Technology Collaboration Programme
by iea

International Energy Agency
Photovoltaic Power Systems Programme

Task 16: Solar resource for high penetration and large scale applications

PVPS

Framework for benchmarking of GHI gap-filling methods

2023

Report IEA-PVPS T16-03:2023

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 6.000 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 TCP's 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 Cop- per 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 (PVPS) 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 forecasts and the resources assessments. To reach this main goal the Task has the following objectives:

- Lowering uncertainty of satellite retrievals and Numerical Weather Prediction (NWP) models for solar resource assessments and nowcasting.
- Define best practices for data fusion of ground, satellite and NWP data (re-analysis) to produce improved datasets, e.g. time series or Typical Meteorological Year (TMY.
- Develop enhanced analysis of long-term inter-annual variability and trends in the solar resource.
- Develop and compare methods for estimating the spectral and angular distributions of solar radiation (clear and all-sky conditions), describing the spatial and temporal variabilities of the solar resource, modelling point to area forecasts, probabilistic and variability forecasting.
- Contribute to or setup international benchmark for datasets and for forecast evaluation.

As the scope of Task 16 is also directly linked to Concentrating Solar Power and solar thermal installations the collaboration with the two TCPs : SolarPACES and SHC.

Authors

- **Main Content:** Philippe Blanc, Rodrigo Amaro e Silva
- **Editor:** Jan Remund

INTERNATIONAL ENERGY AGENCY

PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

Framework for Benchmarking of GHI Gap-Filling Methods

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

Report IEA-PVPS T16-03:2023
June 2023

ISBN 978-3-907281-37-6

COVER PICTURE

Jeff Fenton, Flickr

ISBN 978-3-907281-37-6: Benchmarking of GHI gap-filling methods

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

This report received valuable contributions from several IEA-PVPS Task 16 members and other international experts. Many thanks to: Javier Lopez Lorente (previously at Queen's Univ. of Belfast and the University of Cyprus, now at DNV) and Anton Driesse (PV Performance Lab).

Additionally, it is important to acknowledge the contributions from the two reviewers of this report – Ricardo Aguiar (Directorate-General for Energy and Geology of Portugal) and Yves Poissant (Natural Resources Canada) – which greatly improved its overall quality and potential impactfulness.

LIST OF ABBREVIATIONS

IEA	International Energy Agency
GF	Gap-filling
QC	Quality-check
GHI	Global Horizontal Irradiance
TMY	Typical Meteorological Year
ML	Machine Learning
MSG	Meteosat Second Generation
SCADA	Supervisory Control And Data Acquisition
DSG	Daily Sum with intra-day Gaps
kNN	k-Nearest-Neighbor

EXECUTIVE SUMMARY

An evermore accelerated deployment of photovoltaic (PV) capacity is expected worldwide and in-situ solar irradiance time series play a decisive in supporting such growth: not only they represent the foundation of solar resource assessment and forecasting, but they also drive prospective PV yield studies, can be used as a calibration reference when using satellite data, evaluating PV systems' performance, or even developing forecasting algorithms.

However, such datasets inevitably have gaps (i.e., periods with missing data) – as a result of defaults during data-logging, sensor failures, among others, or from Quality Check (QC) procedures – that can compromise their applicability and value. An additional issue is that data gaps can be further enlarged when computing temporal aggregations, notably for intra-daily to daily, daily to monthly and yearly averages, thus further degrading the dataset.

This has raised the need for gap-filling (GF) methods that can post-process either static historical datasets or more dynamic real-time data streams. Each case is characterized by different constraints, such as the access to data that follows a given data gap or the acceptable time lag for generating the replacement synthetic data.

And while, naturally, a given gap-filling method can be tested for a given location, such analyses are done in a very context-specific manner. Thus, this report aims to propose a GF benchmark framework, as well as evaluate a set of possible baseline algorithms for:

- The GF of intra-hourly time series of global horizontal irradiance. The results are exposed in this report only for time step of 15-min.
- The GF of the corresponding daily sums of irradiation in the presence of with intra-day gaps (DSG*, daily sums with gaps*) using two different approaches: (1) gap-filling the intra-day time series with the previous methods and (2) using directly the incomplete data with methods which are less sensitive to gaps.

It is also important to mention that while the focus is exclusively on:

- global horizontal irradiance (GHI), this report paves the way for potential future studies addressing global tilted irradiance (GTI) and PV time series,
- baseline gap-filling methods, this same benchmark framework can be used by anyone to evaluate more complex, e.g., machine-learning (ML-)based, approaches.

This report is the result of expert discussions during specific sessions of the subtask-2 activity 2.1 of the Task 16 of PVPS and during the dedicated workshop organized during the ICEM 2019 conference: "*Workshop on best practices for automatic and expert-based data quality control methods and for gap filling methods*".

1 MOTIVATION & CONTEXT

Over the last decades, high expectations have been placed in photovoltaics (PV) as an important element of any designed pathways that aims to address climate change (Pathak et al., 2022) and promote fairer and more equitable energy systems (Campos et al., 2022).

This has driven an accelerated deployment of PV capacity, the design of innovative applications (e.g., floating PV, agriPV), but also the valorization of Earth Observation data which is of essence to properly understand and exploit the potential of this technology. Among such data, solar irradiance in-situ measurements are of particular importance for PV, enabling the modelling of predictive and expected operational yields for standard PV configurations as well as the simulation of innovative concepts. From a solar resource assessment and forecasting perspective, in-situ data represent high-quality local information that can also be used as a calibration reference for satellite- and numerical weather prediction-based models.

While larger-scale PV projects are expected to have in place co-located radiation sensor (or reference solar cells) that are installed with the same tilt and orientation as the PV modules, the most generic setup of solar irradiance data is that of a horizontally placed device measuring global horizontal irradiance (GHI). These data are particularly relevant, being used to calibrate satellite and numerical weather prediction products and play a great role in the prospective and operational yield analysis and forecasting of any PV project (using transposition algorithms that transform horizontal data to any intended tilt and orientation with a reasonable level of accuracy). Additionally, for smaller size installations, these are often the only solar resource data available to implement performance analysis protocols.

However, such datasets will inevitably have gaps (i.e., periods with missing data) – as a result of defaults during data-logging, sensor failures, among others, or from Quality Check (QC) procedures – that can compromise their applicability and value. An additional issue is that data gaps can be further enlarged when computing temporal aggregations, notably for intra-daily to daily, daily to monthly and yearly averages, thus further degrading the dataset.

2 PROPOSING A BENCHMARKING FRAMEWORK FOR GHI GF AND DSG METHODS

The main goal for this report is to propose a “generic” framework meant to benchmark gap filling (GF) methods, meant to fill intra-day periods of missing global horizontal irradiance (GHI) data, as well as methods aimed at deriving daily sums from such incomplete time series – here designated as Daily Sum with intra-day Gaps (DSG). Additionally, since the GF methods generate complete time series, the corresponding daily averages will also be assessed from a DSG perspective.

This benchmark framework consists in a Monte-Carlo analysis where realistic GHI data gaps are generated and imputed to high-quality complete GHI daily profiles, considering not only the amount of missing data but also how it is distributed throughout a day. This leads to a setup where we have pairs of GHI time series - complete and incomplete daily profiles – which allow assessing the performance of GF and DSG methods in a robust manner: the incomplete time series are post-processed and compared with either the original daily profile (for GF methods) or the corresponding daily irradiation (for DSG).

In this report, the benchmark is only considering 15-min time resolution, but other time resolutions can be considered if they are multiples of 1-min (e.g., 5-min, 10-min, 1-h, etc.). Additionally, the framework is designed in a way that allows to cross-compare already existing approaches, but also to test new approaches notably based on machine learning (ML) or derived from solar forecasting approaches, for example. Beyond this cross-comparison purpose, this benchmark can be used for recommendations about limits of data gaps, for a given precision for downstream solar processing chains.

2.1 Setting up a reference 1-min GHI database

To be able to implement this benchmark framework, an essential asset is an extensive database of GHI measurements, which contains:

- High-quality measurements, which increases the availability of complete daily profiles and ensures a robust and credible performance assessment of the tested methods,
- Measurements from different latitudes and climates, to ensure a diversity of GHI time series in terms of characteristics such as sunshine duration, intra-day variability but also missing data patterns. The missing data patterns are about when, how long and at what frequency data gaps occur.

Thus, we have considered 1-min GHI measurements from 16 Baseline Surface Radiation Network (BSRN) stations[1] with at least 3 years of data, over the period from 2004 to 2018. An additional criterion when choosing these 16 stations was the availability of concomitant satellite-based GHI estimates from the HelioClim-3v4 database[2,3], since this source of data can be used by a large variety of GF and DSG approaches. In practice, it means that every site is in the field of view (FoV) of the Meteosat Second Generation (MSG) satellite, from which HelioClim-3v4 derives its estimates.

[1] https://bsrn.awi.de/
[2] http://www.soda-pro.com/web-services/radiation/helioclim-3-archives/info
[3] https://www.soda-pro.com/help/helioclim/helioclim-3-overview

The location and altitude of the selected locations is shown in the figure below (c.f. Annex 1 to find exact values).

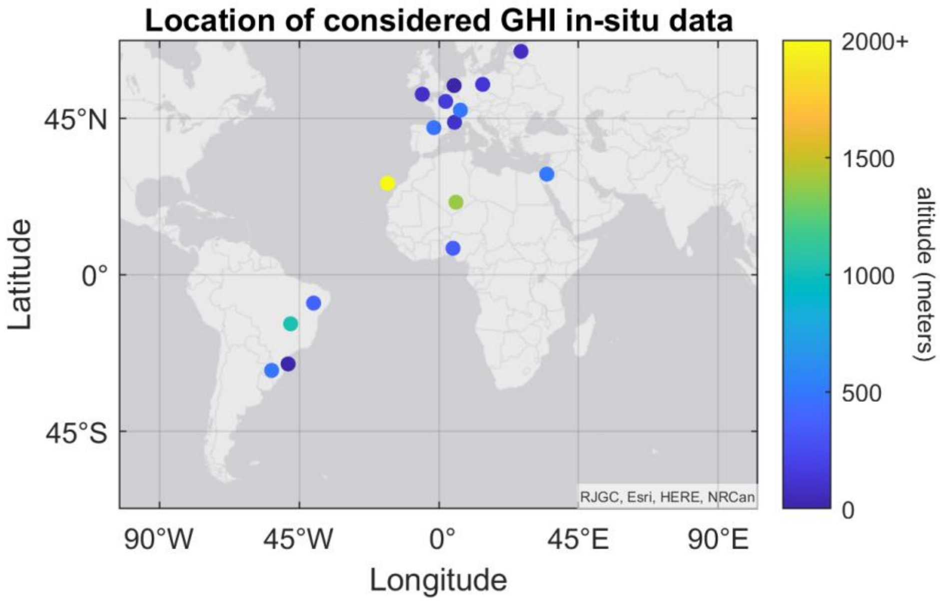

Figure 1: Location and altitude of the locations selected for this report.

The collected data were submitted to automatic QC tests as described by Roesch et al., which you can find in Annex 2. Only totally unflagged 1-min GHI data were kept, meaning that not passing the QC test is considered equivalent to being missing from the dataset.

Figure 2 illustrates the distribution of the percentage of missing 1-min data from the daily profiles for the set of BSRN stations. It corresponds to the histogram, represented in log-scales in both axes, expressing the number of days per bins of percentage of corresponding number of missing 1-min data. It is to be noted that the maximum percentage of missing 1-min data within a day is 99.38 %.

Then, the dataset is complemented with supplementary variables:

- 1-min interpolated GHI from the HelioClim-3v4 database[4] (SAT),
- 1-min top of atmosphere horizontal irradiance (TOA),
- 1-min global horizontal irradiance under clear-sky conditions (CLS) from ESRA (Rigolier et al., 2000),
- Solar Zenith Angle (SZA) and Solar Azimuth Angle (SAA) computed from SG2 (Blanc et al., 2012).

These variables are meant to potentially be used by the GF and DSG methods tested here.

[4] The original estimates have a 15-min resolution http://www.soda-pro.com/web-services/radiation/helioclim-3-archives/info

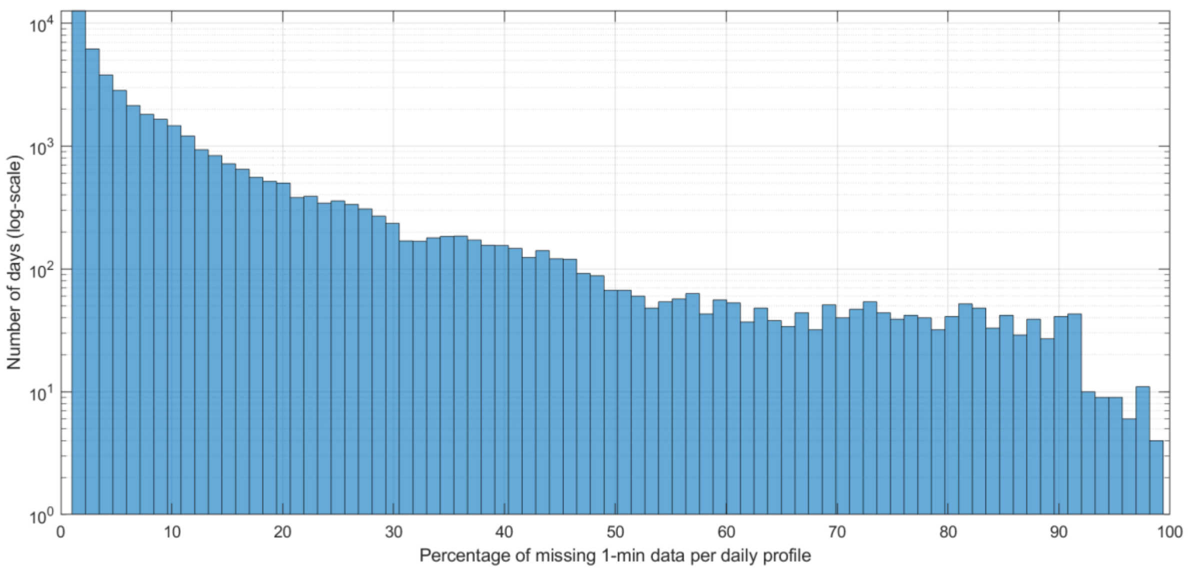

Figure 2: Number of daily profiles (log-scale) for a given daily percentage of missing values.

The following table summarizes the compiled database:

Number of BSRN stations	16 in the field of view of MeteoSat 2nd Generation (MSG)
Temporal resolution	1-min
Number of daily profiles of 1-min GHI	56281
Number of records during daytime (solar elevation greater than 0°)	41067747 min
Number of valid records during daytime (solar elevation greater than 0°)	39784911 min (96 %)
Number of complete daily profiles	28616 (51 %)
Number of incomplete daily profiles	27665 (49 %)
Number of complementary variables	5 (SAT, TOA, CLS, SZA, SAA)

Table 1: Summarized description of the compiled data

This reference database is available online as a NetCDF file db_reference_1min_v0.nc, as well as a support document describing the corresponding CDL structure[5].

[5] https://cloud.mines-paristech.fr/index.php/s/9obUl2ZVGZTUU12

2.2 The reasoning behind the Monte-Carlo based framework

As mentioned before, the main idea behind the proposed framework is to use the prepared reference database with a Monte-Carlo approach to randomly generate many pairs of daily profiles of GHI which is illustrated in Figure 3. Each pair consists in a complete (original) and an incomplete time series with synthetic but realistic data gaps. All these pairs represent realistic examples of daily profiles, coming from a large variety of climates and showing a large variety of missing patterns (both in length, frequency, and distribution throughout a day), which should be in a sense related to the environmental conditions of the instrumental setup.

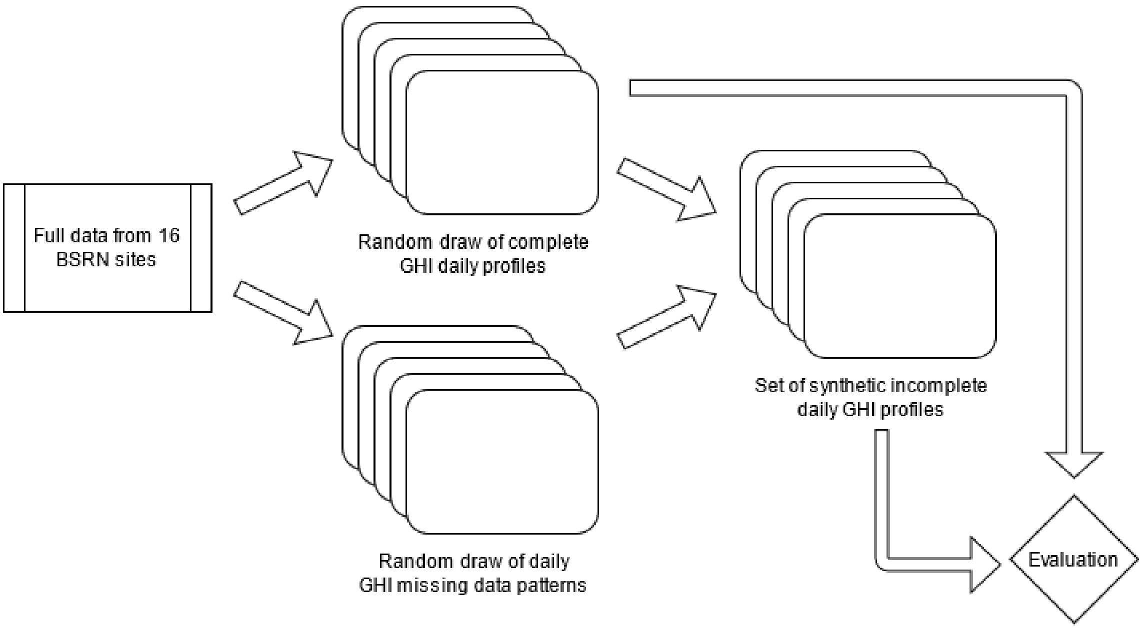

Figure 3: Schematic overview of the implemented Monte-Carlo based benchmark framework. Here, the focus is placed on the generation of synthetic but realistic data gaps and corresponding complete time series.

More precisely, this approach consists in the following steps:

- Step 1: random draws of complete daily profiles of 1-min GHI from BSRN from the reference database, for any site and any day. This ensures a considerably diverse dataset, ensuring an extensive assessment.
- Step 2: for each selected complete daily profile, random draw a missing pattern from incomplete daily profiles with similar daytime duration (±10 %, measured as the number of 1-min values where SZA<90°) and possibly from different sites.
- Step 3: resample the complete daily profile of GHI (W/m^2) for the other resolutions of interest (10-min, 15-min and 1-h) and the corresponding daily sum of irradiation of reference (Wh/m^2).
- Step 4: generate the incomplete daily profile of GHI (W/m^2) at the final time step (1-min, 10-min, 15-min and 1-h) with the corresponding missing pattern.
- Step 5: extract the corresponding additional data: CLS, SAT, SAA, SZA with the corresponding time steps. It is important to note that while the first two are effectively used in the gap-filling methods tested in this report, the others are added for future potential gap-filling approaches.

To further illustrate this, Figure 4 shows an example of actual complete and incomplete daily profiles, as well as the result from transferring the missing data pattern from one to the other.

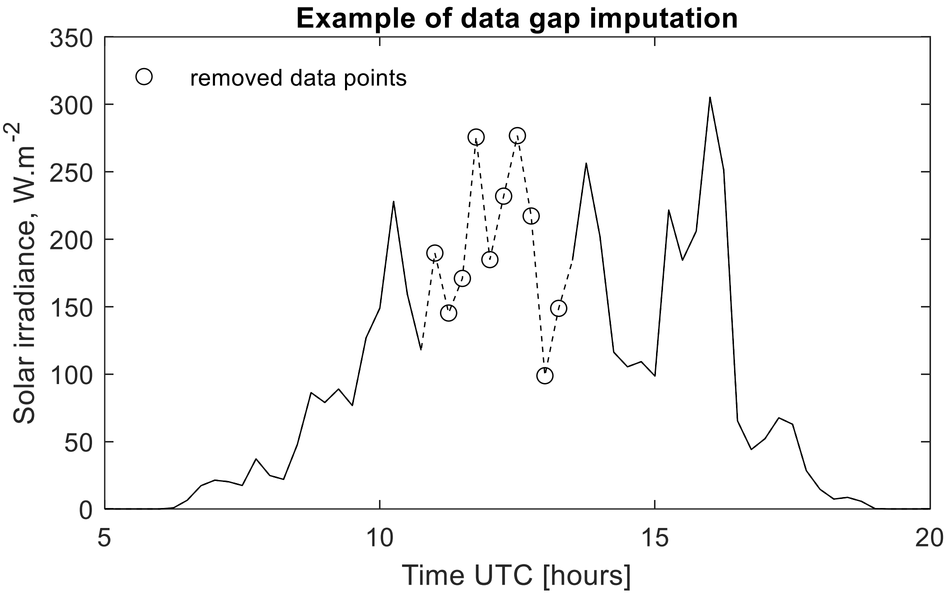

Figure 4: Example of the imputation of a data gap to a complete daily GHI profile. series. As mentioned in the text, the gap is extracted a randomly drawn incomplete daily profile, possibly even from a different location, that shows a similar daytime duration.

It is important to consider that a GF or DSG method may rely on a ML approach which will require an optimization of the associated hyperparameters. Thus, following common practice when optimizing a set of hyperparameters, this Monte-Carlo based procedure can be applied three times to generate three separate datasets:

- TRAIN: One dataset for the training of ML approaches
- VALIDATION: One dataset for the validation of the ML approaches to determine, if any, the optimal value for the hyperparameter set (avoiding overfitting the model)
- TEST: One dataset used exclusively to evaluate the performance of GF and DSG methods.

In order to have three independent datasets, we have randomly separated the years of data into three groups with a respective ratio of 30 % / 15 % / 55 % for the TRAIN, VALIDATION, and TEST subsets (as seen in the table below). Although this partitioning is uncommon, the availability of such long datasets allows us to implement this demanding setup - where more than half of the data is used for testing- which increases the robustness of the obtained results.

Of course, this specific split into TRAIN, VALIDATION, TEST subsets can be modified for the purpose of other benchmarks based on the same database.

Datasets	Years	# of 1-min daily profile pairs[6]
TRAIN	2005, 2010, 2013, 2016, 2018	15 454
VALIDATION	2009, 2012	7 622
TEST	2004, 2006, 2007, 2008, 2011, 2014, 2015, 2017	26 924
	TOTAL	**50 000**

Table 2: Data structuring for the implementation of the benchmark framework

These three Monte-Carlo based datasets were made available online in a NetCDF file for different time resolutions[7]. The structure of the corresponding NetCDF file is described in CDL, which, due to its specificity and technical level, is made available only online in the same web address.

[6] Where one pair of corresponds to a complete daily profile and that same data after imputation of a plausible data gap
[7] https://cloud.mines-paristech.fr/index.php/s/9obUl2ZVGZTUU12

3 SELECTED METHODS FOR THE BENCHMARK

In this chapter, a selection of standard GF and DSG solutions is evaluated using the benchmark framework proposed in Chapter 2. To recap, this framework was designed to provide a comprehensive evaluation of GF and DSG algorithms, by taking into account real measurements from diverse climates to generate synthetic, but plausible (and data-driven), data gaps. The goal of this is twofold: first, to increase the likelihood that the proposed framework is used by readers, by providing results which are actionable, replicable, and, thus, verifiable; then, to propose, and evaluate, a set of models which can be seen as baselines against which we can compare the performance of more complex approaches.

3.1 Symbol glossary

For the brief description of the different methods, the following symbols are used:

- $G(k)$: a given daily profile of global horizontal irradiance averages (W/m^2) with a given time resolution dt expressed in decimal hours ($dt = \frac{1}{60}, \frac{1}{6}, \frac{1}{4}, 1$ respectively for 1-min, 10-min, 15-min and 1-h resolutions),
- $G_{cls}(k)$: the daily profile of global horizontal irradiance under clear-sky condition, obtained using the ESRA model, for the same day as $G(k)$,
- $G_{sat}(k)$: the daily profile of satellite-based global horizontal irradiance, obtained from the HC3v4 database, for the same day as $G(k)$,
- D: set of N_D indices k corresponding to the daytime period within a given day (i.e., a daytime mask)
- M: set of N_M indices k corresponding to the daytime period within a given day where data are missing (missing mask),
- V: set of N_V indices k corresponding to daytime period within the day where data are available (valid mask),
- $Kc(k)$: clear-sky index, which estimates the atmosphere transmissivity to GHI relative to a clear-sky atmosphere.

$$Kc(k) = \begin{cases} \dfrac{G(k)}{G_{cls}(k)} \; if \; k \in V \\ NaN \; otherwise \end{cases} \tag{1}$$

- $\langle X \rangle$ corresponds to the average of available elements of X, which includes the nighttime and disregards values which are missing or have failed the QC test

$$\langle X \rangle = \frac{1}{N_V} \sum_{k \in V} X(k) \tag{2}$$

- H : the effective daily irradiation for a complete G(k) (i.e., $N_M = 0$):

$$H = \sum_{k \in D} G(k) dt = 24 \, \langle G \rangle \tag{3}$$

3.2 Descriptions of the GF methods

The following method descriptions are very brief and do not represent the corresponding efficient code implementation.

3.2.1 GF0: Assuming the nearest available clearness index within the day

Assuming $Kc(n(k))$, i.e., the nearest Kc value available, and multiplying it by the clear-sky expectation G_{cls} for the instants k where data are missing, ensures that the generated data consider the GHI daily seasonality (i.e., the apparent movement of the sun). This is quite similar to a smart persistence model used in solar forecasting (described, for example, in Liu et al.), with the exception that for gap filling we can consider records stored after the period with missing data.

This method needs to have at least one valid data within the current day.

More formally, for any $k \in M$, i.e., for all missing records, let us note n(k) as the nearest element of V with respect to k:

$$n(k) = \underset{l \in V}{\operatorname{argmin}} |l - k| \tag{4}$$

Thus,

$$G_{GF0}(k) = Kc(n(k)) \; G_{cls}(k) \tag{5}$$

This method is considered in this work as a baseline reference for performance, since it is quite easy to implement, and it does not require any additional data source.

3.2.2 GF1: Linear interpolation of available clearness index within the day

Starting from the same premise as *GF0*, where the nearest available Kc values are considered, in GF1 the missing data are derived from a weighted linear interpolation of the last available value before a data gap starts and the first available one after the gap ends, $G(f(k))$ and $G\big(c(k)\big)$ respectively. It is important to note that the contribution of each of these values to the estimation of a given missing value is weighted by the weighting factor $\alpha(k)$ which quantifies the relative distance in time between each of the values and the missing record.

This method needs to have at least two valid data points within the current day.

For any $k \in M$, let's note $f(k)$ and $c(k)$:

$$f(k) = \max\{l \in V \;/\; l < k\} \tag{6}$$

$$c(k) = \min\{l \in V \;/\; l > k\} \tag{7}$$

Thus,

$$G_{GF1}(k) = G_{cls}(k)\left(\big(1 - \alpha(k)\big)Kc\big(f(k)\big) + \alpha(k)Kc\big(c(k)\big)\right) \tag{8}$$

where:

$$\alpha(k) = \frac{k - f(k)}{c(k) - f(k)} \tag{9}$$

3.2.3 GF2: k-Nearest Neighbor-based approach using prior data

In the context of this report, the k-Nearest-Neighbor (kNN)[8,9] can be seen as a simple ML model which estimates the GHI for a period of missing data of a given station as the average, point by point, of the N most similar days. In practice, similarity is quantified as the averaged Euclidean distance[10] between the Kc values from the available records of the day in question and the corresponding values from other days from the training (TRAIN) dataset that belong to the same calendar month.

This method needs to have at least one valid data point within the current day.

Let $T = \{G(k, d)\}$ be a training dataset of N daily complete profiles of GHI, from which we compute the corresponding clear-sky index values $\{Kc(k, d)\}$.

For a given incomplete daily profile G(k), the Euclidean distance is computed for each of the N elements corresponding to the same calendar month (to have approximately the same daylight periods):

$$D_T(d) = \begin{cases} \langle \left(Kc - Kc(.,d)\right)^2 \rangle & \mathit{if\ day\ d\ correspond\ to\ the\ same\ calendar\ month} \\ +\infty & \mathit{otherwise} \end{cases} \quad (10)$$

These distances are then sorted in ascendant order $(d_1, \ldots, d_N)$:

$$D_T(d_1) \le D_T(d_2) \le \cdots \le D_T(d_N) \quad (11)$$

Then, for any $k \in M$, the kNN method estimates the missing value as the simple average of the nearest L neighbors, being L the only parameter of this method,

$$G_{kNN}(k) = G_{cls}(k) \frac{1}{L} \sum_{l=1}^{L} Kc(k, d_l) \quad (12)$$

It is interesting to note that the neighbor search could be extended to also consider historical GHI records from complementary data sources (e.g., a nearby station or concomitant satellite-based estimates, such as the ones used in GF4).

[8] https://medium.com/analytics-vidhya/k-neighbors-regression-analysis-in-python-61532d56d8e4
[9] https://towardsdatascience.com/k-nearest-neighbours-knn-algorithm-common-questions-and-python-implementation-14377e45b738
[10] Other distance metrics could have been considered, c.f. references [8, 9]

3.2.4 GF3: Kernel regression using prior data

A Kernel-based regression is capable of modelling non-linear patterns as a weighted average of the available observations[11,12]. Here, the weights are derived from a kernel function (often a probability density function from a given distribution): for a given new point we want to estimate, the weight of each of the observations is inversely proportional to its distance to the new point. In a sense, the kernel serves as a similarity metric. It can also be said that kernel regression is close to a non-parametric approach, since only the weighting function may require any parameter(s) to be defined.

Here, the weights are attributed based on a Gaussian kernel and on the same distance metric which is used for GF2. This can be seen as a weighted version of the previous GF method with an additional parameter σ for the kernel.

Thus, and using the same notation as GF2, for any $k \in M$, and σ^2 the only parameter for the kernel-based method,

$$G_{GF2}(k) = G_{cls}(k) \frac{\sum_{d=1}^{N} Kc(k,d)\, K\left(\frac{D_T(d)}{\sigma^2}\right)}{\sum_{d=1}^{N} K\left(\frac{D_T(d)}{\sigma^2}\right)} \tag{14}$$

where the kernel K is gaussian:

$$K(x) = e^{-x^2/2} \tag{15}$$

3.2.5 GF4 : Assuming concomitant satellite-based GHI estimation

Being likely the easiest GF approach to implement, GF4 replaces the identified GHI gaps with the concomitant satellite-based GHI estimates (in this report, HelioClim3v4).

Thus, for any $k \in M$

$$G_{GF4}(k) = G_{sat}(k) \tag{16}$$

When using this approach, it is important to be aware of the limitations of satellite-based methods[13]. Additionally, it is important to note that although the satellite data used in this report are made available, HelioClim3v4 is a paid product. However, openly available databases exist, such as the CAMS Radiation service[14] (from Copernicus Atmosphere Monitoring Services) or NSRDB[15], the National Solar Radiation Database from the National Renewable Energy Laboratory (NREL).

---

[11] https://towardsdatascience.com/kernel-regression-made-easy-to-understand-86caf2d2b844
[12] https://towardsdatascience.com/an-introduction-to-kernel-methods-9c16fc8fefd2
[13] Although for a different product, the section 5 of "User Guide to the CAMS Radiation Service (CRS)" is a suggested reading. https://atmosphere.copernicus.eu/sites/default/files/2021-05/CAMS72_2018SC1_D72.4.3.1_2021_UserGuide_v1.pdf
[14] https://www.soda-pro.com/web-services/radiation/cams-radiation-service
[15] https://nsrdb.nrel.gov/

It is also of interest to mention possible variations of this same implementation. First, the satellite estimate can be *a priori* calibrated for the target location based on historical data[16]. Second, alternative concomitant data sources can be considered: such as a neighboring GHI sensor or even a PV installation (which could be used as a variability proxy).

3.2.6 Discussion on the applicability of the tested GF methods

It is important to discuss the context of implementation of a given GF method, so that it is clear which applications a given method can be compatible with. In particular, from a data source and data availability in time perspectives. For example, a given method may require data only from the sensor under study (GF0 to GF3) or from a given satellite database (GF4). Additionally, it may depend on: only historical data which precede the gap (GF2 and GF3), a requirement for live data streams feeding real-time applications; data that come before and after the gap (GF0 and GF1), which benefits from a larger pool of information; or data that are concomittant to the data gap but comes from alternative data sources (GF4).

Regarding the later case, it is important to take into account that the alternative source of data (be it a satellite, a neighboring GHI sensor, or even a PV installation which could be used as a proxy) there is often a delay associated to the data collection, data transfer, and data processing. Thus, even though the data are concomittant, depending on the characteristics of the service, it may be impractical for real-time applications.

Method	Data source		Data are available ______ the gap		
	Sensor itself	Satellite	before[17]	before & after	concomitant to[18]
GF0	X			X	
GF1	X			X	
GF2	X		X		
GF3	X		X		
GF4		X			X

Table 3: Description of tested gap-filling models according to data source and data availability

[16] Because the impact of this calibration is highly dependent on the target location and the amount of available data, this has not been explored in this report.
[17] this would be the case for real-time data streams which feed real-time applications
[18] if neighboring in-situ data are available (GHI or PV as a proxy), and assuming an available and fast enough data transfer, could also be compatible real-time applications

3.3 Descriptions of the DSG methods

First, we test the ability of the gap-filling methods discussed in 2.3 for estimating the daily irradiation H. Since these produce complete daily profiles of GHI, by generating the data for the missing periods, we can calculate the daily sum simply with,

$$H_{GFx} = 24\langle G_{GFx}\rangle \tag{17}$$

Then, three DSG methods are also tested, all of them based on the multiplication of the daily average of available intra-day GHI (nighttime included), or a calibrated version of it, by the number of hours in a day (24). Thus, we note:

$$H_{DSGx} = 24\langle G\rangle_{DSGx} \tag{18}$$

3.3.1 DSG0: Using the simple average of intra-day GHI

Assuming that the daily average of the available data $\langle G\rangle$ (i.e., in our case an incomplete sample of a day) may still be representative of the real daily average, the DSG is estimated by multiplying the simple average of available intra-day GHI, disregarding eventual missing (NaN) values, multiplied by 24 hours.

$$H_{DSG0} = 24\langle G\rangle \tag{19}$$

3.3.2 DSG1: simple average of intra-day GHI calibrated by clear-sky model

Similar to DSG0 but tries to compensate the impact of the missing data by calibrating $\langle G\rangle$ with the ratio of the corresponding daily sum of clear-sky ESRA model with and without the same missing patterns.

$$H_{DSG1} = 24\langle G\rangle \frac{\sum_{k\in D} G_{cls}(k)}{\sum_{k\in V} G_{cls}(k)} \tag{20}$$

This assumption works best when the missing data correspond to clear-sky periods and the corresponding irradiance is well estimated by the ESRA model.

3.3.3 DSG2: Calibrated simple average of intra-day GHI

- Similar to DSG1 but uses the ratio of the corresponding daily sum of satellite-based HC3v4 database with and without the same missing patterns as a calibration factor, instead of the ESRA model.

$$H_{DSG1} = 24\langle G\rangle \frac{\sum_{k\in D} G_{sat}(k)}{\sum_{k\in V} G_{sat}(k)} \tag{21}$$

In comparison, this extends the best working conditions from clear-sky only to all sky conditions, as long as that is accurately modelled by HC3v4.

4 RESULTS OF THE BENCHMARK AT 15-MIN BASIS

The different GF and DSG methods are assessed in terms of standard statistical metrics, as those described for example by Espinar et al. (2009):

- Mean bias error (MBE),
- Root mean square error (RMSE),
- Mean absolute error (MAE),
- Correlation coefficient (CC).

These metrics at the intra-day and daily bases are used to evaluate the overall performance of the GF and DSG methods. Additionally, the dependency between the intra-day GF performance with the length of the period with missing data was further explored. To have a kind of analogy with forecasting, we have in fact considered that the time between a given timestamp with missing data and the nearest valid elements of the daily profile could be understood as the time horizon of the gap filling, named hereinafter the gap time horizon.

$$Hz(k) = |n(k) - k| \quad (22)$$

For DSG assessment, the discriminating element is the fraction of missing values within a day:

$$p_M = N_V / N_D \quad (23)$$

The validation dataset has been used for the following hyperparameters for the ML approaches GF2 and GF3:

- L = 10 (for the kNN approach of GF2),
- σ^2 = 0.0144 (for the Kernel-based approach of GF3).

4.1 Assessment of intra-day GF methods

Over the whole TEST dataset, the satellite-based GF method outperforms the other GF methods:

	NDATA (timesteps)	MREF (W/m²)	MBE (%)	MAE (%)	RMSE (%)	CC (-)
GF0	232556	338.1	-2.9	17.5	33.6	0.923
GF1			-3.6	15.7	29.2	0.941
GF2			3.0	20.6	34.8	0.916
GF3			3.0	20.4	33.6	0.920
GF4			**1.3**	**15.2**	**24.1**	**0.959**

Table 4: Overall performance of tested gap-filling methods

Because this table only shows the overall performance, it may hide some variability in performance in respect to other variables. Here, we explore potential dependencies with the gap time horizons, which are not uniformly distributed in the dataset.

Figure 5 provides the distribution of the number of daily profiles for different bins of gap time horizon.

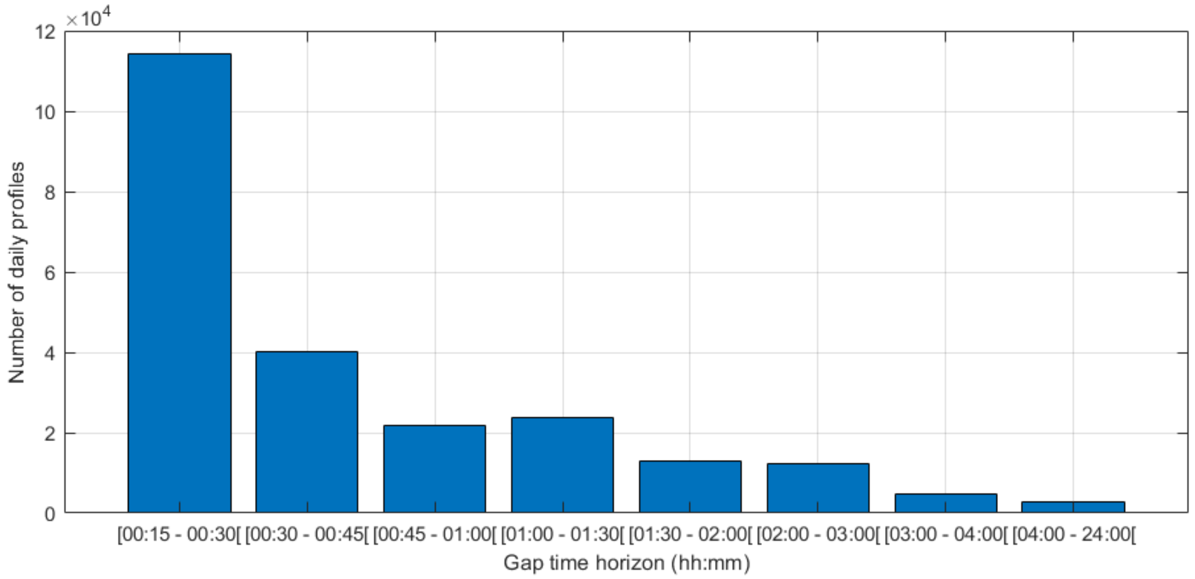

Figure 5: distribution of the number of missing 15-min GHI with respect bins of gap time horizons.

For each of these bins, Figure 6 represents the average daily GHI used for relative RMSE.

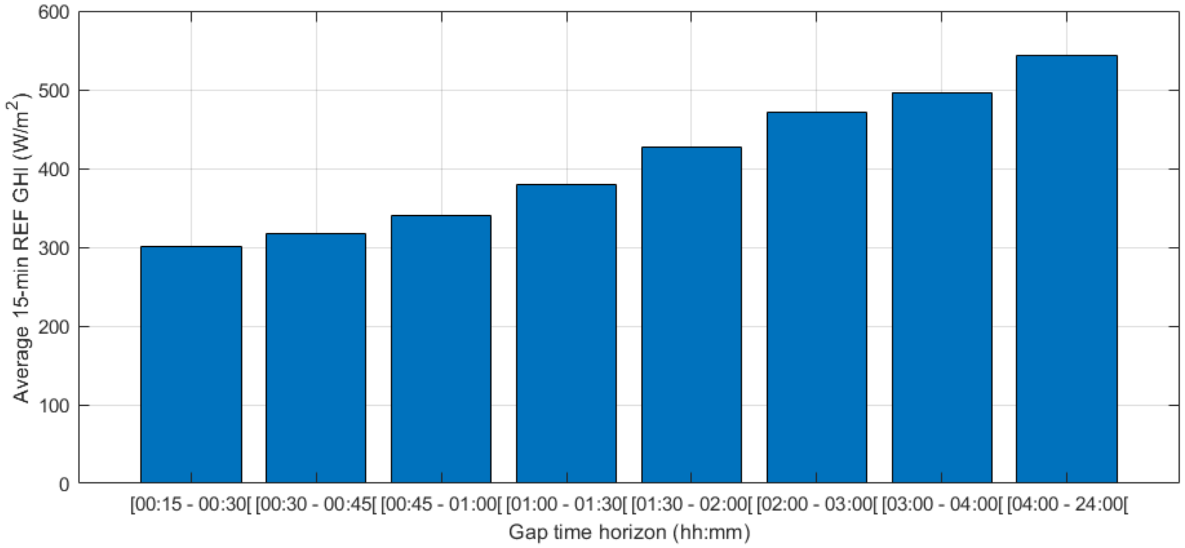

Figure 6: averages 15-min GHI from the complete reference daily profiles for the different bins of gap time horizons.

Figure 7 presents the relative RMSE (%) for each gap-filling method (GF0 to GF4) for 8 different intervals of gap time horizon. For gap time horizons less than 30 min, the best approach is the linear interpolation of clear-sky index (GF1). For larger gap time horizons, satellite data provide (GF4) the best GF approach. However, when satellite data are not available, the kernel-based approach (GF3) is better than the linear approach (GF1) for gap time horizons larger than 2 hours.

The use of the GF3 approach should be carefully evaluated because, like the GF2 k-NN method, it depends on the number daily profiles of the training (TRAIN) dataset used per station that ranges here between 8500+ and 37000+ daily profiles. For such methods, the validation (VALIDATION) dataset must be used in an operational condition to assess the effective running performance of GF2 and GF3 methods, considering the historical dataset at hand and its amount of complete daily profiles.

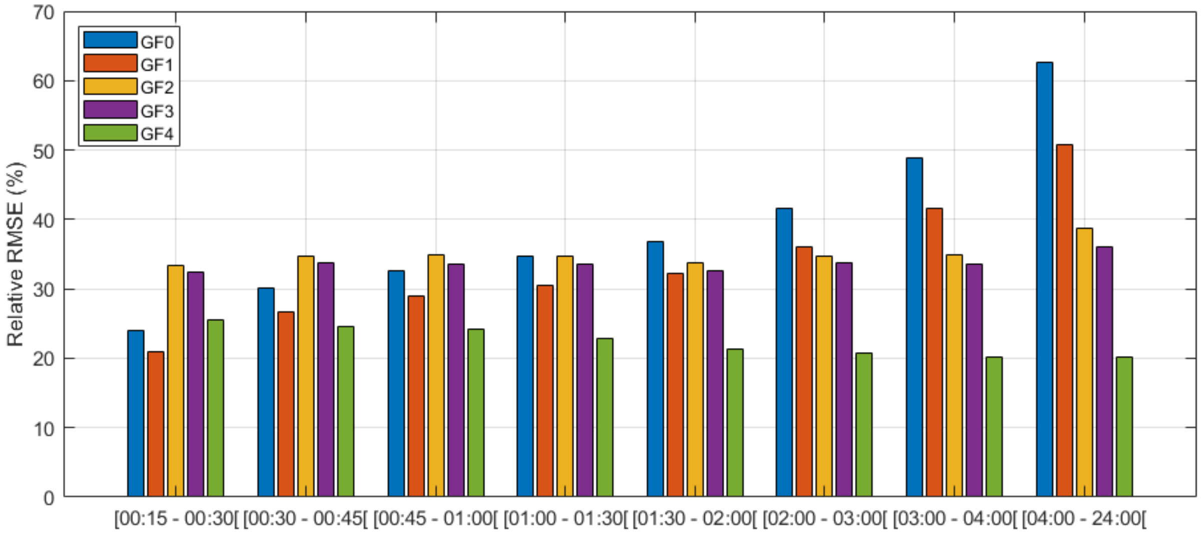

Figure 7: relative RMSE for the different GF methods for the different bins of gap time horizons.

Figure 8 presents with boxplots, per bins of gap time horizon, the variability with respect the 16 stations of the relative difference of RMSE between GF4 based on satellite and GF1 based on linear interpolation of clear-sky index. This relative difference of RMSE can be interpreted as RMSE-skill scores of GF4 compared to GF1.

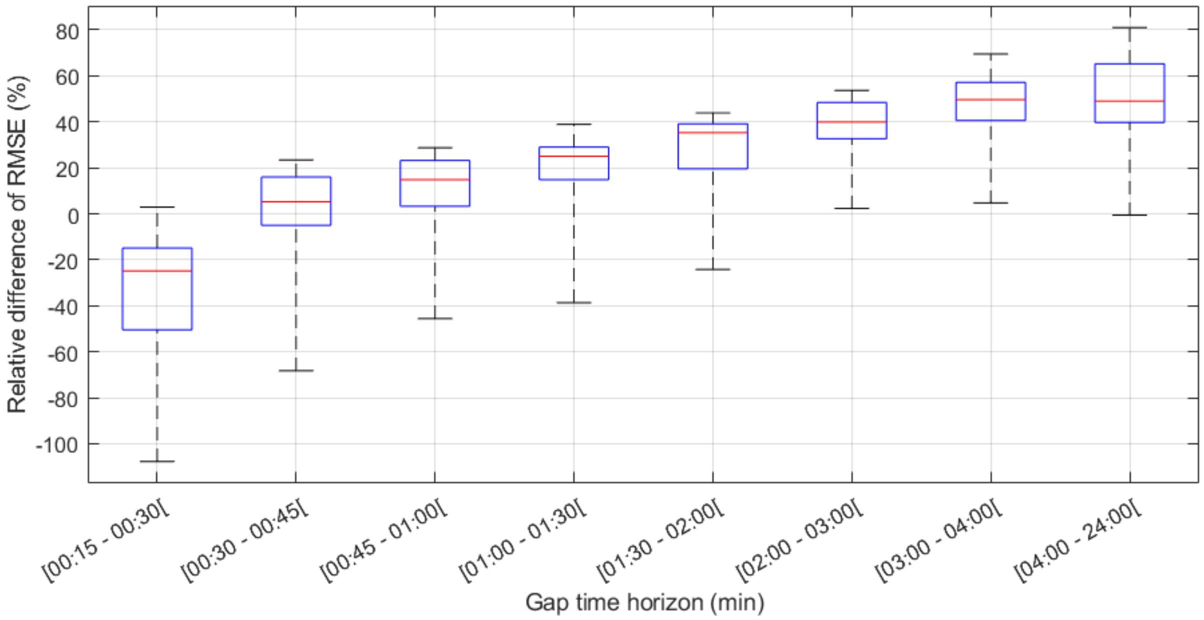

Figure 8: boxplots of station-wise relative difference of RMSE between GF4 and GF1 for the different bins of gap time horizons.

Except for the gap time horizons greater than 2-h, some stations do not present a positive skill-score for GF4 when the gap time horizon is larger than 30 min: the average gain of GF4 with respect GF1 observed in Figure 5 is not uniform with respect the location. The performance of GF4 intrinsically depends on the performance of the satellite database for that location.

Figure 9 shows boxplots of the variability of difference of RMSE between GF4 and GF1 with respect all gap time horizons larger than 30 min, for the different stations. Two stations clearly present under-performance of GF4 compared to GF1: IZA and TOR.

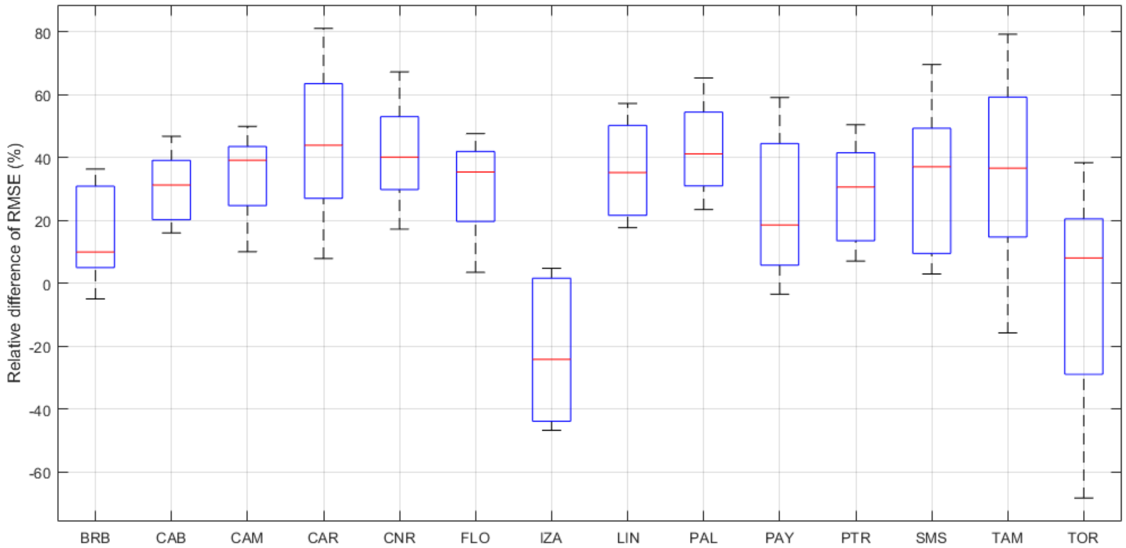

Figure 9: boxplots of difference of RMSE between GF4 and GF1 for all gap time horizons larger than 30 min, per station.

4.2 Assessment of the DSG methods

Over the whole TEST dataset, the satellite-based DSG method outperforms the other methods for the DSG estimations, except for the RMSE criterion for which the satellite-based GF method provides slightly better results.

	NDATA (days)	MREF (Wh/m²)	MBE (%)	MAE (%)	RMSE (%)	CC (-)
GF0→d	26293	3613.1	-0.6 %	2.1 %	6.2 %	0.996
GF1→d			-0.7 %	2.1 %	6.5 %	0.995
GF2→d			0.6 %	2.9 %	6.8 %	0.995
GF3→d			0.5 %	2.8 %	6.5 %	0.995
GF4→d			0.3 %	2.0 %	**3.7 %**	**0.998**
DSG0			-1 %	3.2 %	7.4 %	0.994
DSG1			-0.4 %	2.6 %	6.2 %	0.996
DSG2			**-0.1 %**	**1.8 %**	4.5 %	**0.998**

Table 5: Overall performance of the tested "daily sum with gaps" methods

Just as for the GF methods, limiting the analysis only to the overall performance hides a variability against the percentage of missing data within the daily profiles. Again, we can see a non-uniform distribution of days with respect the percentage of missing data during daylight (Figure 10).

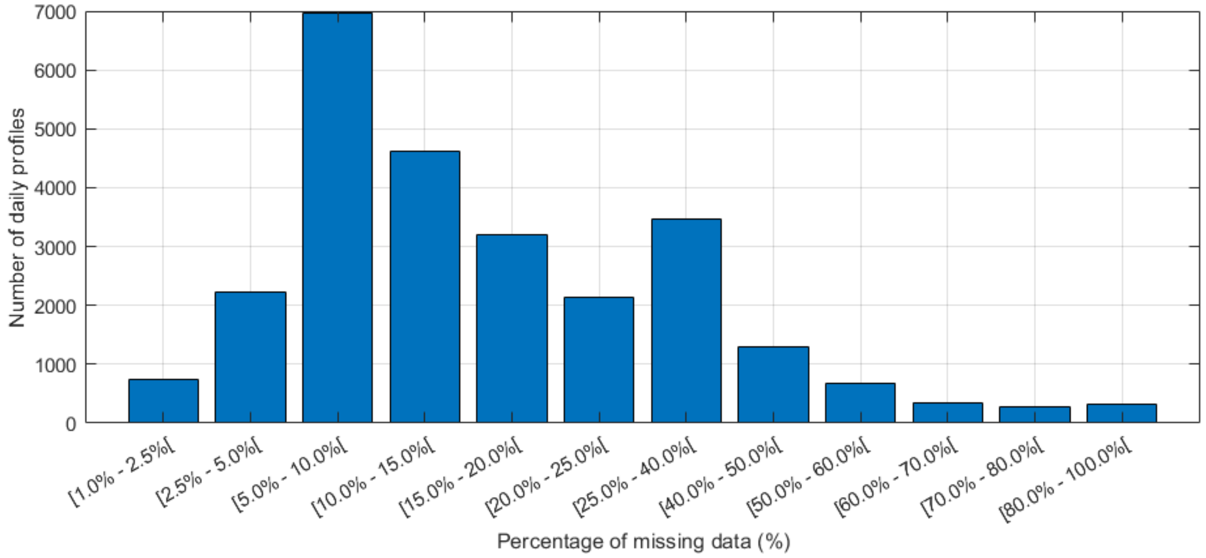

Figure 10: number of daily profiles for the different bins of daily percentage of missing data.

Figure 11 presents the corresponding average of daily sum of irradiation for the same bins of daily percentage of missing data as Figure 8.

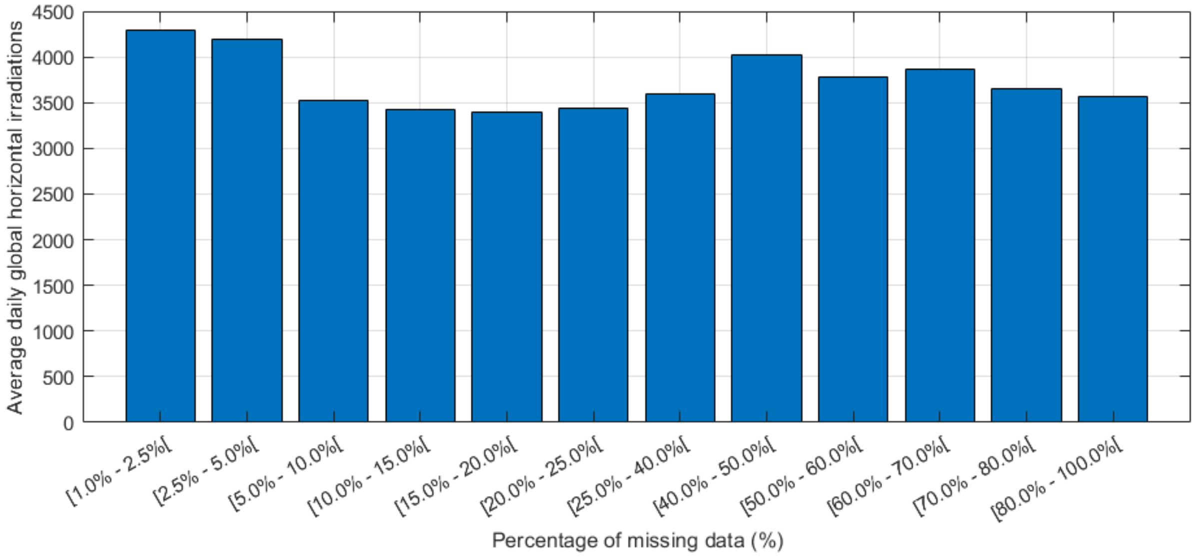

Figure 11: daily global horizontal irradiations from the reference complete daily profiles for the different bins of corresponding daily percentage of missing data.

Owing to Figure 12, having less than 5 % of missing data within a day induces a relative RMSE lower than 1 %, whatever the DSG (or GF) methods. Having less than 20 % of missing data induces an uncertainty between 2 % and 4 %, depending on the method used.

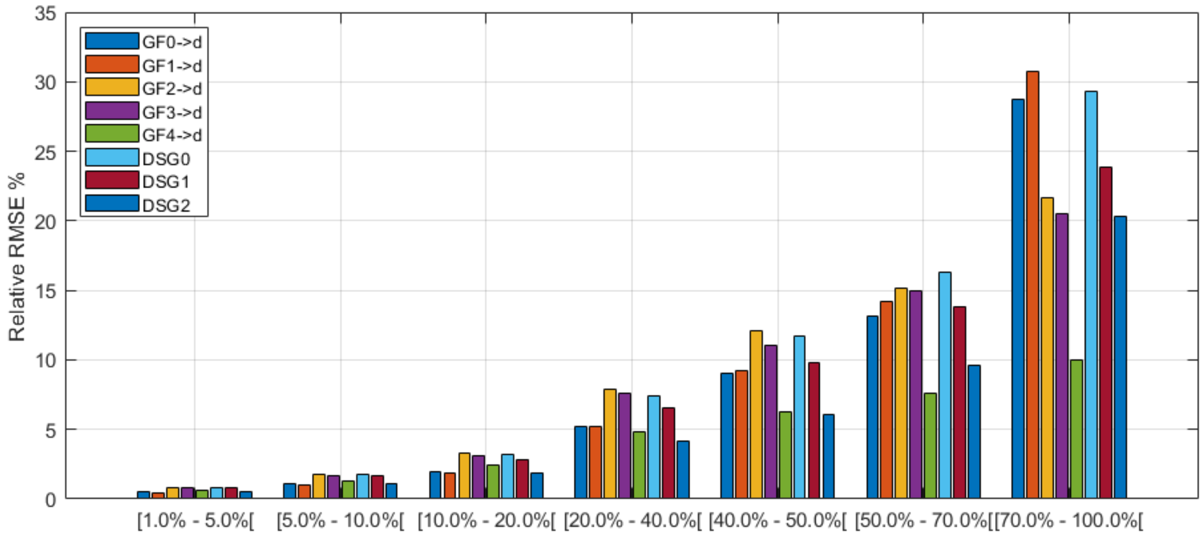

Figure 12: relative RMSE (%) of DSG for the different bins of daily percentages of missing data.

Figure 13Figure 13, presenting the RMSE-based skill values with respect to the GF0→d method, highlights the best approaches for the different bins of daily percentages of missing data. Up to 20 % of daily missing data, the GF1→d method based on linear interpolation of clear-sky index provides the most accurate DSG. After this threshold, satellite-based approaches, DSG2 and GF4→d, provide the most accurate DSG, respectively for daily missing data percentages between 20 % and 50 % and larger than 50 %.

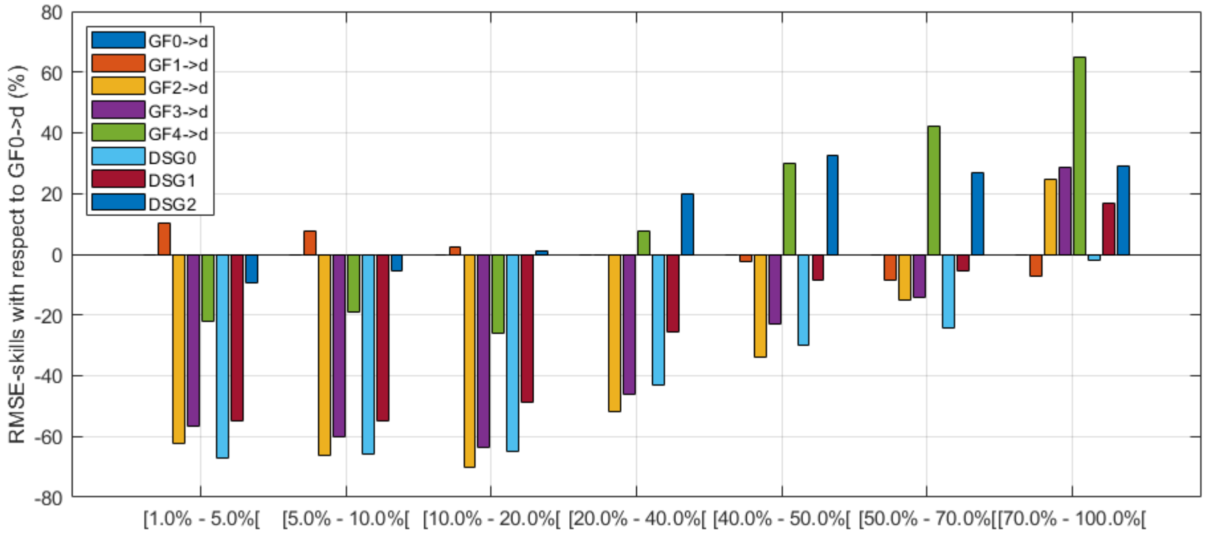

Figure 13: relative RMSE-skills (%) of DSG for the different bins of daily percentages of missing data, compared to the GF0→d method.

Just like the intra-day performances discussed in section 4.1, these performances averaged over the stations hide some disparities within them. Figures 14 and 15 present per station the variability of the RMSE-skills (%) of respectively GF4→d and DSG2 for data daily percentages of missing data larger than 20 %. The RMSE-skills show clear underperformances of the two satellite-based approaches for IZA, TOR and PAY.

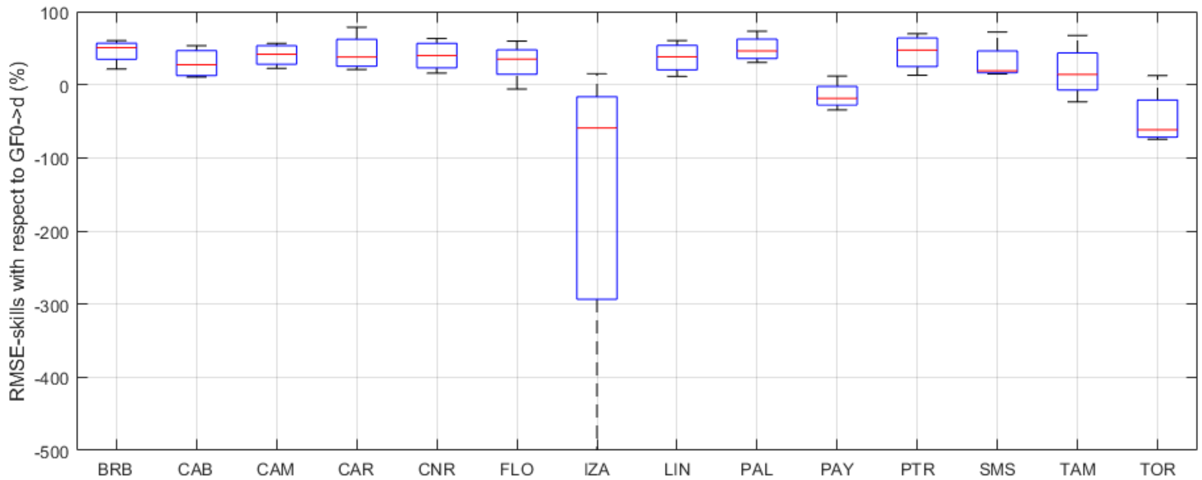

Figure 14: boxplots of difference of RMSE between GF4→d and GF0→d for missing data percentage larger than 20 %, per station.

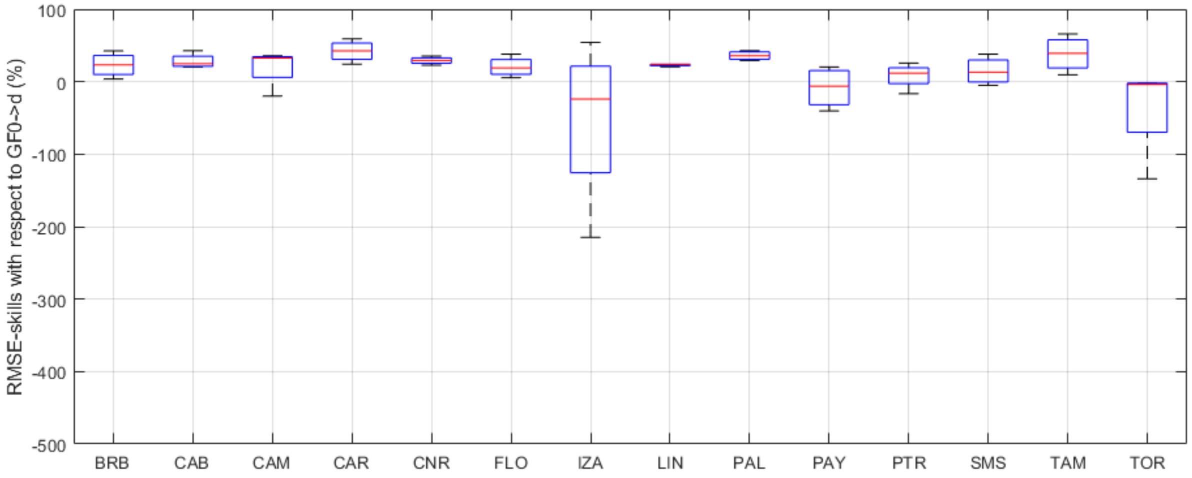

Figure 15: boxplots of difference of RMSE between DSG2 and GF0→d for missing data percentage larger than 20 %, per station.

5 CONCLUSION AND PERSPECTIVES

This report aims to propose a general framework meant for benchmarking gap-filling methods as well as evaluate the performance of a selection of baseline methods. This is done both for intra-day completion and temporal aggregation from intra-daily to daily values of incomplete Global Horizontal Irradiance (GHI) time series.

The general principle of this framework is to use real 1-min data of irradiance from in-situ pyranometric stations to get both high-quality QC-checked complete daily profile of 1-min irradiance along with real patterns of missing data during daytime. Having these two reference datasets, a Monte-Carlo procedure is used to generate synthetic daily profiles of irradiance with different time resolutions (by aggregation of the original 1-min data) both complete and with various data missing patterns. These synthetic profiles can be then used respectively as the reference and the input of gap-filling methods under test.

This general framework is implemented in this work from high-quality QC-checked daily profiles of 1-min GHI time series provided by 16 BSRN stations over the time period 2004-2018, along with satellite-data from Helioclim-3v4 and solar position related information from the SG2 library. This 1-min NetCDF database is available online[19], as well as the corresponding databases for the time steps 10-min, 15-min and 1-h. Of course, these databases can be extended to a larger set of pyranometric stations with available corresponding satellite data.

To apply this general framework, the report presents the approaches and the performances of 5 gap-filling methods meant to complete time series of 15-min GHI and 8 gap-filling methods to provide daily sums of irradiations from daily profiles with missing data. The second case includes the 5 intra-day completion models, which can simply be aggregated after the gap-filling is done, as well as 3 other methods that can compute the compensated daily sums without explicitly gap-filling the intra-day profiles beforehand.

From the obtained results, below is the overall ranking performance-wise (from best to worst, RMSE-wise):

- GF4, GF1, GF0 and GF3 (tied), and GF2, for the intra-day completion of daily profiles,
- GF4→d, DSG2, GF0→d and DSG1 (tied), GF1→d and GF3→d (tied), GF2→d, DSG0, for calculating the daily sums of irradiations from daily profiles with missing data.

However, it is important to note that not only this ranking may change depending on the chosen statistical performance metric, but also that this performance depends on the target location and the horizon of the data gap.

For example, for large data gaps, the use of GF4 (concomitant satellite data) demonstrates, on average, the highest performance but some significant under-performance may arise for some specific location. Thus, using the proposed Monte-Carlo-based benchmark framework with synthetic daily profiles using historical data can be very useful to quantify the local performance gains (or losses) of using satellite-based gap-filling methods compared to simpler approaches like the linear interpolation of clear-sky index.

[19] "*db_benchmark_1min_v0.nc*" file in https://cloud.mines-paristech.fr/index.php/s/9obUl2ZVGZTUU12

It was also shown that for gap time horizons less than 30 min, the best approach is in fact GF1 (linear interpolation of clear-sky index), whereas GF4 (using concomitant satellite data) performs best for higher horizons. However, when satellite data are not available, GF3 (the kernel-based approach) is better than GF1 for gap time horizons larger than 2 hours. It is important to note that for real-time applications the gap horizon is not known beforehand. For such contexts, either the best overall performing model or the one that better suits the most frequent gap horizon for a target location can be deployed. A simplistic alternative could be to implement a model switch could be put in place that changes from one approach to another as the gap horizon surpasses a given threshold.

Nonetheless, it is relevant to acknowledge that a RMSE-based ranking of the gap-filling approaches under test may induce oversimplistic and inconsistent conclusions. For example, for some methods, notably based on satellite data, the average good performance may hide some local significant under-performances (e.g., underestimation of solar resource local variability). Additionally, it disregards the robustness and implementation complexity of each method, as well as specificities of the application context. While the (un)availability of satellite data are an easy example to illustrate this, two more differentiated cases are provided:

- the added value of the generated synthetic data and which statistical indicator is a better proxy for its practical value greatly depend on the application at hand,
- real-time data streams and applications can only use past data when a data gap occurs and may prioritize factors such model latency or time-of-response.

Regarding potential developments for this benchmark, it could be extended to:

- include larger reference databases (both in time and in number of locations),
- integrate more gap-filling approaches,
- integrate more performance tests, for example one which differentiates “clear” from “overcast” or “cloudy” days.

Additionally, the principle of the framework itself can be of course questioned. First, this Monte-Carlo implementation is based on random draws of already existing data gap patterns but disregards their external causes (meteorological parameters, maintenance constraints, etc.): some unrealistic scenarios may then arise that may create some biases in the performance analysis. Pure random generation of data gaps based on time of occurrence and duration for each location could be implemented, as well to check the consistency of the results, compared to the existing approach.

6 REFERENCES

Baseline Solar Radiation Database (BSRN). https://bsrn.awi.de/

Blanc, P., & Wald, L. (2012). The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. Solar Energy, 86(10), 3072–3083. https://doi.org/10.1016/j.solener.2012.07.018

Ching (2020). K-Nearest Neighbours (kNN) Algorithm: Common Questions and Python Implementation. Towards Data Science (online)

Campos, I., Brito, M., de Souza, D., Santino, A., Luz, G., Pera, D., (2022). Structuring the problem of an inclusive and sustainable energy transition – A pilot study. Journal of Cleaner Production, 365. https://doi.org/10.1016/j.jclepro.2022.132763

Espinar, B., Ramirez, L., Drews, a, Beyer, H. G., Zarzalejo, L. F., Polo, J., & Martin, L. (2009). Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. Solar Energy, 83(1), 118-125. https://doi.org/10.1016/j.solener.2008.07.009

Gschwind, B.. GitHub repository for SG2 package for sun position calculations. https://github.com/gschwind/sg2

Ivanov, G. (2021). An Introduction to Kernel Methods. Towards Data Science (online)

Muhajir, I. (2019). K-Neighbors Regression Analysis in Python. Medium (online)

NREL. National Solar Radiation Database. https://nsrdb.nrel.gov/

Pathak, M., Slade, R., Shukla, P.R., Skea, J., Pichs-Madruga, R., Ürge-Vorsatz, D. (2022). Technical Summary. In: Climate Change 2022: Mitigation of Climate Change. Contribution of Working Group III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK and New York, NY, USA. https://doi.org/10.1017/9781009157926.00

Pramanik, N. (2019). Kernel Regression – with example and code. Towards Data Science (online)

Rigollier, C., & Bauer, O. (2000). On the clear sky model of the ESRA – European Solar Radiation Atlas – with respect to the HELIOSAT method. Solar Energy, 68(1), 33–48. https://doi.org/10.1016/S0038-092X(99)00055-9

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., & Zhang, T. (2011). Assessment of BSRN radiation records for the computation of monthly means. Atmospheric Measurement Techniques, 4(2), 339–354. https://doi.org/10.5194/amt-4-339-2011

SoDa – Solar Radiation Data. CAMS Radiation Service. https://www.soda-pro.com/web-services/radiation/cams-radiation-service

SoDa – Solar Radiation Data. Helioclim-3 (HC3) overview. https://www.soda-pro.com/help/helioclim/helioclim-3-overview

SoDa – Solar Radiation Data. Helioclim-3 Archives - Web Service. https://www.soda-pro.com/web-services/radiation/helioclim-3-archives/info

Liu, W., Liu, Y., Zhou, X., Xie, Y., Han, Y., Yoo, S., Sengupta, M. (2021). Use of physics to improve solar forecast: Physics-informed persistence models for simultaneously forecasting GHI, DNI, and DHI. Solar Energy, 215, 252-265. https://doi.org/10.1016/j.solener.2020.12.045

7 ANNEX 1: LOCATION OF CONSIDERED GHI DATA

BSRN Code	Station name	Latitude	Longitude	Elevation
BRB	Brasilia	-15.6010°	-47.7130°	1023 m
CAB	Cabauw	51.9711°	4.9267°	0 m
CAM	Camborne	50.2167°	-5.3167°	88 m
CAR	Carpentras	44.0830°	5.0590°	100 m
CNR	Cener	42.8160°	-1.6010°	471 m
FLO	Florianopolis	-27.6047°	-48.5227°	11 m
ILO	Ilorin	8.5333°	4.5667°	350 m
IZA	Izana	28.3093°	-16.4993°	2373 m
LIN	Lindenberg	52.2100°	14.1220°	125 m
PAL	Palaiseau	48.7130°	2.2080°	156 m
PAY	Payerne	46.8150°	6.9440°	491 m
PTR	Petrolina	-9.0680°	-40.3190°	387 m
SBO	Sede Boqer	30.8597°	34.7794°	500 m
SMS	Sao Martinho de Serra	-29.4428°	-53.8231°	489 m
TAM	Tamanrasset	22.7903°	5.5292°	1385 m
TOR	Toravere	58.2540°	26.4620°	70 m

Table 6: Description of the considered sites with GHI measurements from the BSRN network.

8 ANNEX 2: IMPLEMENTED QUALITY-CHECK PROTOCOL

As mentioned in the body of this report, the quality-check protocol considered in this report can be found in Roesch et al. (2011). It aims to flag observations that are suspected of being erroneous, so that for a given application the end user may choose if they wish to disregard such data. The authors propose two possible procedures when only GHI data are available: one defining physically possible limits for GHI observations, considering the time and location at hand; the other, more demanding, which defines extremely rare limits that GHI would only surpass over very short periods and under very rare conditions.

For this report, the second procedure was implemented, and it consists in the following accepted ranges:

$$-4\ W.m^{-2} \leq GHI \leq 1.5 \times S_0 \times \mu^{1.2} + 100\ \ W.m^{-2} \tag{24}$$

where S_0 is the solar constant adjusted for Earth-Sun distance and μ is the cosine of the solar zenith angle. As also mentioned in the report, both variables were calculated using the sg2 library (Blanc et al., 2012) which is available in GitHub[20].

[20] https://github.com/gschwind/sg2

Technology Collaboration Programme
by iea

ISBN 978-3-907281-37-6