STANDARDIZING AND BENCHMARKING OF MODELED DNI DATA PRODUCTS

Richard Meyer¹, Chris Gueymard² and Pierre Ineichen³

¹ Suntrace GmbH. Address: Brandstwiete 46, 20457 Hamburg, Germany. Phone: (+49) 40 767 9638 0 E-mail: r.meyer@suntrace.de
² Solar Consulting Services, P.O. Box 392, Colebrook, NH 03576, USA
³ University Genève, Battelle Bat D-7 rte de Drize, CH-1227 Carouge, Switzerland

Abstract

Modeled direct normal irradiance (DNI) can be either derived from satellite data or from numerical weather prediction models. Such modeled datasets are available at continental scale and provide continuous long-term time series, but are known to fall short in quality over some areas, compared to high-quality ground measurements. The uncertainty in DNI may be locally so high that CSP projects cannot be financed. The CSP industry would obviously benefit from a comprehensive and large-scale benchmarking of the existing modeled DNI datasets. This would help CSP developers select the most appropriate dataset for a given region, and would also provide due-diligence or financial analysts with the desired information on the expected accuracy of the data. This contribution investigates how the benchmarking study should be conducted, and evaluates the difficulties that can be encountered. A large set of ground stations with anticipated good- to high-quality measurements has been identified, most of which reporting public-domain data. Various criteria that such measured datasets must fulfill to be usable in the benchmarking study are discussed. Automatized quality control methods are also proposed for quality assessment purposes, and are described in some detail. The most important statistics to be used for the benchmarking process are discussed, with an emphasis on those that appear of particular importance to the CSP industry. The full-scale benchmarking study should now follow, assuming proper funding is secured.

Keywords: Concentrating Solar Power (CSP), solar resource, Direct Normal Irradiance (DNI), DNI uncertainty, satellite-derived DNI, quality control

1. Introduction

Direct normal irradiance (DNI) is the ‘fuel’ of concentrating solar power (CSP) systems. The intensity and distribution of DNI are dominating factors for the design and performance of CSP plants. Due to the significant short-term and long-term variability in DNI, CSP analysts must obtain DNI data time series—not just mean annual averages. It has been shown that a relatively accurate annual value might be predicted by satellite-based models, but with large errors in monthly values over vast regions [1]. Using probabilistic performance simulations, which also take the uncertainty in many technical parameters into account, it has also been shown that the uncertainty and variability in annual DNI at a site is the single greatest cause of uncertainty when predicting the energy production of a CSP plant [2]. In addition to the need for verifiable accuracy in monthly-average DNI, realistic performance simulations of a CSP installation depend on reliable time-series of DNI and ancillary meteorological variables, such as temperature. Many simulation tools use only a synthetic year, referred to as TMY. To avoid power output prediction errors of typically ≈10 %, such TMY data sets must well represent the hourly frequency distribution of DNI [3], but this is not always the case. Moreover, recent studies (e.g., [4]) suggest that, when simulating the transient behavior of CSP plants, larger errors on the total energy produced may occur when using datasets whose time resolution is not better than one hour. As the characteristics of the available DNI time series have a strong impact on the design and financing of large solar plants, it is necessary to quantify the accuracy of such data products under various aspects, from long-term averages down to sub-hourly time-steps.

The validation of available satellite-derived DNI data is currently insufficient, due in part to the lack of measurement sites with good-quality data, particularly over regions with large CSP potential but limited infrastructure. Validation exercises take time and must be done by experts and, hence, require sufficient financial resources. So far, such resources have been scarce in general. The CSP industry, which would directly benefit from such studies, prefers to install weather stations at the specific sites where their own CSP projects are planned. The datasets collected at these stations would be essential for validation studies, but are most often inaccessible since they are proprietary. Often, the data producers themselves do the only existing validation, which is against best practices for due-diligence work, for instance. Recently, Süri et al. [5] intercompared five satellite-derived DNI datasets and found very strong differences between them. The standard devi-
ation of the differences between various long-term data series reached more than 24% in some regions. Such an uncertainty is much too high for bankability. On average over the study’s region (Europe), the standard deviation was in the range of 10%, which is way beyond the requirements for CSP project development. From the results reported in [1] and [5] it is expected that differences can be even higher in other regions than Europe. To gain better understanding of the magnitude of the uncertainty in modeled DNI and its dependence on geography, and to help data providers eventually improve their products, a more general study would be desirable, with worldwide coverage and for as many datasets as possible. A shortcoming of the European study [5] is that the average of all data sets is used as the reference instead of ground-based measurements. Rather than identifying the absolute quality of the datasets, this only displays the fact that there is substantial disagreement, which leaves many unanswered questions.

Now that the CSP industry has matured, it urgently needs an in-depth analysis of the currently available satellite-derived datasets, so that it can continue its expansion in various parts of the world where the current modeled datasets may not be of sufficient quality. Ideally, validation studies should be performed by independent experts with various datasets, using reference measurements that have not been used by the data providers to train or calibrate their model.

For the above reasons, the Executive Committee of SolarPACES decided to start funding a research project to benchmark those DNI datasets whose participation would be authorized by their developer. This report describes the first phase of this project. It gives an overview of current DNI data products, provides criteria for the selection of DNI measurements that will be appropriate for validation and benchmarking, and finally proposes a general methodology to benchmark existing (or future) modeled DNI datasets.

2. Overview of available modeled DNI data products

Several satellite-derived DNI products are now available. NASA-SSE, NREL-CSR, NREL-SUNY or SolarLight are examples of public-domain data sources. SoDa/HelioClim and DLR-Solemi are commercial products that also offer samples of free data. Data from other providers, such as CleanPower, EnMetSol, focus solar, GeoModel, IrSOlA, s2m, or 3TIER are only of a commercial nature, and are reportedly based on better radiative models and input data. It has to be shown whether this translates into higher-accuracy DNI results on a global (continental or worldwide) basis, however. Like the CSP industry, solar radiation data providers face a fierce competition, and are thus under pressure to provide the best possible products. Indeed, significant progress in modeling has been made in recent years (e.g., [6–10]).

In principle, DNI can also be calculated from numerical weather prediction models. Usually such models are used for forecasting purposes, but they can also be run in re-analysis mode to create long historical time series. Since the industry has requested that such kind of data be analyzed too, it has been decided to open up the planned benchmarking study to such products, even though they do not directly use satellite-based cloud data as input. Another request from the CSP industry is to evaluate other types of data products, not directly based on satellite or reanalysis data, but on climate data and spatial/temporal interpolation. A typical product in this category is the Meteonorm software [11], which is often used for design purposes at least.

3. Worldwide measurement sites for ground truth

Since ground-based DNI measurements are to be used as reference for validation and benchmarking, they must respond to some important criteria:

1. Each station must be at relatively low latitudes (between -45° and 45°) to be of interest for CSP
2. A large number of stations must be considered, so as to increase the significance of the results
3. They should be well spread geographically and climatically over all regions of interest
4. The data quality must be as high as possible, which means they must be obtained with first-class instrumentation, using high standards for measurement method, calibration, maintenance, quality control, etc.
5. As much as possible, measurements should be made at stations that are not likely to have been used by modeled data developers or providers for empirical fine-tuning of their models.

Criterion #5 is delicate, since not all data providers resort to empirical fine-tuning, but some of them do. When they do, the list of stations actually used for that purpose is usually proprietary information, and will most likely not be available to third parties. However, it is safe to assume that the best-known stations represent the “low hanging fruit” that would be picked by modelers for this exercise. Consequently, an important aspect of this task was to “discover” a large number of lesser-known stations in all possible climatic areas, to also satisfy Criteria 2 and 3 above. Hence, of high interest to help with Criterion 5 are stations whose data is only available through personal contact with the station’s caretaker, or not available publicly at all. Personal
contacts have identified many “private” stations whose data could be made available to us, generally on the basis of a “personal communication” and/or with the signature of a non-disclosure agreement (NDA). Some station caretakers rather asked to be associated to this research, and their case is pending, since it depends on legal and budget aspects that are not resolved yet.

Although the CSP industry undertakes resource assessment measurements at various sites, such datasets are extremely sensitive and would usually not be accessible to any third party, even under an NDA. We still try to obtain such permissions for the benchmarking exercise, with the argument that it will benefit the industry in the first place, and the owner of such private datasets most particularly.

Figure 1 shows the location of the 185 “public-domain” measuring stations currently known. About 60 private stations have also been identified. It is anticipated that a few more stations will be added in the near future, since we continue contacting people who may know about public or private measured data sources. It is obvious that some regions of the world (e.g., USA and Spain) have a good density of stations, whereas essentially the rest of the world is poorly represented.

![Fig. 1: Location of public stations (blue diamonds), superimposed on a map of mean annual DNI based on the NASA-SSE dataset.](image)

Criterion 4 is another delicate issue because “quality” is hard to define precisely, and may vary a lot during the life of each single station. Degradation of quality has been observed at many stations and in many countries, generally because of decreasing budgets, policy changes, or priority changes at research institutions. During this part of the study, personal contacts with many caretakers have revealed that they could not guarantee the quality of the data, or that some periods were of questionable quality, due to insufficient maintenance or quality control procedures. The relatively high frequency of this issue indicates that a thorough quality assessment (QA) of the data from each station should be made prior to their qualification as “ground truth” for the benchmarking process per se. QA techniques exist, and are reviewed in Section 4. Since a rigorous QA is time consuming and the techniques used may be somewhat site specific, it was not attempted at this stage of the study due to stringent time and budget limits. Therefore, not all stations selected here may ultimately be available for use as ground truth. One important tool toward a successful QA is the simultaneous availability of all three radiation components: direct normal irradiance (DNI), global horizontal irradiance (GHI) and diffuse horizontal irradiance (DIF). Not all stations are in this ideal case, unfortunately.

There are many discussions among experts, and in some publications (e.g., [12–16]), about how the type of instrumentation affects irradiance data quality. It is our opinion that, by design, thermopile radiometers normally provide the best response, if well maintained, which is often not the case. However, such instruments are expensive and have stringent requirements: relatively high power consumption for the trackers and ventilators, and frequent cleaning (ideally every day). Some thermopile pyranometers are also known for their significant thermal imbalance, which particularly affect global and diffuse measurements under clear skies, even if the instrument is properly calibrated [14]. To avoid these high costs and potential issues, many recent radiometric stations, particularly those installed by or for CSP developers, include Rotating Shadowband Irradiance sensors (RSI) constructed from solid-state silicon sensors. Although they are much less expensive, and have much lower power requirement and a much faster response time than thermopiles, their drawback is
their imperfect spectral response. Empirical corrections must be introduced a posteriori to bring their measurement’s quality close to that of thermopiles. Field experience has shown that RSIs need less frequent cleaning than thermopiles [17]. This can be a decisive advantage for remote stations. The difficult issue we face here, as third-party users of data, is that we generally do not know how well each instrument was maintained, and particularly how frequently it was cleaned during any given period. As mentioned above, budget restrictions and other issues frequently lead to lack of maintenance, and thus to rapid data degradation. Although under ideal (laboratory-type) conditions, thermopiles are considered reference instruments and have the edge over silicon sensors, performance wise, the situation is not obvious under field conditions. It is then often considered that a rotating shadowband instrument with a regular 7-day cleaning cycle should provide better-quality data than a thermopile with such infrequent cleaning. With all these considerations, we have opted to include stations equipped with both types of instrument.

All stations have gaps in their data, which is unavoidable. To limit biases in reference data and incorrect results, extra care will be necessary. For instance, when calculating hourly averages from original data points at higher-frequency, periods with more than a few missing data points should be discarded. The validation of data series available only with a daily or monthly time increment may well be biased, depending on the fraction of missing original data points they tolerate. Finally, historical stations that have ceased operation before 1995 are of no interest, since the large majority of modeled databases only start after that date.

4. Measured data quality

Sensor calibration is the key point for precise acquisition in the field of solar radiation. The radiation sensors should be calibrated by comparison against a sub-standard before the beginning of the acquisition period, and then every year. Due to a variety of possible errors, a precise calibration correction (or virtual calibration, based on the available measured data) is extremely difficult to conduct a posteriori.

Controlling the observed data quality is the first step to perform in the process of validating models against observations. This essential step should be properly devised and efficiently automated, in order to rapidly detect significant instrumental problems, like sensor failure, inconsistencies, or errors in calibration, orientation, leveling, tracking, etc. This quality-control process should be done by the institution responsible for the measurements. Unfortunately, it is not the case at all stations, or at all times during the recording period. Even if some quality control procedure has been implemented, it might not be sufficient to catch all errors, or the data points might not be flagged to indicate the source of the problem. A stringent control quality procedure must therefore be adopted in the present context. Its various elements are summarized in what follows.

If the three irradiance components—DNI, GHI and DIF—are available, a fundamental consistency test can be applied, based on the closure equation that links them. No a posteriori quality control can detect all acquisition problems that could have happened, however. The remaining elements to be assessed are threefold: (i) Measurement’s time stamp (needed to compute the solar geometry); (ii) Sensor calibration coefficients used to convert the acquired data into physical values; and (iii) Coherence between irradiance components.

4.1 Time stamp

To detect any possible time shift in the data, the clear-sky irradiance symmetry (with respect to solar noon) must be checked visually. Hourly GHI and DNI values are plotted versus the sine of solar elevation angle, \( h \), for specific clear days. If the time stamp is correct, the afternoon curve should lay over the morning curve. Exceptions do occur, however, at sites where the atmospheric turbidity changes during the day, due for example to topography-induced effects. (This is the typical situation at Golden, Colorado, for instance, where the clear-sky irradiance is significantly lower in the afternoon than in the morning.) Another verification can be done with the help of the clearness indices \( K_c \) and \( K_b \), which are respectively defined from:

\[
K_c = \frac{G_h}{I_o \cdot \sin(h)} \quad \text{and} \quad K_b = \frac{B_h}{I_o} \tag{1}
\]

where \( G_h \) stands for GHI, \( B_h \) stands for DNI, and \( I_o \) is the extraterrestrial irradiance (i.e., the solar constant corrected for the actual sun-earth distance). Hourly values of each clearness index are then plotted for morning and afternoon separately. The upper envelope of the morning and afternoon data is representative of clear-sky conditions. An example appears in Fig. 2, based on one year of GHI data acquired at Carpentras (France). Ideal hourly clear-sky values are plotted in blue on the same graph.

When these two conditions (symmetry around solar noon and consistency of envelope) are fulfilled, the time stamp of the data bank can be considered correct, and the solar geometry can be precisely calculated. This test is very sensitive since a time shift of only a few minutes will induce a noticeable asymmetry.
4.2 Sensor calibration

Each sensor calibration must be verified using clear-sky data. For each day, the highest hourly value of GHI or DNI is selected from the measurements and plotted against the day of the year. These points are normally representative of the clearest conditions. For GHI, however, it frequently happens that higher-than-clear-sky values are obtained under partly cloudy, high-sun conditions. On such graphs, data from different sites or from different years for the same site can be compared. For DNI, the clear-sky index is defined as:

$$K_{bc} = \frac{B_n}{I_n e^{-M (\delta_{cd} + \delta_w)}}$$

(2)

where $\delta_{cd}$ is the broadband clean-and-dry atmospheric optical depth, $\delta_w$ is the water vapor optical depth, and $M$ is the air mass. These two broadband optical depths can be evaluated following [18]. The calibration correctness can then be assessed by direct comparison with data from a nearby site. In such a favorable case, it can be assumed that the two sites have similar atmospheric conditions, and thus similar clear-sky irradiances. Otherwise, this test can alternatively be conducted with the help of a clear-sky radiative model, if the atmospheric aerosol optical depth (AOD) and precipitable water ($w$) are known. The upper limits of the compared plots should then coincide. Alternatively, AOD and $w$ data may be retrieved from a nearby sunphotometer. For instance, the stations of the Aeronet network automatically retrieve AOD and $w$ at 15-minute intervals. In case $w$ is not measured in the vicinity, it can be evaluated from ground ambient temperature ($T_a$) and relative humidity ($RH$) through the use of an appropriately selected empirical model, whose validity must first be verified over the area of study. Alternatively, $w$ can be obtained from satellite data (such as MODIS) or reanalysis data. These alternate methods are approximate, but spatially extrapolating actual measurements also introduces errors, so that there is no perfect solution. These AOD and $w$ values are then used as inputs to a high-performance clear-sky radiative model (e.g., [19, 20]) to evaluate the clear-sky hourly $G_b$ and $B_n$ values.

For a quantification of the calibration factor correctness, the selected clear-sky hourly values between the modeled and measured data points are linearly regressed. This is illustrated in Fig. 3 for both GHI and DNI, again using data from Carpentras (where an Aeronet station is collocated with the radiometric station). In this case, the slope of the regression line is close to 1, suggesting a correct calibration.
4.3 Consistency

Consistency tests are conducted with the help of the global and beam clearness indices defined above. The hourly beam clearness index is plotted as a function of the corresponding global index. This is illustrated in Fig. 4 for Carpentras. Clear-sky irradiance predictions are also represented for four different a priori values of AOD. The corresponding Linke turbidity coefficient for an air mass of 2, $T_{Lam2}$, is then calculated from the value of $B_c$ thus obtained:

$$B_c = I_e e^{-M \cdot (\Delta_e + \Delta_o)}.$$  \hspace{1cm} (3)

The correspondence between $T_{Lam2}$ and AOD is also indicated on the graph. Any important deviation between the predicted and measured clear-sky values can be caused by a variety of conditions: calibration errors, pyrheliometer misalignment, soiled or shaded sensors, mis-categorization of clear-sky conditions, etc.

4.4 Statistical measures for benchmarking

Quality measures for solar resource products have been recently introduced under the European MESoR project [21]. Quality indicators are separated into “first order” and “second order” statistical parameters, which typically characterize the quality of solar resource products.

The “first order” measures involve the commonly used root mean square difference, RMSD, and mean bias difference, MBD. In the literature, these are usually expressed in absolute values (W/m²), percent, or both. For better intelligibility, particularly to non-technical people, the use of relative values expressed in percent is preferable. These values relate to the reference irradiance for each site under scrutiny, averaged over some
clearly indicated period. This will be the method of choice for reporting model performance, since this will be understandable by analysts in financial institutions, for instance.

Furthermore, it is clear that RMSD strongly depends on the time resolution of the analyzed time series. Up to now, most performance simulations of solar power systems were done with a 60-min time step. Newer satellite data and models now offer the option to derive irradiance outputs at 15-min time steps. Such predictions could be further processed by statistical or stochastic methods to derive data at 1-min intervals. The CSP industry now starts to realize that, to obtain realistic production results, high-performance simulation models must be run at much higher frequency than one hour. For instance, a dynamic model was developed [22] to simulate the most important transient effects of parabolic-trough plants, such as those typically being built in Spain. By feeding this dynamic model with meteorological input data at various time resolutions, it was found that the calculated electrical production of a CSP plant could be significantly overestimated when applying only hourly data, especially during partly cloudy days. For benchmarking of solar resource products, such results give a clear indication that irradiance datasets need to be analyzed at various time resolutions.

“Second order” statistics involve measures like the Kolmogorov-Smirnoff Integral (KSI), OVER or Combined Performance Index (CPI), which provide information on the quality of the frequency distribution of irradiance over time. It has been established [4] that the most important factor that influences the performance of CSP plants is the frequency distribution of DNI. These second-order statistics are therefore of paramount importance. The OVER and CPI statistics are both based on KSI, but consider only differences above a certain threshold [23, 24]. The OVER and CPI statistics have been specifically developed so that a single number could represent how well a frequency distribution of solar irradiance predictions agrees with a reference frequency distribution at a specific site.

It is recognized that more detailed tests are necessary to satisfy the industry needs. For instance, low-sun conditions are associated with low DNI and increased shading issues, and are therefore rarely utilizable under realistic CSP operating conditions. Such time periods can be advantageously filtered out, thus tailoring the benchmarking process to provide performance results optimized for CSP applications. For these reasons, we propose to filter out all data points for which \( h < 10^9 \). Similarly, most CSP systems cannot operate below some threshold value of DNI. The statistics would thus have more value if calculated only for DNI values above 250 W/m², which often is the value, at which CSP plants shut off operations.

5. Conclusions and outlook

In recent years, many new DNI-measurement stations have been installed worldwide with public or private funds. Measurement quality improves and data records get longer, while satellite-based modeled data are now offered in 15-min time steps, in addition to the conventional hourly frequency. Irradiance time series at even higher frequency could also become reality within a few years. In the future, it is thus expected that detailed CSP performance simulations will be done with 1- to 10-min time steps to improve accuracy under transient conditions due to clouds.

New DNI products are currently being developed. Especially for Europe, Africa and the Middle East, the Meteosat Second Generation satellite is now able to provide the necessary inputs to radiative models at 15-min time intervals. Preliminary validation studies that were conducted at a limited number of sites have shown that significant deviations could affect the modeled DNI predictions, particularly in areas like North Africa, the Middle East and Asia, thus confirming the importance of a full-fledged benchmarking study.

Several powerful statistical indicators will be proposed and used to characterize the performance of DNI time series for CSP applications. An advanced methodology will be used to inter-compare measurements and modeled time series, after detailed quality assessment of the measured data used as reference. Assuming this project will be properly funded, all interested data providers will be invited to deliver their datasets at various time resolutions, for areas where either public or private measurement sites are available. More sources of measured data (public or private) from sunny sites would also be an essential asset in this research study.

Acknowledgements

We thank SolarPACES for sponsoring and providing partial financial support for this work.

References


2. Ho C.K. and G.J. Kolb, 2010. Incorporating uncertainty into probabilistic performance models of con-


