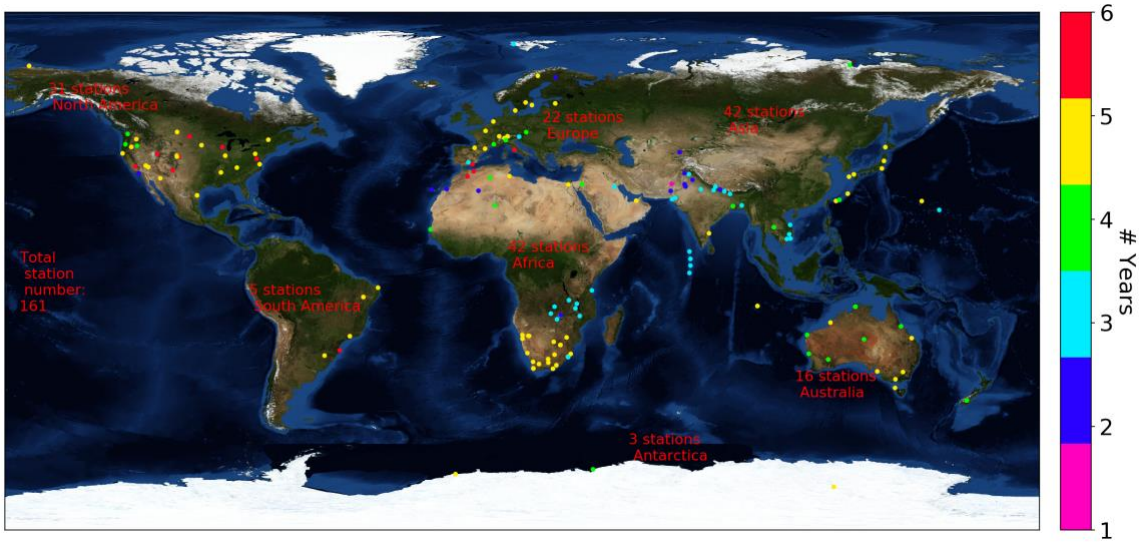


Figure 4: Location and number of years per reference station. In total, 686 station-calendar years of the original pool of 161 different stations were quality-controlled.

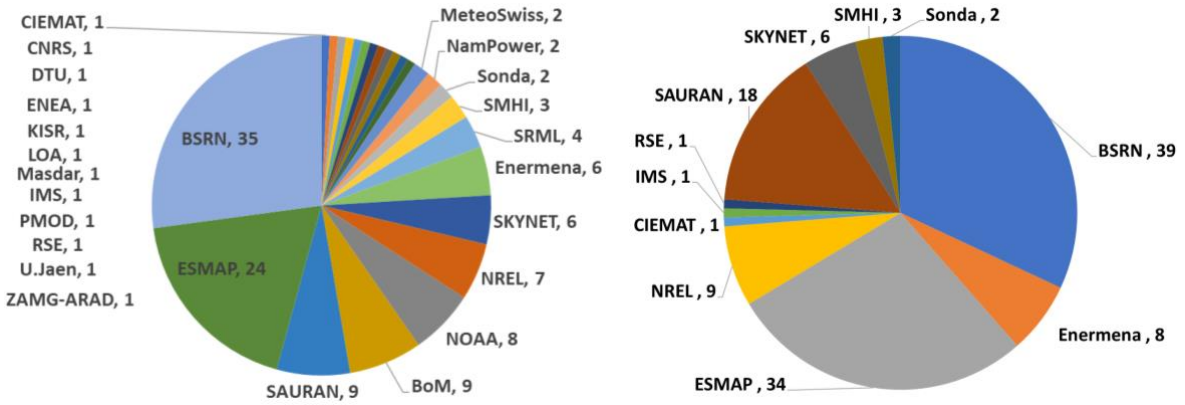


Figure 5: Source of the 129 selected reference data sets for the benchmark (left) and the 122 published data sets with quality control flags (right) (Forstinger, et al. 2021: link). Note that the number of stations per provider is different in the two data sets.

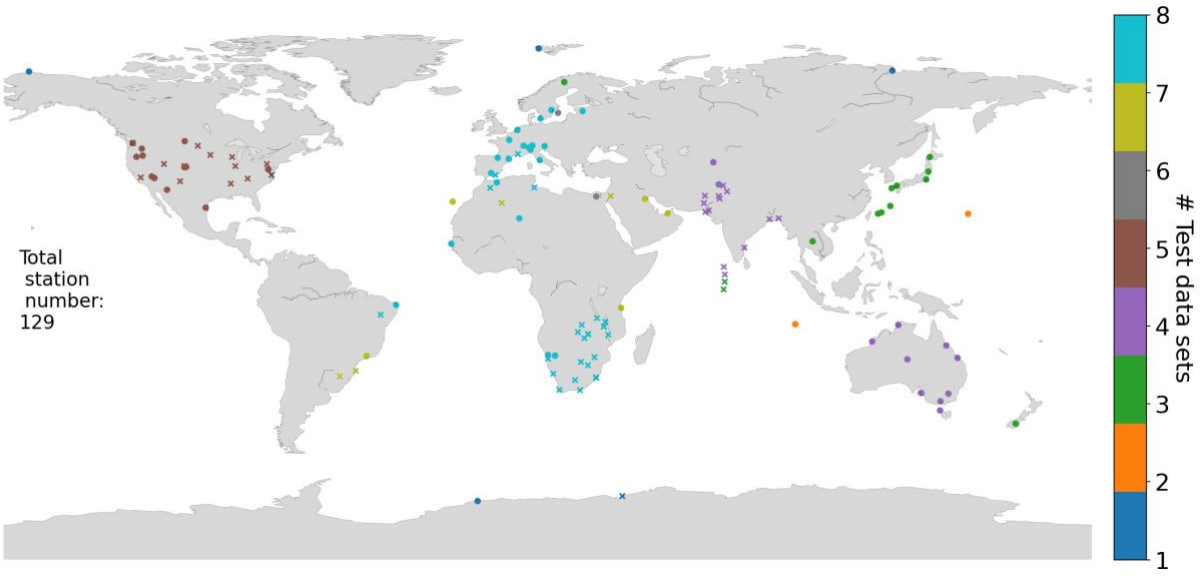


Figure 6: Location and number of test data sets per reference station. Stations that were used by one or two providers for post-processing are marked with crosses.

2.1 Test data sets

Currently, there are many regional and global, public or private, modelled solar irradiance data sets, and new ones are created on a regular basis. For more details, the reader is referred to the extensive review of solar resource databases in another report of this Task (Sengupta et al. 2021).

Ten data sets are evaluated in the present benchmark. This represents a good sample of what is available today for solar resource assessment, verification of solar forecasts, or other applications. Not all commercial data providers accepted to participate in this study, however.

All data sets, with their data provider, main data sources, and spatial and temporal coverage, are described in Table 1. The 10 data sets originate from 9 different data providers because the Copernicus Atmosphere Monitoring Service (CAMS) contributed two different versions of

its CAMS radiation database. The surface irradiance in these data sets is modelled mainly from geostationary satellite images, with however two exceptions: (i) One pure NWP data set that provides global coverage (ACCESS G3, Australian Community Climate and Earth-System Simulator) and (ii) a global data set that is mainly based on imagery from polar satellites (CERES, Clouds and the Earth's Radiant Energy System). Multiple test data sets use Meteosat Second Generation (MSG) satellites as main data source.

Some modelled data sets use imagery from more than one satellite to reach global coverage, whereas other data sets only evaluate a part of the satellite field of view. The geostationary satellites and their field of view are shown in Figure 7. The areas close to the poles are not covered by most data sets, which is primarily caused by the poor viewing angle of geostationary satellites at these latitudes, but are covered by NWP and polar-orbiter-based models. Table 2 provides further details on the main data sources and methods, as well as on the spatial and temporal resolutions. The table also provides the different resolutions of the input data sets of the models for some data sets (e.g., CAMS). The test data providers kept the responsibility to properly use their data for the creation of the 60-min averages that are evaluated in this benchmark.

Table 1: Overview of the properties of the test data sets. Note that the spatial and/or temporal coverages might have been extended since the submission of the test data and that the now available version might have been updated since the submission.

Provider	Dataset or model	Main data source	Spatial coverage	Temporal coverage	Availability
DWD	SARAH-2.1	MSG satellites	Full disk MSG	Since 1983	Gridded data, 30min: freely available link
CAMS	CAMS v3.2	MSG satellites	Europe / Africa / Middle East / Atlantic Ocean (MSG field of view, -66°N to 66°N) (clear-sky data available globally)	Since 2004	Freely available
	CAMS pre-v4				
Meteotest	Meteotest, various sat.	GOES-16, MSG-4, IODC, HIMAWARI-8, Meteotest NWP model MOS	Global (-66°N to 66°N)	MSG since 2005; other sat: since 2018	Commercially available
CSIRO	CSIRO	Himawari-8	Australian continent	Since Jul. 2016	Freely available
NREL (NSRDB)	Physical Solar Model Version 3	GOES	GOES: covering longitudes between 25°W to the east and 175°W to the west as well as latitudes between 21°S to the south and 60°N to the north (i.e., contiguous United States, part of Alaska, southern Canada, Central America, and part of South America.	GOES: 1998–2019	Freely available
Solargis	Solargis v2.x	Various satellites	Global (60°N to 45°/55°S), land area and adjacent sea and oceans. regions between 60–65°N on request.	Since 1994 for Europe and Africa; since 1999 for central Asia and America except of >50° S (2018 there); since 2007 for other regions.	Commercially available
BoM	BoM APS3 ACCESS-G3	NWP	Global	Since Jul. 2019	Freely available
NASA	CERES SYN1deg	Various satellites	Global	Since 2000	Freely available

KNMI	MSG-CPP algorithm v1	MSG satellites	Full-disk Meteosat	Since 2015	Freely available

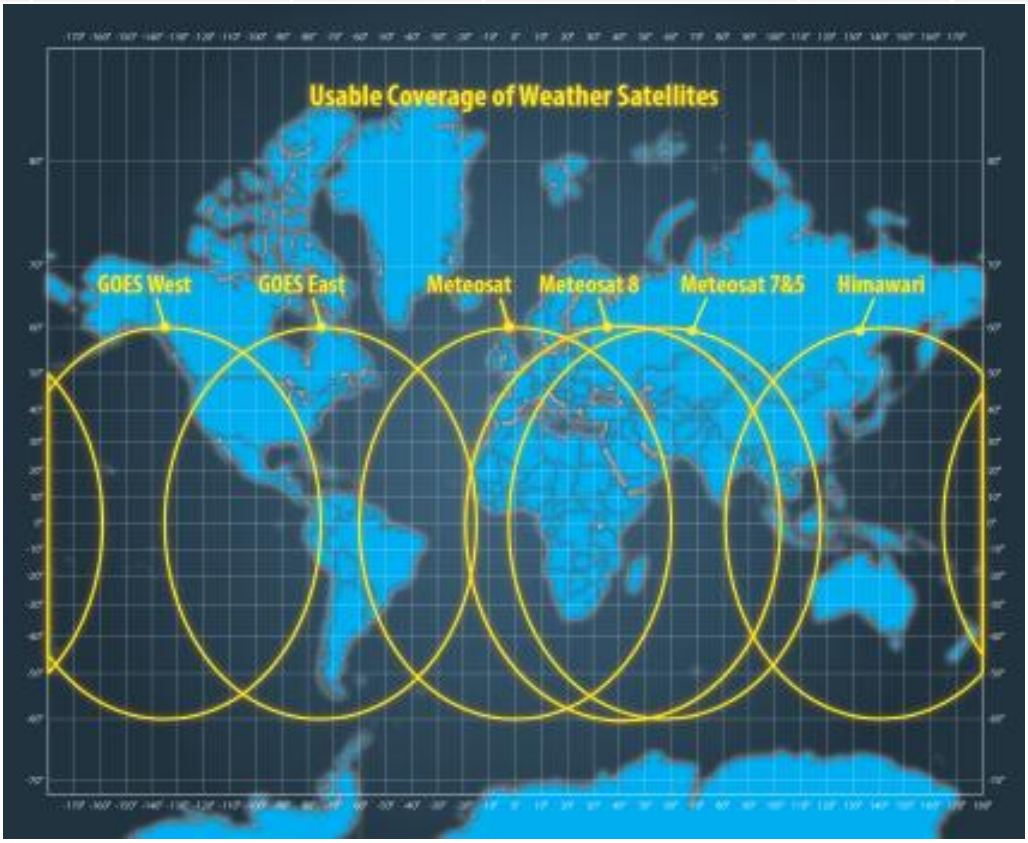


Figure 7: Location of the current geostationary satellites that provide coverage around the globe. Meteosat corresponds to the coverage of the Meteosat Prime satellities, Meteosat 8 and Meteosat 7&5 correspond to two slightly different coverages of the Meteosat IODC (Indian Ocean Data Coverage) satellites, depending on period. Image from NREL.

All data sets, except ACCESS G3, include direct irradiance estimates. In the case of the CERES data set, the direct horizontal irradiance is provided rather than DNI. The conversion from direct horizontal irradiance to DNI at hourly resolution constitutes a source of error because of the variation in solar zenith angle (SZA) during 1-hour time intervals. In the present case, DNI is derived by dividing direct horizontal irradiance by the cosine of the zenith angle at the center of the hour (e.g., 12:30 for an irradiance average corresponding to the hour from 12:00 to 13:00). SZA is obtained at each instant from a sun position algorithm. For this study, the SZA provided with the reference data is used.

To better understand the possible effects of location-specific or regional model post-processing techniques that are often used to improve surface irradiance estimates, the data providers were specifically asked about their possible reliance on such methods. Consideration for this issue certainly plays a non-negligible role in any validation study. The CAMS v3.2 data used a field-of-view-wide bias correction. Similarly, CSIRO applied a continental-wide spatial calibration for its data set, which only covers Australia. Solargis used 23 stations for regional improvement of the model. Meteotest applied data from 34 stations for post-processing using

interpolation. The list of stations that have been used by any of these providers has been considered for the data analysis and are marked in Figure 6.

The selection of the time intervals and evaluated sites is discussed in Section 4.

Table 2: Properties of the test data sets. Note that various resolutions are used for the individual input data sets that are utilized to derive the irradiance and that information on these resolutions is only provided in some cases.

Provider	Model/main data sources	Spatial resolution	Temporal resolution
DWD	SARAH-2.1, Meteosat satellites, MVIRI + SEVIRI, doi: 10.5676/EUM_SAF_CM/SARAH/V002_01	0.05° gridded sat. data (~5.5 km)	1 min (based on 30-min satellite data), 30 min, daily, monthly
CAMS	CAMS v3.2 and experimental pre-v4 APOLLO_NG/Heliosat-4(DLR) method, MSG satellites for clouds, clear-sky from CAMS integrated forecasting system	Output interpolated to location of ground station, input data at various resolutions: 3–10 km (sat. pixel), DTM up to ~100 m; aerosol, water vapor, ozone: 0.4°; ground albedo: 6 km	output: 1 min, 15 min, 60 min, 1 d, monthly; input: 15-min clouds, 3-h aerosols/water vapour/ozone, monthly ground albedo
Meteotest	Meteotest MOS GOES-16, MSG-4, IODC, HIMAWARI-8	1/16° (~7 km)	15 min
CSIRO	Himawari-8	2 km	max 10 min
NREL (NSRDB)	Model: Physical Solar Model Version 3 GOES	1998–2019, gridded segments (4- km), and for 2018 and 2019, 2-km spatial resolution	1998–2019: 30 min; 2018–2019: 5 min for continental US and 10–15 min full disk.
Solargis	Solargis model v2.x; GOES, Meteosat MSG and MFG (PRIME and IODC positions), Himawari and MTSAT satellites; Aerosols from CAMS atmospheric model	Final result 250 m, satellite data 2–4 km	10 and 15 min depending on satellite, 1 and 5 min on request. 15-min data used for benchmark
BoM	BoM APS3 ACCESS-G3	~12 km	1 h (23-07-2019 to 2020)
NASA	CERES. MODIS on Terra & Aqua (polar sat.) + geostationary sat. (GOES, Meteosat, MTSAT, Himawari)	1°x1° (111 km)	1 h
KNMI	MSG-CPP algorithm v1. Input: MSG sat. (SEVIRI data: all channels except HRV); multi-year mean climatologies of water vapor, ozone, aerosol (ECMWF/CAMS), surface albedo (MODIS)	full disk, satellite pixel size (~3 km)	15 min

3 EVALUATION METHOD

The present evaluation compares the modelled test data from each data set and reference station to the corresponding data points that were determined to be valid according to the quality-control procedure (explained in Section 4). Various metrics are used to characterize the deviations, as explained below. The time resolution of this comparison is 1 hour. The evaluation method was proposed and discussed in detail by the evaluation team and further revised based on the input of the participants of PVPS Task 16.

The evaluation metrics used in this benchmark are summarized in Table 3. The metrics are calculated for GHI and DNI at each station using the reference data (r) and the modelled estimates (s). Mean values are noted as μ. The total number of valid data points at each station is noted as N. Individual data points are noted with the subscript "i", which varies between 1 and N. The metrics are either expressed in irradiance unit (W/m^2) or as a relative value in percent, as indicated by an "r" prefix. To distinguish the metrics in irradiance units better from the relative deviations, the former are also referred to as "absolute" metrics, indicated by an "a" prefix. Note that, here, the term *absolute* does not refer to the distance to zero (the absolute value or modulus), but to the units. The formulas provided in this section mainly stem from (Gueymard 2014).

Table 3: Metrics for prediction error evaluation.

Mean bias deviation (aMBD)	$$\mathrm{aMBD} = \frac{1}{N}\sum_i (s_i - r_i)$$
Mean bias deviation relative to mean value of reference data (rMBD)	$$\mathrm{rMBD} = \frac{100}{\mu_r}\frac{1}{N}\sum_i (s_i - r_i)$$
Root mean square deviation (aRMSD)	$$\mathrm{aRMSD} = \sqrt{\frac{\sum_i (s_i - r_i)^2}{N}}$$
Root mean square deviation relative to mean value of reference data (rRMSD)	$$\mathrm{rRMSD} = \frac{100}{\mu_r}\cdot\sqrt{\frac{\sum_i (s_i - r_i)^2}{N}}$$
Standard deviation (Stddev)	$$\mathrm{stddev} = \sqrt{\frac{\sum_i (s_i - r_i - \mathrm{mean}(s_i - r_i))^2}{N}}$$
Mean absolute deviation (aMAD)	$$\mathrm{aMAD} = \frac{1}{N}\sum_i \lvert s_i - r_i\rvert$$
Mean absolute deviation relative to the mean value of reference data (rMAD)	$$\mathrm{rMAD} = \frac{100}{\mu_r}\frac{1}{N}\sum_i \lvert s_i - r_i\rvert$$
Kolmogoroff-Smirnoff Index (rKSI), defined in text	$$\mathrm{rKSI} = \frac{100}{A_C}\cdot\int_{X_{min}}^{X_{max}} D_n dx$$
rOVER , defined in text	$$\mathrm{rOVER} = \frac{100}{A_C}\cdot\int_{X_{min}}^{X_{max}} \max(D_n - D_C, 0)\, dx$$

Relative Combined Performance Index (rCPI)	$\mathrm{rCPI} = \frac{1}{4}(\mathrm{rKSI} + \mathrm{rOVER} + 2 \cdot \mathrm{rRMSD})$

In the definition of the rKSI and the rOVER the following parameters are used:

- D_n: absolute difference between the normalized cumulative distributions of the test and reference irradiance data in a specific irradiance interval
  $$D_n = \left|\mathrm{normdistr}_{cum,s} - \mathrm{normdistr}_{cum,r}\right|$$
  - $\mathrm{normdistr}_{s/r} = \frac{\mathrm{hist(s/r,bins=100)}}{\mathrm{sum(hist(s/r,bins=100))}}$, where hist describes the histogram of the irradiance data sets s or r, respectively, using 100 bins.
  - $\mathrm{normdistr}_{cum,s/r}$ is the cumulative distribution according to $\mathrm{normdistr}_{s/r}$
- $D_n = \left|\mathrm{normdistr}_{cum,s} - \mathrm{normdistr}_{cum,r}\right|$n: irradiance interval number
- x: irradiance
- $X_{min/max}$: minimum and maximum values of the irradiance time series
  - $X_{max} = \max(r)$
  - $X_{min} = \min(r)$
- $D_C = \frac{\Phi(N)}{\sqrt{N}}$ with the approximation $\Phi(N) = 1.63$
- $A_C = D_C \cdot (X_{max} - X_{min})$

By design, rKSI is 0 if the test and reference data distributions can be considered identical.

rOVER describes the relative frequency of exceedance situations, when the normalized distribution of test data points in specific bins exceeds the critical limit that would make it statistically undistinguishable from the reference distribution.

A small rCPI indicates a good performance of the test data set.

The above metrics are calculated for each evaluated test data set for each year (YYYY-01-01 to YYYY-12-31) and each month. Furthermore, the mean annual values are used to calculate the weighted average metrics. The applied weight for each year is the number of available hours per year. Station years with <1000 h/year are discarded.

Only hours considered as valid are used for the benchmark. The definition of a valid hour is presented in the following section. The minimal solar elevation for the benchmark is 10°. A time interval that includes some data with lower elevation is processed if the remaining valid data represent more than 83% of this time interval (i.e., less than 10 missing minutes). In this case, the data points <10° are also included.

If data points are missing in the test data sets, these data points are excluded and reported by specifying the number of missing points and the irradiation sum related to the gaps that is derived from the reference data.

4 QUALITY CONTROL AND DATA SELECTION

Ground-based irradiance measurements typically contain time intervals with significant errors caused, for example, by instrument malfunction, maintenance issues, or problematic environmental conditions (e.g., rain drops, dew, or ice on the sensor). Hence, a quality control (QC) procedure is needed to detect such erroneous (or potentially erroneous) data and ultimately exclude them from the reference data used in high-accuracy applications such as this benchmark exercise. QC methods of various kinds have been proposed, e.g., (Espinar et al. 2011; Long and Dutton 2002; Maxwell et al. 1993; Long and Shi 2008). Each one of these consists of a suite of automatic tests. However, automatic tests alone are insufficient because they typically miss certain types of errors, and often mislabel valid data as erroneous. In the vast majority of cases, an expert visual inspection step must be added to automatic QC to obtain the best possible results (Forstinger et al. 2021).

A harmonized QC procedure is used here for the benchmark in the form of a “best-of” method based on a combination of a variety of tests that have already been published and widely recognized, while adding expert visual inspections.

In order to perform QC of such a large database, several radiometric stations were assigned to a number of experts, all co-authors of this report. Their QC results differed to a certain degree for various reasons. For instance, experts might have different opinions on what constitutes a bad data point, or they might have practical experience with a specific instrument model or with unusual measurement situations. More pragmatically, coding errors might have been inadvertently introduced by one expert in some cases. The evaluators implemented the QC method individually so that any difference in implementation could be traced back for further improvements and better documentation. The deviations of the results between different evaluators were then compared to the variation in the fraction of usable data for the 161 stations. The comparison of the evaluators’ results showed sufficient consistency of the method, as described in more detail in (Forstinger et al. 2021). The QC method is described in the following subsection. The application of the QC results to determine if an hour is used in the benchmark is explained in subsection 4.2. The QC results are included in the published reference data sets mentioned in subsection 3.2 ((Forstinger et al. 2021), link).

4.1 QC methodology for 1-minute data

The QC methodology consists of many tests that are selected from the literature and applied in stepwise progression. In addition to these automatic tests, the evaluators also reviewed and reported the available information on instruments, calibration, maintenance, and records of any special events at each station if such detailed information was available. The visual inspection of such a large database (686 station-years of 1-min data from 161 stations) constitutes an important accomplishment, at a scale never attempted before.

Several sets of QC tests have been specifically designed for historic radiation databases, such as BSRN (Long and Dutton 2002), SERI QC (Maxwell, Wilcox, and Rymes 1993), QCRad (Long and Shi 2008), MESOR (Hoyer-Klick et al. 2008; Hoyer-Klick et al. 2009), ENDORSE (Espinar et al. 2011), RMIB (Journée and Bertrand 2011), or MDMS (Geuder et al. 2015). In these existing methods, the various tests use different types of threshold limits for the three individual irradiance components—DNI, GHI, and DIF— as well as parameters derived from these components together with additional quantities such as solar position angles or clear-sky irradiance. The types of limits are:

- physical possible limits
- extremely rare limits.
- rare limits.

The existing QC tests have been compared and critically discussed by experts within the framework of IEA PVPS Task 16. Considering the diversity of monitoring stations currently existing in the world, two separate methods have been devised, (i) for the ideal case when measurements of all three irradiance components (GHI, DIF, DNI) are available; and (ii) for the case when only two components are measured (GHI and either DIF or DNI). The latter case is typical of remote solar resource stations that are equipped with an RSI; see details in (Sengupta et al. 2021). Whenever DIF is rather measured with a thermopile pyranometer equipped with a manually-operated shadowband attachment, more QC tests should be implemented (Nollas, Salazar, and Gueymard 2023), but such stations were not included in this study.

Each QC test generates a specific flag for each timestamp. Each flag can take one of three possible values: “data point seems fine”, “data point seems problematic”, or “test could not be performed”. The latter situation can occur because of a missing timestamp/data or because the test requirements were not met (e.g., the irradiance was not above the required threshold), and thus the test could not be applied.

The visual inspection of the data is important to detect any “bad” point that was not detected by the automatic tests, and manually assign a specific flag. This step also includes checking the metadata, if available (logbook with maintenance schedules, reported issues, calibration information, general comments, etc.). Visual inspection can also help determine if the timestamps refer to the start or the end of the averaging interval (e.g., 1-min, 10-min or 1-h averaging), since this information is often not provided (or can be erroneous). The correct interpretation of the timestamps is essential for practically all QC tests but also for any validation or benchmarking exercise to ensure that the reference and test data align rigorously. Furthermore, errors in the correct time zone or station coordinates can also only be identified through visual examination by an expert.

The applied QC tests are defined and described in detail below. All test results are visualized using appropriate public-domain software and provide automatically generated flags. Manual flagging is also permitted, thus providing a way to flag data that passed the automated QC tests. The applied tests are:

- Missing timestamps
- Missing values
- K-Tests (Geuder et al. 2015; Gueymard 2017)
- BSRN’s closure tests (Long and Dutton 2002)
- BSRN’s extremely rare limits test (Long and Dutton 2002)
- BSRN’s physically possible limits test (Long and Dutton 2002)
- Tracker-off test, improved from (Long and Shi 2008)
- Visual inspection, including
    - shading assessment,
    - closure test,
    - AM/PM symmetry check for GHI, and
    - calibration check using the clear-sky index (GHI divided by clear sky GHI).

All the automatic tests, as well as the visual review, are discussed in more detail in what follows.

Missing timestamps

Missing timestamps, which might occur during a data logger reset or data acquisition failure, are identified and filled in with the "not a number" data type (NaN). This ensures that, at the end of the QC procedure, all data files are serially complete.

Missing values

After adding any missing timestamps, the total number of missing data can be determined to provide an overview of the data completeness of each station.

K-Tests

Various studies, e.g., (Geuder et al. 2015; Gueymard 2017), have defined a number of tests to verify that each data point is within physical limits and to detect possible tracker issues. These tests are based on the clearness indices K_n, and K_t, the diffuse fraction K, and their physical relationships. These normalised quantities are defined as

$$K_n = \frac{DNI}{ETN} \qquad \text{(eq. 1)}$$

$$K = \frac{DIF}{GHI} \qquad \text{(eq. 2)}$$

$$K_t = \frac{GHI}{ETN \cdot \cos(SZA)} \qquad \text{(eq. 3)}$$

where ETN is the extraterrestrial irradiance at normal incidence, and SZA is the solar zenith angle. ETN is obtained as the product of the solar constant, 1361.1 W/m^2 (Gueymard 2018), and the sun-earth distance correction factor, which is calculated by the sun position algorithm. The suite of K-tests is applied within each appropriate domain; the corresponding flag names are indicated in Table 4. If the condition is not fulfilled and the data point is within the appropriate domain, the point is flagged with the corresponding flag name. Because the measured GHI at 1-minute resolution can be much higher than the corresponding clear-sky value during cloud-enhancement periods (Gueymard 2017), the upper threshold for K_t is adjusted here for the use of 1-min data. It might have to be decreased for data with a lower resolution (e.g., 5-min or 10-min resolution).

Table 4: Performed K-Tests. ALT denotes the station altitude above mean sea level (AMSL) expressed in m. ETN is the extraterrestrial irradiance in W/m².

Condition	Domain	Flag name
$K_n < K_t$	$(GHI > 50\ W/m^2$ and $K_n > 0$ and $K_t > 0)$	flagKnKt
$K_n < (1100 + 0.03 * ALT)/ETN)$	$(GHI > 50\ W/m^2$ and $K_n > 0)$	flagKn
$K_t < 1.35$	$(GHI > 50\ W/m^2$ and $K_t > 0\ )$	flagKt
$K < 1.05$	$(SZA < 75°$ and $GHI > 50\ W/m^2$ and $K > 0\ )$	flagKlowSZA
$K < 1.1$	$(SZA \geq 75°$ and $GHI > 50\ W/m^2$ and $K > 0\ )$	flagKhighSZA
$K < 0.96$	$(K_t > 0.6$ and $GHI > 150\ W/m^2$ and $SZA < 85°$ and $K > 0\ )$	flagKKt

BSRN's closure tests

To test the expected correspondence between the GHI, DNI, and DIF irradiance components, i.e., deviation from the ideal closure, the BSRN closure tests are applied (Long and Dutton 2002). If the conditions described in

Table 5 are not fulfilled in the corresponding domain, the data point is flagged with a descriptive flag.

Table 5: Application of the three-component closure test

Condition	Domain	Flag name
$\left\lvert \frac{\mathrm{GHI}}{\mathrm{DNI}\cdot\cos(\mathrm{SZA})+\mathrm{DIF}} - 1 \right\rvert \leq 0.08$	$(\mathrm{SZA} \leq 75° \text{ and } \mathrm{GHI} > 50\ \mathrm{W/m^2})$	flag3lowSZA
$\left\lvert \frac{\mathrm{GHI}}{\mathrm{DNI}\cdot\cos(\mathrm{SZA})+\mathrm{DIF}} - 1 \right\rvert \leq 0.15$	$(\mathrm{SZA} > 75° \text{ and } \mathrm{GHI} > 50\ \mathrm{W/m^2})$	flag3highSZA

BSRN's extremely rare limits test

The three irradiance components are also tested in comparison with extremely rare limits (Long and Dutton 2002). If the condition for each component is not fulfilled for a data point, that point is flagged with the corresponding flag name, as described in Table 6.

Table 6: Application of the extremely rare limits tests

Condition	Domain	Flag name
$-2\ \mathrm{W/m^2} \leq \mathrm{GHI} \leq 1.2 \cdot \mathrm{ETN} \cdot \cos^{1.2}(\mathrm{SZA}) + 50\ \mathrm{W/m^2}$	all data	flagERLGHI
$-2\ \mathrm{W/m^2} \leq \mathrm{DIF} \leq 0.75 \cdot \mathrm{ETN} \cdot \cos^{1.2}(\mathrm{SZA}) + 30\ \mathrm{W/m^2}$	all data	flagERLDIF
$-2\ \mathrm{W/m^2} \leq \mathrm{DNI} \leq 0.95 \cdot \mathrm{ETN} \cdot \cos^{0.2}(\mathrm{SZA}) + 10\ \mathrm{W/m^2}$	all data	flagERLDNI

BSRN's physically possible limits test

In addition to the extremely rare limits, the physically possible limits of each component are tested as well (Long and Dutton 2002). Considering the high-quality requirement for the benchmark application envisioned here, both tests are required. If the condition for each component is not fulfilled for one data point, the point is flagged with the corresponding flag name (Table 7).

Table 7: Application of the physically possible limits tests

Condition	Domain	Flag name
$-4\ \mathrm{W/m^2} \leq \mathrm{GHI} \leq 1.5 \cdot \mathrm{ETN} \cdot \cos^{1.2}(\mathrm{SZA}) + 100\ \mathrm{W/m^2}$	all data	flagPPLGHI
$-4\ \mathrm{W/m^2} \leq \mathrm{DIF} \leq 0.95 \cdot \mathrm{ETN} \cdot \cos^{1.2}(\mathrm{SZA}) + 50\ \mathrm{W/m^2}$	all data	flagPPLDIF
$-4\ \mathrm{W/m^2} \leq \mathrm{DNI} \leq \mathrm{ETN}$	all data	flagPPLDNI

Tracker-off test

Since, for most stations, the direct and diffuse components are obtained with a tracker equipped with a pyrheliometer and a pyranometer with shading disc or ball, a tracker failure results in incorrect values for both measurements. The causes of such failures include electromechanical problems within the tracker, loss of power, misalignment or timestamp errors, etc. Detecting such problems is critical, but can be difficult, particularly in the case of slight mistracking. The tracker-off test involves comparisons with rough estimates of the coincident clear-sky irradiance components ($\mathrm{GHI_{clear,s}}$, $\mathrm{DNI_{clear,s}}$, $\mathrm{DIF_{clear,s}}$), which are here obtained as a fixed fraction of the extraterrestrial irradiance at horizontal incidence, ETH = ETN cos(SZA) (Table 8). If all conditions described in Table 8 are not fulfilled for any data point, it is flagged with the corresponding flag name.

Table 8: Application of the tracker-off test

Conditions	Definitions	Flag name
$\frac{(\mathrm{GHI}_{\mathrm{clear,s}} - \mathrm{GHI}_{\mathrm{measured}})}{(\mathrm{GHI}_{\mathrm{clear,s}} + \mathrm{GHI}_{\mathrm{measured}})} < 0.2$ $\frac{(\mathrm{DNI}_{\mathrm{clear,s}} - \mathrm{DNI}_{\mathrm{measured}})}{(\mathrm{DNI}_{\mathrm{clear,s}} + \mathrm{DNI}_{\mathrm{measured}})} > 0.95$ $\mathrm{SZA} < 85°$	$\mathrm{GHI}_{\mathrm{clear,s}} = 0.8 \cdot \mathrm{ETH}$ $\mathrm{DIF}_{\mathrm{clear,s}} = 0.165 \cdot \mathrm{GHI}_{\mathrm{clear,s}}$ $\mathrm{DNI}_{\mathrm{clear,s}} = \frac{\mathrm{GHI}_{\mathrm{clear,s}} - \mathrm{DIF}_{\mathrm{clear,s}}}{\cos(\mathrm{SZA})}$	flagTracker

Visual inspection with a multi-plot

All test results and selected irradiance data are compiled into a single multi-plot arrangement for easy visualization. Such a plot is made for each year and for each station (see, e.g., Figure 8 for the Visby station, 2016). For a larger examplary image of the multi-plot and example Python code, please refer to https://github.com/AssessingSolar/solar_multiplot. More specifically, these plots not only include visualization of the test results discussed above, but also:

(1) visualization of the deviation of the measured DNI by the pyrheliometer from the DNI calculated from DIF and GHI (i.e., closure error);

(2) an overview of the diurnal variation of DNI and GHI as a function of time and solar position;

(3) the clear-sky index calculated as the ratio between the measured GHI and the clear-sky GHI from the public-domain McClear v3's database (Lefèvre et al. 2013; Gschwind et al. 2019; Qu et al. 2017);

(4) a comparison between the pyranometer GHI observation and that calculated from DNI and DIF;

(5) comparisons of the pyranometer GHI measurements before and after solar noon to identify possible levelling or timestamp errors; and

(6) visualization of the data points in K-space with the applied limits.

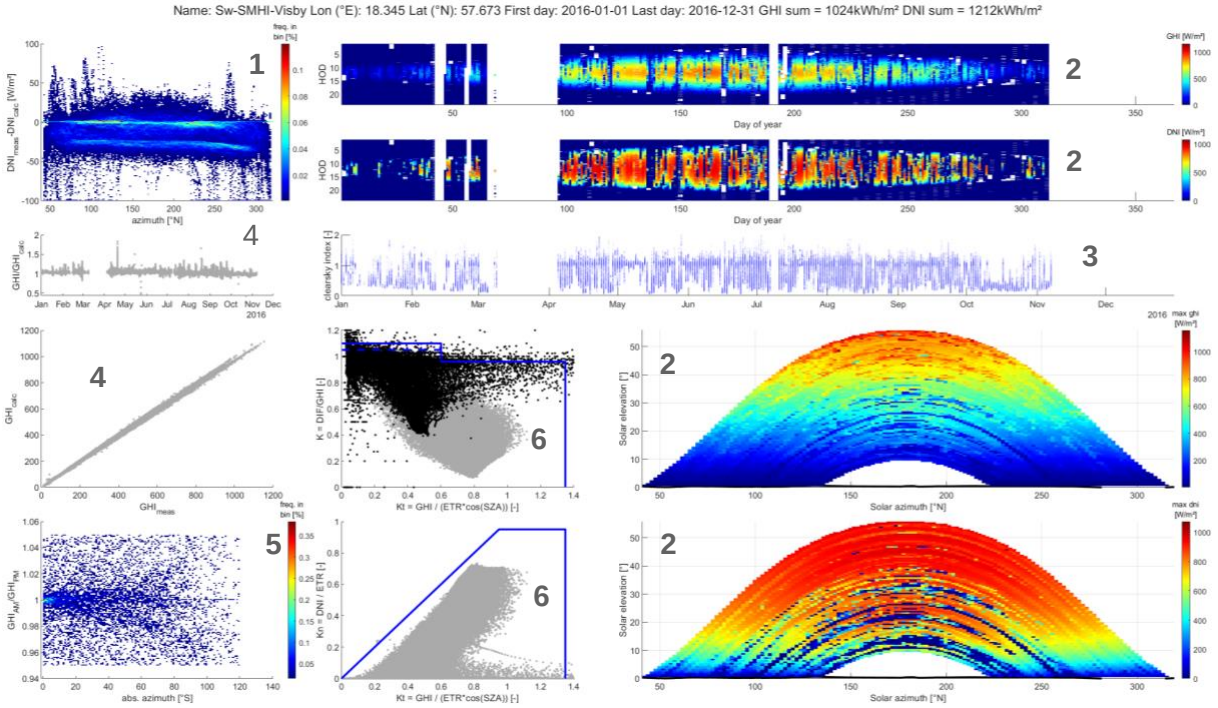


Figure 8: Visualization of various QC tests used to evaluate the quality of irradiance data at one station (Visby, Sweden, 2016). Numbers in boldface refer to the description in the text.

A multi-plot like the one shown in Figure 8 is created not only from the raw (pre-QC) data, but also from the data points that pass the automatic flagging (as an intermediate result in order to visualize the data flagged by the automatic checks), and finally from the data points that pass the complete QC, including manual revisions. Sorting out the already detected suspicious data points before the plotting step allows for a better visual control of the remaining data points. If suspicious data points are found, further visualizations can be used to confirm whether those points are invalid, in which case an overriding manual flag is set that can be used to exclude such points from processing. For each station and year, the three kinds of multi-plots (raw data, data points that pass the automatic flagging, and final selection) just described, offer a complete overview of the station data.

Plot (1) in Figure 8 shows the deviation between the measured and calculated DNI with respect to the sun's azimuth angle. In this case, two distinct levels appear over the year. To detect if a specific issue existed at the station (e.g., long periods without cleaning or with a tracker issue), one needs further visualization of the data. One example is shown in Figure 9, which describes the diurnal variation of DNI for each day of a complete year. The day of the year appears on the x-axis, whereas the time of day is shown vertically, using *true solar time* to emphasize the expected symmetry around solar noon. The upper plot shows a clearly different deviation pattern in the later part of the year (black rectangle). Consultation of the log book available for that site led to the conclusion that this is not a station issue per se, but the result of a sensor change. That change resulted in a slightly different configuration in terms of levelling, alignment, instrument response, and overall performance. The lower plot of Figure 9 is for the same station but a different year. It shows a change in deviation caused by sensor soiling, which remained noticeable over a long period. Whereas the sensor changes in the upper plot of Figure 9 do not lead to an exclusion of the data, the sensor soiling shown in the lower plot of Figure 9 may lead to data exclusion. This demonstrates the necessity of manual expert QC

and the general need for station log books, in which cleaning intervals and sensor changes are recorded.

The GHI clear-sky index time series (3) is helpful to reveal whether the GHI sensor's calibration is outdated, incorrect or its sensitivity drifts over time. The clear-sky index is expected to be ≈1 under clear-sky conditions. However, this is rarely the case in the real world. One main reason is that the clear sky GHI is only an approximation at any instant. Nevertheless, cases where the clear-sky index remains constant and well below 1 can be an indication of a calibration issue. Similarly, an abrupt or step-like change of the clear-sky index under clear conditions is typically the signature of a change of calibration factor by a substantial amount, or of the result of cleaning a dirty sensor. Again, an expert is needed to decide whether a calibration issue is likely at the station. In Figure 8, the clear-sky index is constant and well below 1 in the later part of the year, but this is the result of cloudy weather conditions rather than a calibration issue. This is apparent when comparing the heat maps of GHI and DNI (plot (2) in Figure 8).

If the expert detects issues with individual data points, those are flagged with "flagManual". The results of the individual tests, in the form of a quality flag per test, are properly documented (metadata) and packed into one single file per site and year, which also includes the solar irradiance observations. Finally, all flags are combined into a single usability code, indicating an objective level of quality for each data point.

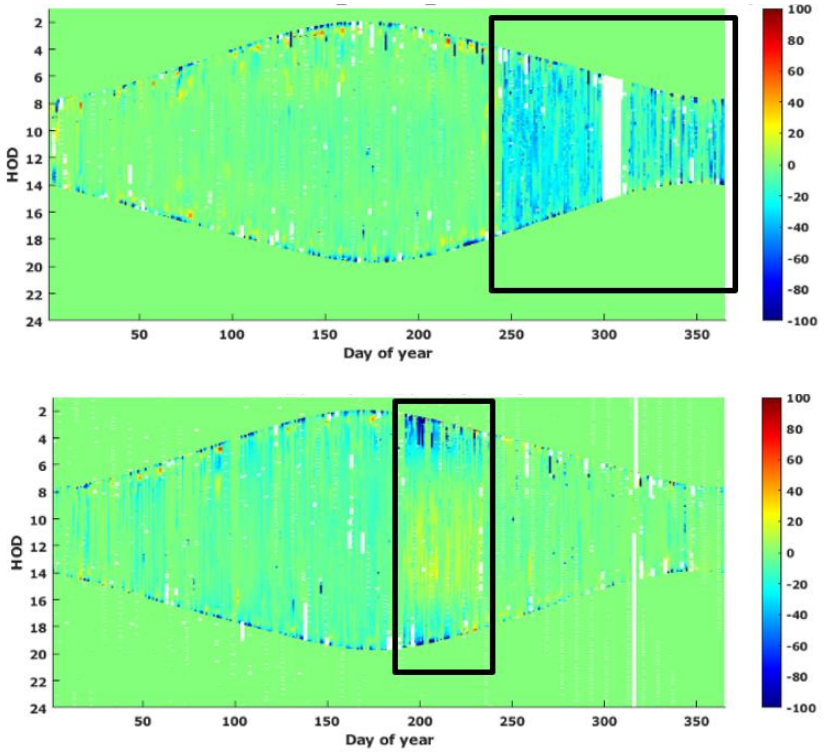


Figure 9: Heat maps of the difference between measured and calculated DNI in W/m² with respect to day of the year (x-axis) and solar time of the day (y-axis) for different station-years. Top: Visby 2015; Bottom: Visby 2019.

4.2 Data selection

Since the benchmark is carried out at hourly resolution, the 1-minute QC flags have to be combined to indicate whether the hourly averages are valid or not. The following method is applied to determine the validity of hourly averages:

- Split each hour in 12 “5-min intervals” (e.g., 1 to 5 min; 6 to 10 min, …)
- Count the number of good samples in each 5-minute interval (minimum 0, maximum 5). Each 1-min sample is labelled as good if it has not failed any QC test and if all three radiation components are present.
- Count the number of 5-minute intervals in an hour containing at least 3 good samples. If at least 3 good samples are found, the 5-minute interval is classified as “OK”, otherwise as “NOT OK”.
- The hour is only included in the benchmark if at least 10 5-minute intervals are “OK” (83% of the complete hour).

A special case had to be considered for the ACCESS G3 modelled data set because it does not provide DNI data. The normal rule that all components must be available is simply not enforced in that case.

Regarding the data selection and exclusion process, it is important to ensure that the remaining data set still represents the conditions at the site from a statistical standpoint. Because data collected under complex and variable cloud conditions might be flagged as suspicious or erroneous more frequently than under clear skies, there is a risk of artificially biasing the data set toward less cloudy conditions than what actually occurs. It is possible to confirm that, in terms of cloudiness, the final data sets are close to the original data sets (before QC) by comparing the histograms of the GHI clear-sky indexes before and after data exclusion. The changes in the histograms caused by the data exclusion resulting from the method described above can be considered negligible.

All measured GHI, DNI, and DIF valid data points (i.e., that have passed the QC tests) are used for the calculations.

5 BENCHMARK RESULTS

This section presents an overview of the results of the benchmark and allows the reader to analyse specific stations or groups of stations in more detail. This overview uses further plots and results that are included in the data Annex of the report (DOI: 10.5281/zenodo.7867002). A presentation-style approach is used because the answer to the key question addressed here (*Which data set is most adequate at site X*?) depends on the site location, the application considered (e.g., accuracy requirements), and other technical, practical, or even subjective standpoints. Hence, the user should rely on a contextual assessment rather than uniquely on statistical results. Because it is impossible to discuss each individual site or region in detail here, the objective is to provide guidance so that the reader can be empowered and able to successfully analyse their case based on the specific information in the data Annex.

The results of the benchmark are presented in world maps with color-coded dots for each station and in color-coded tables for station subgroups, as explained in the following subsections. The subgroups are continents and/or climate zones. The statistical metrics are defined above in Section 3. Furthermore, scatter density plots of modelled vs. reference data for each station, year, and radiation component are included for a quick evaluation of the dispersion.

Among all metrics, rMBD is of paramount interest for the analysis as it is directly related to the overall under- or over-estimation of the solar energy resource at any given site. Hence, the examples provided below are mostly for rMBD. Moreover, the variation of rMBD from year to year and site to site within a certain region is of interest to estimate the reliability of a data set. Similarly, the distribution of the deviations and the histograms of the annual irradiance data are also of interest. The quality of these distributions is described by the other metrics and visualized in the scatter plots.

Obviously, anyone interested in the data quality of modelled data sets for a site that is included in this benchmark will consider the results for this specific site as most important. Such results can be seen best in the result tables and scatter plots. To estimate the data quality for a site that is geographically close and in a climate similar to any site covered in this benchmark, the results can be analysed in the same way.

If the objective is rather to estimate the data quality for a specific region or climate zone, the result tables for those groups are appropriate. Additionally, the world maps can be analysed by focussing on the specific regions of interest.

If one or more specific modelled data set is of interest, e.g., to find where in the world it is most accurate, the corresponding world maps, result tables, and scatterplots should be used and compared.

The different presentations of the results are described in the following sections based on examples. Results for all stations and parameters are found in the data Annex.

5.1 Scatter density plots

Scatter density plots have been created for all stations, all years, and the two essential radiation components (GHI and DNI). The plots show the measured reference irradiance on the x-axis, the modelled test irradiance on the y-axis, and the color denotes the number of data points within a data bin. The x- and y-axes are the same for all plots , whereas the range of the color bar is specific to each plot. Because of the large quantity of possible scatterplots, it

is virtually impossible to scrutinize all those that are related to a specific case. To help the reader into using the scatterplots more efficiently and on a wider scale, various groups of such plots are combined into single ensemble charts. They are provided as images (in png format), as follows:

- Charts showing results from all data providers at one station, year, and component in a single file
- Charts showing results from a single data provider at all stations, for each year and component in a single file.

The left and right parts of Figure 10 are two examples of the first type of these combined plots for the station of Cabauw (CAB) for 2016. The second type is created in two different ways, which differ depending on the number of stations covered by each data provider. In one option, the plots are placed such that each station is always in the same position within the figure, irrespective of the number of stations covered by the specific data provider. This leads to small-size scatter density plots per station because 129 of them are analyzed here. As a remedy for this issue, the second option rather shows all scatterplots as big as possible to fit into one such single figure. In that case, the drawback is that the position of a station in the figure changes from one data provider to another.

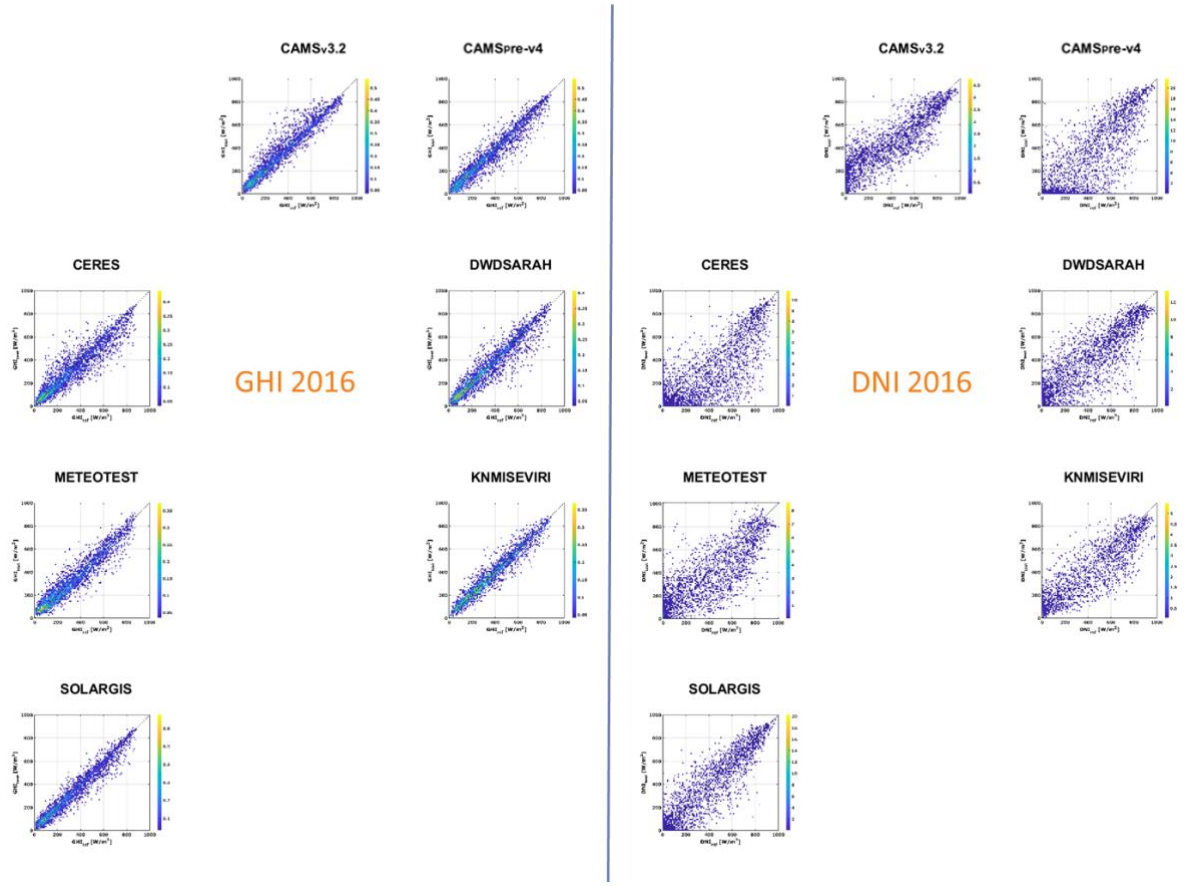


Figure 10: Exemplary scatterplots for Cabauw in 2016: GHI (left) and DNI (right).

Examples of GHI and DNI scatterplots for 2016 from various data providers are also shown for the station Izaña (IZA) in Figure 11. That station has been selected as an example as it sticks out from other stations, as previously noted in (Yang and Gueymard 2021), which motivates a closer analysis with the scatter density pots. It is obvious that GHI is modelled more accurately

than DNI and that some models show much stronger deviations than others. IZA is located on the Canary Islands at 2373 m AMSL. The high altitude brings along modelling complications because it is difficult to ascertain whether the station is below or above clouds from satellite images alone. Moreover, snow cover cannot be easily distinguished from cloud cover. There might also be deviations in the modelled data sets caused by strong variation in altitude within the spatial resolution of the various input data sets—most importantly in relation to aerosols and water vapor. The issues found in the IZA case are likely to apply at similar locations (mountain sites, low latitudes). These issues and resulting uncertainties also have to be considered when evaluating the performance of any satellite-based model over surrounding areas, which might often result in incorrect conclusions.

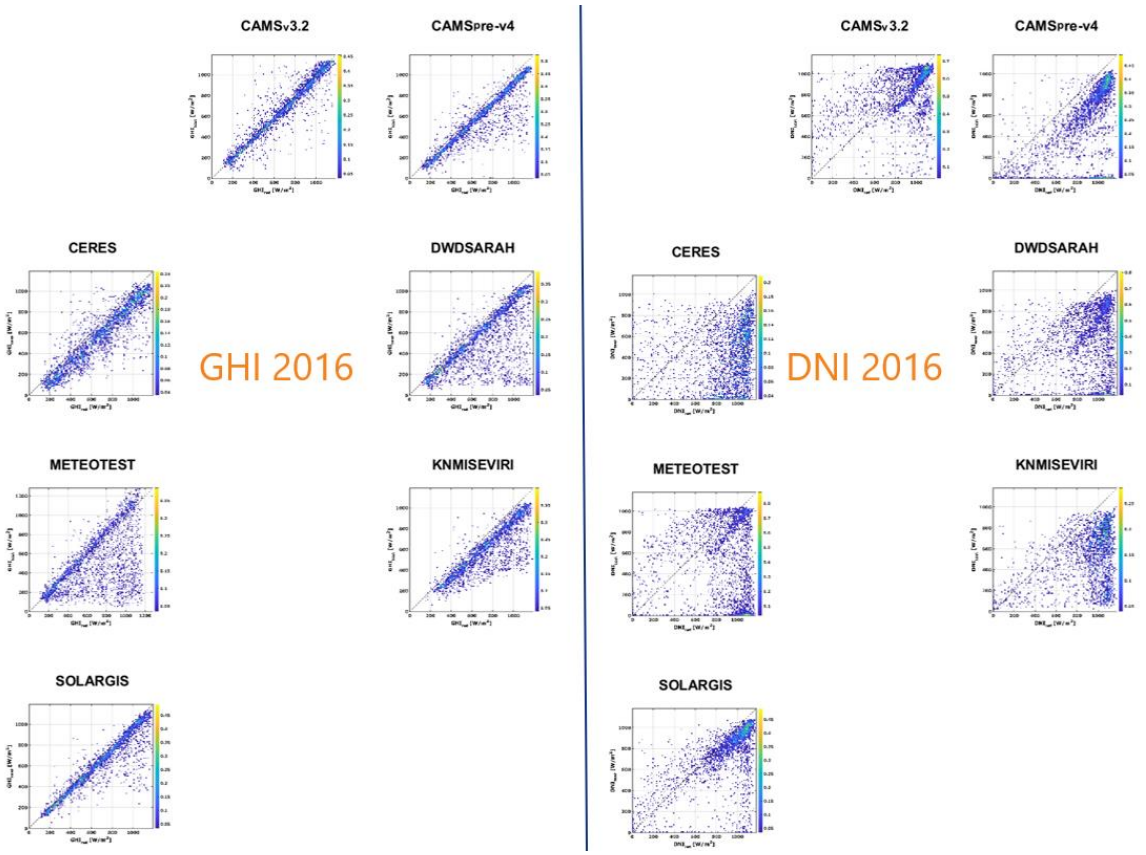


Figure 11: Exemplary scatterplots for Izaña in 2016: GHI (left) and DNI (right).

5.2 World maps

To obtain an overview of the benchmark results, world maps with color-coded dots for the analysed error metrics are used here. Figure 12 shows rMBD, rRMSD, rMAD, and rKSI for GHI for all data providers, stations, and years in a single plot. To combine the results for different years, the weighted averages of all years for each specific metric are used. The point size corresponds to the total number of data points from all considered years included in the analysis.

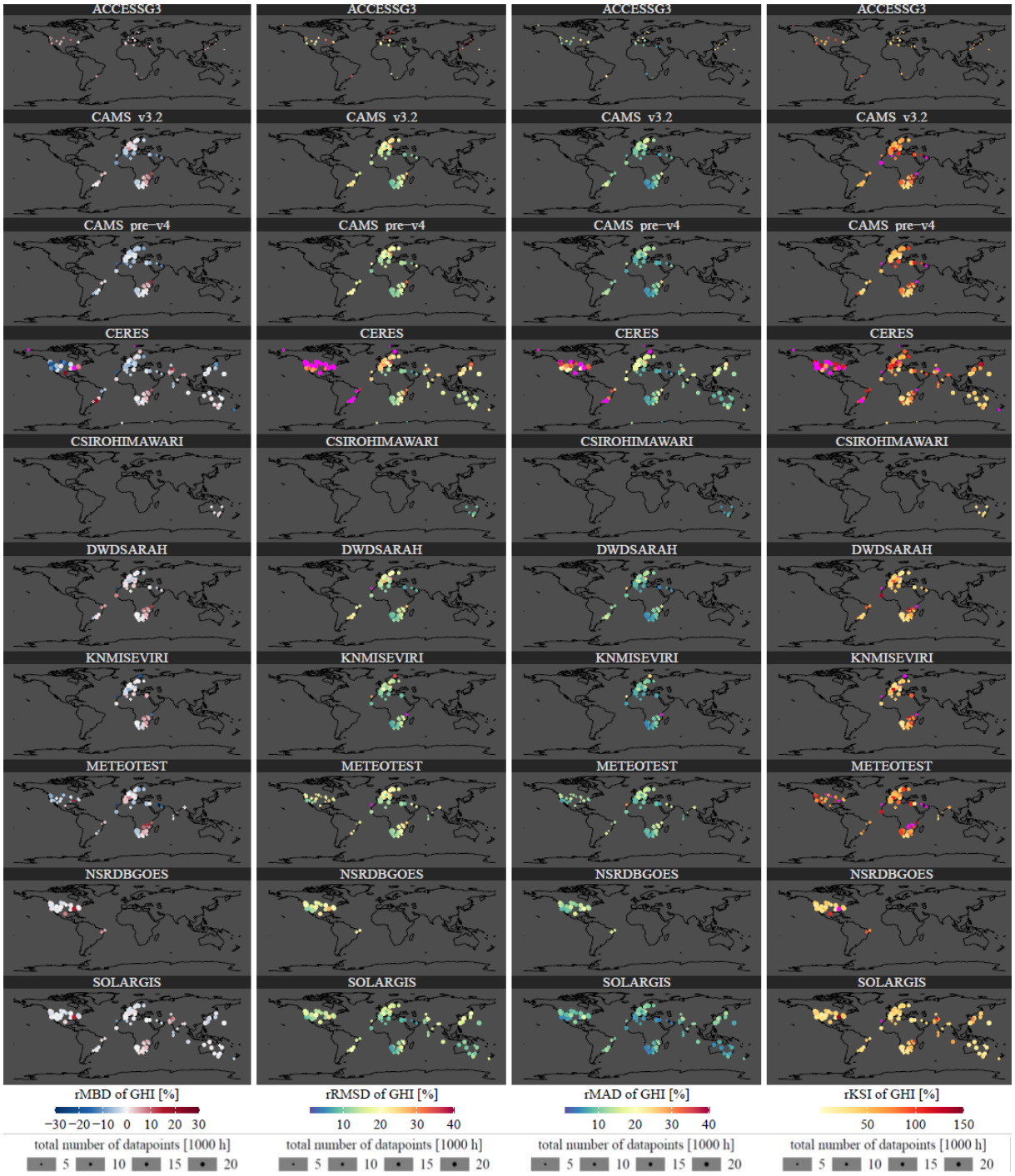


Figure 12: GHI benchmarking results for all stations and years, using four error metrics. Magenta color indicates stations out of the color bar range.

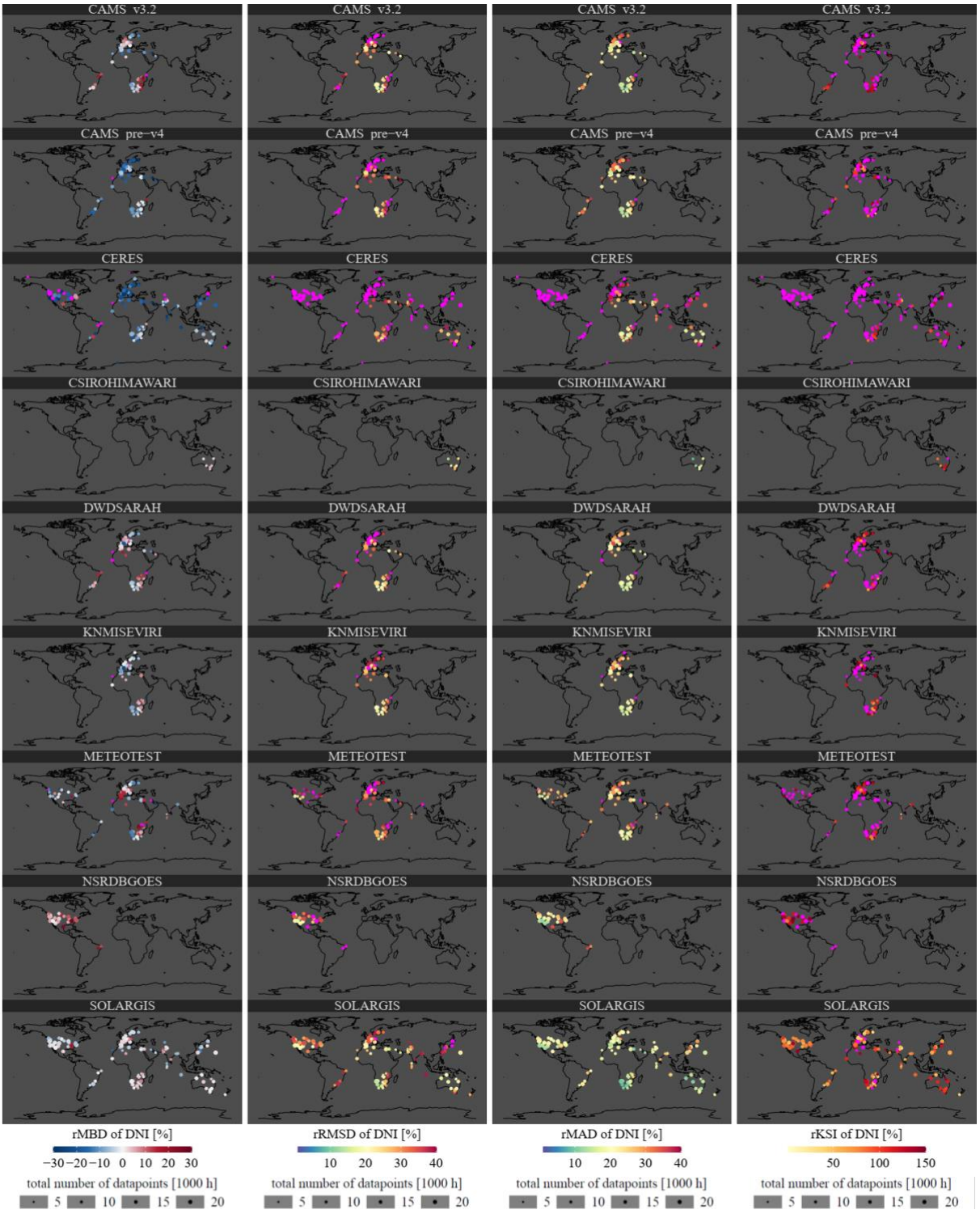


Figure 13: DNI benchmarking results for all stations and years, using four error metrics. Magenta color indicates stations out of the color bar range.

The overview plots in Figure 11 exemplify the variation of the results for different stations, clearly indicating that some stations are affected by less accurate estimates from multiple models. For some models, a clear dependency on region or continent is also visible. Moreover, the different data set size per station and provider can be seen. Various general tendencies of

some data sets are also found. For example, the NWP ACCESS 3G data set shows almost exclusively positive bias. The higher deviations of the CERES data sets compared to the other satellite-based data is apparent, at least in part because of the much larger size (≈110x110 km) of its pixels compared to that of models based on cloud data from geostationary satellites (≈4x4 km). Interestingly, these deviations appear generally highest over the Americas.

Over southern Africa, all data sets generally benefit from lower rRMSD and rMAD than they do elsewhere. Some data sets also show below-average rRMSD and rMAD results over Australia and the southwestern USA. These results are most likely related to the high irradiance levels found over those regions, which are affected by only low and relatively stable attenuation from clouds and aerosols. In other regions with low cloudiness but higher aerosol loads, such as northern Africa, fewer data sets show below-average rRMSD or rMAD.

The same metrics as in Figure 12 are shown for DNI in Figure 13, with the same colour bars as for GHI and each metric. It is obvious that the deviations are typically higher for DNI than for GHI, and that large deviations exist at more stations. This was expected and corroborates findings of previous studies. This result can be explained by a stronger error propagation effect on the modelled DNI caused by uncertain input data (in particular for clouds and aerosols), which is made even stronger by any significant discrepancy between station elevation and the pixel's mean elevation. Examples stations with higher deviations are Izaña (Canary Islands, mountain site with high in-pixel topographic inhomogeneity) and Dar es Salaam (Tanzania, site with prevalent high aerosol load and smog).

World map plots for additional metrics can be found in the data Annex.

5.3 Results overview per continent

At continental scale, it is important to obtain a more detailed overview of the results, which in turn facilitates a better comparison of different data sets. Specific tables with color-coded metrics are created to that effect. There is one table per continent, radiation component, and metric. Such tables are also provided for each climate zone, radiation component, and metric.

As examples, two such tables are shown in Table 9 and Table 10. The same general structure is used for all other cases. The table title specifies the sub-group (continent or climate zone), the radiation component, and the metric. Below the title, a list of the (abbreviated) site names is shown in alphabetical order from left to right. Tier-2 stations are marked with an asterisk (*) to distinguish them from Tier-1 stations. The reference data points from Tier-2 stations are typically associated with higher uncertainty because not all QC tests can be performed and the instrumentation's accuracy is typically lower. These reasons might lead to measurement-induced higher error metrics compared to Tier-1 stations. The uncertainty of the applied reference data is expected to vary from one station to another and even from time stamp to time stamp for a specific station. An individual uncertainty analysis per station and time interval is extremely complex. After the detailed quality control and the rejection of suspicious data, a significant part of the variation of the uncertainty with station and time is removed. As a simplification, we estimate the uncertainty of Tier-1 and Tier-2 stations and DNI and GHI based on the literature (Sengupta et al. 2021).

In well maintained field measurement campaigns with fully functional instruments, the DNI measurements of Tier-1 stations are associated with a standard uncertainty (1-sigma, 68%) of about 1.5% and GHI measurements with about 2% (Sengupta et al. 2021). In problematic cases and the most relevant zenith angles, the quality control removes data with deviations of calculated and measured GHI of 8%. We expect the uncertainties of the used data to be lower

than this limit defined by the quality control and close to the mentioned uncertainties for good maintenance.

For well-maintained Tier-2 stations, the standard uncertainty is estimated as 5% for DNI and for GHI as 4% (Sengupta et al. 2021). As the three component test cannot be performed for the Tier-2 stations, the upper uncertainty limit that is related to the quality control for these stations is higher than in the case of the Tier-1 stations, and we expect it to be about 10%. As in the case of the Tier-1 stations we also expect the uncertainty to be close to the estimations for well-maintained stations.

In parallel, stations that have been used for post-processing during the production of some of the modelled data sets (Solargis and Meteotest) are marked with the "^" symbol. The first column from the left provides the list of the modelled data providers that supplied irradiance data at each site. In the color-coded part of the tables, the weighted average of the metric considering the mean annual results (e.g., rMBD in percent for Tables 9 and 10) of each test data set at each site is shown. The weighting factor for each year is the number of data points evaluated in the year for the specific modelled data set. Note that, depending on the provider, the evaluation at any specific site might be based on a different number of data points or years, if the test data contains gaps or covers a different time period.

The columns titled “Mean” and “Std” to the right of the color-coded part of the table provide the mean and standard deviation of the corresponding metric for each modelled dataset under scrutiny, as calculated using all the available reference stations. These values are evaluated for a modelled data set only if more than 75% of the sites of the subgroup (continent or climate zone) is included in it. In Table 9, for example, the average and standard deviation are not calculated for ACCESSG3 because this data set does not provide data for at least 75% of the stations. Moreover, only those stations that are covered by all modelled data sets with >75% of the sites are included in the calculation of the average and standard deviation. For example, still in Table 9, the stations KIR, NYA, and VIS are neither used to calculate the average nor the standard deviation.

Table 9: rMBD for GHI over Europe.

Europe: GHI, rMBD (%)

site / model	CAB	CAR^	CAS	CEN	DAO	DAV	JAE*	KAZ	KIR	LOC	LYN	MIL	NOR	NYA	ODE	PAL	PAY	TAB^	TOR	VIS	Mean	Std	Abs_Mean	Abs_Std
SOLARGIS	-2.5	0.6	0.3	-0.1	-2.1	-1.0	1.0	2.6		0.5	0.9	1.9	-1.6		1.7	-1.8	-1.7	0.2	-2.3	-2.6	-0.2	1.6	1.3	0.8
METEOTEST	-3.6	0.1	-0.7	0.9	3.2	5.1	2.7	2.1		0.5	1.7	0.0	-0.7		8.8	-0.2	-2.2	3.4	-4.9	-7.1	0.9	3.2	2.4	2.3
CAMS_v3.2	2.1	-1.1	-0.0	-2.5	-8.1	-6.7	-1.3	4.8		-2.2	0.5	-1.8	-1.0		7.3	2.5	-2.7	-4.2	-3.2		-1.0	3.8	3.1	2.4
CAMS_pre-v4	-4.1	-0.9	-0.8	-2.6	-10.2	-9.0	-2.0	0.8		0.3	-4.5	-1.4	-4.4		-0.2	-3.3	-3.2	-1.8	-8.0	-4.4	-3.3	3.2	3.4	3.1
KNMISEVIRI	0.7	-0.9	0.2	-0.2	-17.7	-17.3	-1.4	-2.3	-19.1	-3.0	1.9	2.7	0.3		-7.7	0.4	-1.3	-2.7	0.2	-0.8	-2.8	6.0	3.6	5.5
DWDSARAH	-1.2	0.9	1.7	-1.0	-21.5	-20.4	-1.5	-0.7		-1.4	0.2	3.5	-1.6		-5.4	1.7	-3.7	1.7	-2.7		-3.0	7.1	4.2	6.4
CERES	-3.2	-8.4	-3.8	-3.5	-8.0	-5.9	-5.1	-4.1	4.3	-7.7	3.2	-1.2	-0.1	-38.8	-7.0	-0.7	-10.0	-6.2	-8.3	-2.4	-4.7	3.5	5.1	2.9
ACCESSG3			2.0		3.2	4.1				0.2	6.4		4.7		6.0	0.2		2.3		-0.9				

Table 10: rMBD for DNI over Europe.

Europe: DNI, rMBD (%)

site	CAB	CAR^	CAS	CEN	DAO	DAV	JAE*	KAZ	KIR	LOC	LYN	MIL	NOR	NYA	ODE	PAL	PAY	TAB^	TOR	VIS	Mean	Std	Abs_Mean	Abs_Std
model																								
SOLARGIS	-1.6	-0.5	-1.5	2.2	13.1	14.6	10.2	10.9		1.3	-3.3	7.3	-2.5		2.2	-1.6	0.3	1.8	-5.4	-2.5	2.8	6.1	4.7	4.7
CAMS_v3.2	9.5	-2.6	0.3	-1.2	-10.6	-10.1	3.1	7.6		-4.2	-4.2	4.7	-9.1		3.5	7.4	-1.7	-4.4	-11.5		-1.4	6.7	5.6	3.6
METEOTEST	-5.5	0.7	-3.1	8.9	20.8	21.7	16.8	11.7		1.9	-1.3	1.8	-3.5		14.3	1.2	3.3	14.1	-9.2	-11.8	5.6	9.5	8.2	7.1
DWDSARAH	4.3	-1.2	-0.2	-1.8	-31.4	-31.0	2.7	-2.8		-1.7	-4.3	12.9	-5.4		-19.7	6.3	-3.8	1.8	-8.8		-4.9	12.0	8.2	9.9
KNMISEVIRI	1.1	-6.3	-4.6	-4.7	-34.8	-35.3	-2.2	-5.8	-0.0	-6.2	-1.1	9.1	-1.1		-23.5	-0.9	-3.5	-9.2	-5.5	-3.9	-7.9	12.0	9.1	11.1
CAMS_pre-v4	-18.6	-7.9	-4.7	-14.1	-19.3	-15.7	-5.2	-9.7		-2.1	-17.8	1.9	-21.0		-18.4	-13.8	-10.2	-4.9	-29.5	-19.0	-12.4	8.1	12.6	7.7
CERES	-20.4	-24.9	-18.5	-11.4	-17.4	-16.3	-17.3	-21.3	-7.2	-27.8	-11.3	-8.5	-18.0	-68.8	-27.8	-14.2	-24.4	-21.7	-19.8	-21.4	-18.9	5.6	18.9	5.6

Regarding the bias metrics (rMBD and MBD), the average and standard deviation of the absolute values of the biases for each station and data set are also shown (see columns titled “Abs_mean” and “Abs_Std” at the far right). Beside the change in how the absolute value is derived, the same procedure as above is applied to calculate the average and standard deviation.

The list of modelled data sets is sorted according to their performance ranking only if they provide data for more than 75% of the stations. For all metrics except the bias, the ranking is based on the average metric shown in the column “Mean”. For the bias (rMBD and aMBD), the ranking is made according to the mean of absolute biases (“Abs_mean”); this is necessary to avoid that a model with both high positive and high negative biases obtains a good ranking if those biases compensate each other when averaging. Modelled data sets that are not ranked because of a too low number of validation sites appear in alphabetic order below the ranked data sets and are separated from these by a black horizontal line.

Although the ranking is of great interest in general, it should not be overvalued, and rather has to be regarded as only one piece of information amongst many. For example, the standard deviation of the metric used for each ranking (column “Std” or “Abs_Std”) is also of importance because it indicates how far from the average the performance for a specific site could be. A modelled data set that would perform well on average (low deviations at many sites), but would show huge deviations at some sites, might constitute a risk in practice. A data set with slightly higher average deviations, but less dispersion overall, could be more recommendable.

Some interesting findings are that (i) the rankings are different for GHI and DNI; (ii) a ranking one might derive according to the metrics for a single station does not always follow the shown ranking that is based on the average; and (iii) different results are obtained for different subgroups (continents or climate zones). One of the data sets (Solargis) is often associated with the lowest average deviation metrics. This corresponds generally well to the performance ranking that one can obtain for many individual stations, but the best data set for a specific station is at times provided by another model.

Because the post-processing performed by some of the providers might lead to too optimistic results, it is important to analyze the general ranking after removing the particular stations used by Meteotest and Solargis to that effect, and to derive a new ranking covering only the remaining sites. This new ranking is found almost unchanged. There are exceptions, however, when only a small number of stations per subgroup exist, but in such cases the ranking is not considered representative anyway because of the low number of remaining stations (e.g., South America, where 3 of 5 sites are removed).

Regarding the analysis for climate zones, it is important to emphasize that less models are ranked because the models share less stations when grouped into climate zones than into continents. This is a direct consequence of the spatial coverage of the modelled data sets.

6 CONCLUSIONS AND SUMMARY

A worldwide benchmark of various modelled radiation data sets of both global irradiance (GHI) and direct normal irradiance (DNI) has been carried out extensively at 129 high-quality radiometric stations on an hourly basis. The measured reference data were submitted to a stringent quality control procedure, performed at 1-min resolution.

The benchmark results have shown noticeable deviations in performance between the various modelled data sets. In particular, it was found that the most appropriate data set actually depends on both site and climate or continent of interest. Some stations are especially challenging for some models, as evidenced by the high deviations observed for several data sets in difficult environments (e.g., high mountains or coastal areas).

The modelled errors and deviations between data sets were found generally much higher for DNI than for GHI, as expected, because of the higher sensitivity of the former to aerosols, clouds, elevation, and other factors.

All modelled data sets evaluated here are based on satellite observations, with the exception of ACCESS G3, which is based on NWP modelled data. The NWP data set does not provide DNI predictions, and generally showed positive bias in its GHI predictions. The CERES global data set derived from polar satellites and geostationary satellites had significantly higher deviations than all the other satellite derived data sets, at times even higher than the NWP data. An essential reason is its coarse resolution (1°), which results in pixels (cells) that are much larger than those of the other databases. The deviations of its DNI predictions are even substantially more pronounced than that of other models.

The deviation metrics of data sets based mainly on geostationary satellite imagery are closer to each other than to the ACCESS G3 or the CERES data sets. This difference is more pronounced for the standard deviation than for the bias, and also more obvious for DNI than for GHI. The lowest average deviation metrics are often achieved by a single data set (Solargis). It also performs best at many individual stations, but the best data set for a specific station is at times provided by another model.

From a methodological standpoint, this benchmark underlined the importance of the reference data quality. Without a stringent quality control procedure, no real validation can be done, with the risk of obtaining invalid results.

The large volume of benchmark results obtained here have been organized in tables, figures, and supplementary material in a data Annex. Users are encouraged to consult this dataset to to identify the most appropriate data set(s) for a specific application or over a specific region.

Future work will further analyse the expected positive impact of various post-processing methods (known as “site adaptation”) on the data evaluated in this benchmark. Furthermore, it is recommended that a similar benchmark be conducted for more reference stations, including sites over regions that are so far not covered well or at all. Moreover, future work should also involve new modelled data sets and updated versions of the current data sets. Such work is currently planned as part of the activities of IEA PVPS Task 16 on solar resource data and forecasting. The validation of global tilted irradiance predictions is desirable too, but both modelled data sets and high-quality reference data are much less common.

DESCRIPTION OF THE DATA FILE ANNEX

The data annex (DOI: 10.5281/zenodo.7867002) includes:

- StationList.xlsx: list of all stations including their coordinates, climate zone, station code, continent, altitude AMSL, data source, number of available test data sets, station type (Tier-1 or Tier-2), and available calibration record.
- Result tables in folder “ResultTables”: Folders “climate_zones” and “continents” contain the tables described in Section 5.3. The filenames are “Component_metric_in_subgroup.html” with “component” DNI or GHI, “metric” describing the metric (see Table 3), and “subgroup” describing the continent or climate zone.
- World maps: The folder “Resultmaps” contains world maps of the metrics described in Section 5.2. Either four or three metrics, depending on map, are included in each pdf. A legend describing the meaning of the point size is also included.
- Scatter plots of test vs. reference irradiance: The folder “Scatterplots” contains two folders, “DNI” and “GHI”, for the two investigated components. Three subfolders are also contained in these two folders:
    - The subfolders “plotsPerSiteYear” contain plots named “scatOverviewCOMPONENT_SITEYYYY.png”, where “COMPONENT” is either DNI or GHI, SITE is the three-letter site abbreviation, and YYYY is the evaluated year. The png plots include the scatterplots for all test data sets evaluated for the case specified by the filename.
    - The subfolders “plotsPerTestdataProvider” contain plots named “scatOverviewTESTDATASET_COMPONENTYYYY.png”, where “TESTDATASET” describes the test data set, “COMPONENT” is either DNI or GHI, and YYYY is the evaluated year. The png plots include the scatterplots for all sites evaluated for the case specified by the filename.
    - The subfolders “plotsPerTestdataProviderSamePosPerStat” contain the same scatterplots as “plotsPerTestdataProvider”, but using a slightly different visualization method. Here, the position of each scatterplot for a given site within the plot is always the same. Although this yields many empty subplots and small scatterplots, it can be helpful to rapidly browse through the plots if only one or a few stations are of interest.

REFERENCES

Amillo, A. G., L. Ntsangwane, T. Huld, and J. Trentmann. 2018. "Comparison of Satellite-Retrieved High-Resolution Solar Radiation Datasets for South Africa." *Journal of Energy in Southern Africa* 29 (2): 63–76. https://doi.org/10.17159/2413-3051/2018/V29I2A3376.

Andreas, A., and T. Stoffel. 1981. "NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data)." *NREL Report No. DA-5500-56488*. https://doi.org/10.5439/1052221.

———. 2006. "University of Nevada (UNLV): Las Vegas, Nevada (Data)." *NREL Report No. DA-5500-56509*. https://doi.org/10.5439/1052548.

Andreas, A., and S. Wilcox. 2010. "Observed Atmospheric and Solar Information System (OASIS); Tucson, Arizona (Data)." *NREL Report No. DA-5500-56494*. https://doi.org/10.5439/1052226.

———. 2012. "Solar Resource & Meteorological Assessment Project (SOLRMAP): Rotating Shadowband Radiometer (RSR); Los Angeles, California (Data)." *NREL Report No. DA-5500-56502*. https://doi.org/10.5439/1052230.

Brooks, M.J., S. du Clou, J.L. van Niekerk, P. Gauche, C. Leonard, M.J. Mouzouris, A.J. Meyer, N. van der Westhuizen, E.E. van Dyk, and F Vorster. 2015. "SAURAN: A New Resource for Solar Radiometric Data in Southern Africa." *Journal of Energy in Southern Africa* 26: 2–10.

Driemel, A., J. Augustine, K. Behrens, S. Colle, C. Cox, E. Cuevas-Agulló, F. M. Denn, et al. 2018. "Baseline Surface Radiation Network (BSRN): Structure and Data Description (1992–2017)." *Earth Syst. Sci. Data* 10: 1491–1501. https://doi.org/10.5194/essd-10-1491-2018.

Espinar, B., L. Wald, P. Blanc, C. Hoyer-Klick, M. Schroedter Homscheidt, and T. Wanderer. 2011. "Project ENDORSE - Excerpt of the Report on the Harmonization and Qualification of Meteorological Data:Procedures for Quality Check of Meteorological Data." https://hal-mines-paristech.archives-ouvertes.fr/hal-01493608.

Forstinger, A., Y.-M. Saint-Drenan, S. Wilbert, A. Jensen, B. Kraas, C. Fernández Peruchena, C. Gueymard, D. Ronzio, D. Yang, E. Collino, J. Polo Martinez, J. Ruiz-Arias, N. Hanrieder, and P. Blanc. 2021. "IEA-PVPS Task-16 Reference Solar Measurements." Lionel Menard. https://doi.org/10.23646/3491b1a6-e32d-4b34-9dbb-ee0affe49e36.

Forstinger, A., S. Wilbert, B. Kraas, C. Gueymard, D. Ronzio, D. Yang, E. Collino, J. Polo Martinez, J. Ruiz-Arias, N. Hanrieder, P. Blanc, and Y.-M. Saint-Drenan. 2021. "Expert Quality Control of Solar Radiation Ground Data Sets." *ISES Solar World Conference*, no. October.

Geuder, N., F. Wolfertstetter, S. Wilbert, D. Schüler, R. Affolter, B. Kraas, E. Lüpfert, and B. Espinar. 2015. "Screening and Flagging of Solar Irradiation and Ancillary Meteorological Data." *Energy Procedia* 69 (May): 1989–98. https://doi.org/10.1016/J.EGYPRO.2015.03.205.

Gschwind, B., L. Wald, P. Blanc, M. Lefèvre, M. Schroedter-Homscheidt, and A. Arola. 2019. "Improving the McClear Model Estimating the Downwelling Solar Radiation at Ground Level in Cloud-Free Conditions - McClear-V3." *Meteorologische Zeitschrift* 28 (2): 147–63. https://doi.org/10.1127/metz/2019/0946.

Gueymard, C. 2014. “A Review of Validation Methodologies and Statistical Performance Indicators for Modeled Solar Radiation Data: Towards a Better Bankability of Solar Projects.” *Renewable and Sustainable Energy Reviews* 39 (November): 1024–34. https://doi.org/10.1016/J.RSER.2014.07.117.

———. 2017. “Cloud and Albedo Enhancement Impacts on Solar Irradiance Using High-Frequency Measurements from Thermopile and Photodiode Radiometers. Part 1: Impacts on Global Horizontal Irradiance.” *Solar Energy* 153 (September): 755–65. https://doi.org/10.1016/J.SOLENER.2017.05.004.

———. 2018. “A Reevaluation of the Solar Constant Based on a 42-Year Total Solar Irradiance Time Series and a Reconciliation of Spaceborne Observations.” *Solar Energy* 168 (July): 2–9. https://doi.org/10.1016/J.SOLENER.2018.04.001.

Gueymard, C., J. Bright, X. Sun, J. Augustine, S. Baika, L. Brunier, S. Colle, et al. 2022. “BSRN Data Set for IEA-PVPS Task-16 Activity 1.4 Quality Control.” PANGEA. https://doi.org/https://doi.pangaea.de/10.1594/PANGAEA.939988.

Hoyer-Klick, C., H.G. Beyer, D. Dumortier, M. Schroedter-Homscheidt, L. Wald, M. Martinoli, C. Schilings, et al. 2008. “Management and Exploitation of Solar Resource Knowledge.” In *EUROSUN 2008, 1st International Conference on Solar Heating, Cooling and Buildings, Lisbon, Portugal*.

Hoyer-Klick, C., H. Beyer, D. Dumortier, L. Wald, C. Schillings, B. Gschwind, L. Menard, et al. 2009. “MESoR – MANAGEMENT AND EXPLOITATION OF SOLAR RESOURCE KNOWLEDGE.” In *SolarPACES 2009, Berlin : Germany (2009)*.

Ineichen, P. 2014. “Long Term Satellite Global, Beam and Diffuse Irradiance Validation.” *Energy Procedia* 48 (January): 1586–96. https://doi.org/10.1016/J.EGYPRO.2014.02.179.

Journée, M., and C. Bertrand. 2011. “Quality Control of Solar Radiation Data within the RMIB Solar Measurements Network.” *Solar Energy* 85 (1): 72–86. https://doi.org/10.1016/j.solener.2010.10.021.

Lefèvre, M., A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, et al. 2013. “McClear: A New Model Estimating Downwelling Solar Radiation at Ground Level in Clear-Sky Conditions.” *Atmospheric Measurement Techniques* 6 (9): 2403–18. https://doi.org/10.5194/amt-6-2403-2013.

Long, C. N., and E. G. Dutton. 2002. “BSRN Global Network Recommended QCtests, V2.0.” *Baseline Surface Radiation Network*. https://bsrn.awi.de/fileadmin/user_upload/bsrn.awi.de/Publications/BSRN_recommended_QC_tests_V2.pdf.

Long, C N, and Y. Shi. 2008. “An Automated Quality Assessment and Control Algorithm for Surface Radiation Measurements.” *The Open Atmospheric Science Journal* 2: 23–37.

Marchand, M., A. Ghennioui, E. Wey, and L. Wald. 2018. “Comparison of Several Satellite-Derived Databases of Surface Solar Radiation against Ground Measurement in Morocco.” *Advances in Science and Research* 15 (April): 21–29. https://doi.org/10.5194/ASR-15-21-2018.

Maxwell, E., S. Wilcox, and M. Rymes. 1993. “Users Manual for SERI QC Software, Assessing the Quality of Solar Radiation Data.” *Solar Energy Research Institute, Golden, CO*.

Nollas, F., G. Salazar, and C. Gueymard. 2023. “Quality Control Procedure for 1-Minute Pyranometric Measurements of Global and Shadowband-Based Diffuse Solar Irradiance.” *Renewable Energy* 202 (January): 40–55. https://doi.org/10.1016/J.RENENE.2022.11.056.

Polo, J., S. Wilbert, J. A. Ruiz-Arias, R. Meyer, C. Gueymard, M. Súri, L. Martín, et al. 2016. "Preliminary Survey on Site-Adaptation Techniques for Satellite-Derived and Reanalysis Solar Radiation Datasets." *Solar Energy* 132 (July): 25–37. https://doi.org/10.1016/J.SOLENER.2016.03.001.

Qu, Z., A. Oumbe, P. Blanc, B. Espinar, G. Gesell, B. Gschwind, L. Klüser, et al. 2017. "Fast Radiative Transfer Parameterisation for Assessing the Surface Solar Irradiance: The Heliosat-4 Method." *Meteorologische Zeitschrift* 26 (1): 33–57. https://doi.org/10.1127/metz/2016/0781.

Ramos, J., and A. Andreas. 2011. "University of Texas Panamerican (UTPA): Solar Radiation Lab (SRL); Edinburg, Texas (Data)." *NREL Report No. DA-5500-56514*. https://doi.org/10.5439/1052555.

Salazar, G., C. Gueymard, J. Bezerra Galdino, O. de Castro Vilela, and N. Fraidenraicha. 2020. "Solar Irradiance Time Series Derived from High-Quality Measurements, Satellite-Based Models, and Reanalyses at a near-Equatorial Site in Brazil." *Renewable and Sustainable Energy Reviews* 117 (109478).

Sengupta, M., A. Habte, S. Wilbert, C. Gueymard, and J. Remund. 2021. "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition." National Renewable Energy Laboratory. https://www.nrel.gov/docs/fy21osti/77635.pdf.

Šúri, M, J. Remund, T Cebecauer, D Dumortier, L. Wald, T Huld, and P. Blanc. 2008. "First Steps in the Cross-Comparison of Solar Resource Spatial Products in Europe." In *Eurosun 2008*, 7–10. http://hal-ensmp.archives-ouvertes.fr/hal-00587966%0Ahttp://hal.archives-ouvertes.fr/hal-00587966/.

Vignola, F., and A. Andreas. 2013. "University of Oregon: GPS-Based Precipitable Water Vapor (Data)." *NREL Report No. DA-5500-64452*.

Yang, D., and C. Gueymard. 2021. "Probabilistic Post-Processing of Gridded Atmospheric Variables and Its Application to Site Adaptation of Shortwave Solar Radiation." *Solar Energy* 225 (September): 427–43. https://doi.org/10.1016/J.SOLENER.2021.05.050.

ISBN 978-3-907281-44-4