SolarPACES Report

Standardizing and Benchmarking of Model-Derived DNI-Products
Phase 1

Richard Meyer (Suntrace GmbH)
Chris Gueymard (Solar Consulting Services)
Pierre Ineichen (University Genève)

March 2014

Document Status

SolarPACES Report

Report Title:	Standardizing and Benchmarking of Model-Derived DNI-Products - Phase 1
Authors:	Richard Meyer, Chris Gueymard, Pierre Ineichen

Document history

Version: 1

Drafted:	2012-09-08	
Reviewed:	2013	by SolarPACES review team
Revised:	2014-03-26	by authors
Approved:	2014	by SolarPACES Executive Committee

Acknowledgements

This report is the result of the project "Standardizing and Benchmarking of Satellite-Derived DNI-Products - Phase 1", which is funded by the IEA SolarPACES ExCo. It builds on results achieved during the course of SolarPACES Task V “Solar Resource Knowledge Management”, which is a collaborative Task with the IEA programme “Solar Heating and Cooling” (SHC). At SHC, this task is known as SHC Task 46 and managed by its Operating Agent, Dave Renné. We acknowledge the EC-funded project MESOR (European Union CA – Contract No. 038665), which provided valuable input to this report.

Summary

Modeled direct normal irradiance (DNI) can be either derived from satellite data or from numerical weather 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. A full-scale benchmarking study should now follow, assuming proper funding is secured.

For such a study it is proposed to take as many high-quality measurement stations as possible for reference purposes. The search for such stations preferably should focus on latitudes below 45°. This covers the regions where most of the CSP power is going to be deployed, but was not a focus in the few satellite solar radiation validation studies that have been published so far. Ideally, the measurement data to be used for validation purposes should not be in the public domain, to guarantee that these datasets have not been available to the model developers to train or adapt their models. Only this ideal scenario would fulfill the conditions for truly independent results, because otherwise the validation process would suffer from data incest. The rapid development of solar projects over many regions has sparked a proliferation of private weather/radiometric stations, and thus the existence of many "secret" datasets that could prove extremely useful for the proposed task. The great difficulty, however, is to find ways to motivate the data owners to make their exclusive data available for such a scientific study.

An other challenge is the anticipated arduous effort to thoroughly quality check the measured data series to be taken as a reference. In the authors' experience, this process (to be conducted a posteriori, without sufficient knowledge of the measurement conditions, etc.) is overwhelmingly the most critical issue. This report discusses several procedures to screen data automatically, but experience proves that manual and visual screening by experienced human analysts is absolutely necessary before any measured dataset can be considered a valid reference for such a validation. In particular, special attention must be given to the possible degradation of instruments through lack of calibration, maintenance or regular cleaning. Soiling, for instance, can rapidly make the data measured with the best instruments of little value, if not promptly detected and corrected by the radiometric station supervisor. All the issues just mentioned induce systematic biases in the reference data, and hence may lead to the wrong conclusions about the quality/accuracy of model-derived data.

Contents

1 Introduction

When energy yields of a concentrating solar thermal power (CSP) plant must be calculated, the beam irradiance—here called direct normal irradiance (DNI)—is the input parameter having the highest variability and uncertainty (Ho et al., 2009, Chhatbar & Meyer, 2011). In the SolarPACES project named guiSmo (Guidelines for CSP performance modeling) conducted by Eck et al. (2011) the availability of DNI is addressed in work package 8 on meteorological inputs, whereas DNI's accuracy is part of work package 5, which covers the topic of yield uncertainty.

Gueymard (2010, 2011) showed that, at potential CSP sites in Northern Africa, the deviation of annual sums of DNI from various modeled DNI data sets could reach up to 45%. Ineichen (2011) analyzed five satellite data products and validated them against 23 ground stations for IEA. That recent investigation can be considered as a limited-scale prototype of the more ambitious study that the present research task is now tackling.

The EU-funded project called MESoR (Management and Exploitation of Solar Resource knowledge; Hoyer-Klick et al., 2009) also initiated the process of benchmarking satellite-derived global horizontal irradiance (GHI) and DNI data. As further detailed in Section 2.3, various conventional or non-conventional statistics were defined as useful in such a context. A still open question is how important these statistics are in relation to the energy yield of CSP plants, and which ones need to be more critically studied in a benchmarking exercise.

It is estimated that around 10 different providers of modeled DNI-data sets are available today, but not all of them can provide data for all possible CSP-favorable areas in the world. Based on the experience collected during previous validation studies, such as those of Schillings et al. (2004) or Šúri et al. (2009)—none of which being as thorough as desirable for many reasons, however—it is possible to evaluate what could be an "exhaustive" benchmarking study, open to all data providers, and which would attempt to qualify their data independently. The main shortcoming of the above-mentioned studies is that the model developers have been directly involved in the validation studies. Although we assume none of them was doing any false representation, a self-evaluation tends to have a higher probability for unintended oversight of shortcomings. The ideal scenario of using previously unreleased measurement data as reference for validation, cannot be satisfactorily implemented if the modelers/developers are themselves directly involved in the validation process.

Some of the above estimated 10 potential providers likely could send various types or versions of their products—including data at various time steps, since these can be of interest in CSP applications. In addition to various integration times, various spatial resolutions could also be considered for the same product, because experience shows that a coarser resolution of around 10 km x 10 km leads to better correlation (with hourly values) than, e.g., high-resolution 1 km x 1 km data. Overall, and for each test site, it is expected that up to 50 different "products" would need to be inter-compared to measured data.

The investigation of available DNI measurement sites shows that data can be obtained at well over 100 worldwide. This number increases steadily because of new stations being commissioned or because of the very existence of unpublicized older stations being discovered. Most of these stations are at latitudes below 45°, and can be considered suitable for CSP projects in principle. In many cases, these data sets are available at time resolutions of 10-minute, or even 1-minute, and over several years. Therefore, the amount of data points to manage in any benchmarking activity can be quite extensive.

Overall, a serious international benchmarking of DNI data could require an intercomparison of around 5 000 different time-series. To manage such a large number of data sets it is essential that all model-derived data sets be delivered to the benchmarking team in well-defined file formats. An "ideal" file format for such DNI time-series data has still not been established. Therefore, potential file formats have been checked for suitability, and a recommendation is given in this report about how such data should be delivered for efficient processing. If the selected data format appears suitable to the energy simulation community, it could also be used as input data format for CSP performance simulation tools in the future.

This report does not contain information of the *actual* quality of DNI data. The present goal is to prepare an international benchmarking investigation of DNI data products, as a follow-on project with more substantial funding. To reach this goal, this report sets up criteria for the benchmarking methodology, proposes suitable standardized data formats, and reviews the availability of ground-based measurements that might serve as reference for validation and benchmarking. An important discussion is offered about how these reference measurements should be selected and quality checked. Finally, an evaluation is provided on how the actual benchmarking activity can be executed in the course of the anticipated follow-on research project.

2 Criteria for the benchmarking of modeled DNI products

This chapter describes the kind of model-derived products that should be benchmarked. It defines the criteria that should be applied and how the benchmarking process should be conducted to guarantee accurate and significant results to the CSP industry. The most important DNI product required by the CSP industry is an accurate time-series data set. Time-series data sets are typically used first at the level of pre-feasibility studies, where performance simulations are necessary to determine the basic layout of CSP plants. Typical Meteorological Years (TMY) are usually applied for this purpose. Such products are conventionally supplied with a hourly time-resolution, but some commercial vendors have started offering a 10-min resolution. At the next level (feasibility studies), more detailed engineering is done, for which higher time resolutions are needed, such as (ideally) 1-min resolution. This again can be done with TMY-type data. For thorough and bankable financial analyses, however, additional constraints exist because one needs to evaluate the impact of worst-case scenarios, not just "average" years, which imply long-term DNI data series of at least 15 years.

For the planning of an effective benchmarking of modeled DNI data, the number of quality parameters to be analyzed should be reduced to what is reasonable. For CSP applications, this means that one needs to determine which quality measures can give good indications about how well the predicted CSP yields will ultimately match the actual yields.

From the discussion above, it is clear that a variety of different DNI time-series products is necessary in the DNI industry. Therefore, it is important that all users of such data time-series know about their expected accuracy beforehand, and for each possible application. So far, only few studies have provided uncertainty estimates of modeled DNI data series, which is understandable since the prediction uncertainty is highly variable over space and time. The reasons for this are explained, albeit on a qualitative basis, by Cebecauer et al. (2011).

2.1 DNI time-series products to be benchmarked

2.1.1 Types of model-derived DNI

Modeled DNI-time series can be delivered by algorithms that calculate DNI based mainly on observations of the atmosphere from satellite-based sensors. These products are referred to as *satellite-derived DNI* in what follows. Alternatively, DNI may be derived from global or regional atmospheric weather or climate models, or interpolated from nearby ground-based measurements. Building on the respective strengths of all these methods, new products are being developed, in which satellite-derived DNI data are optimally combined with numerical weather models or chemical transport models to take into account the influence of atmospheric constituents, like water vapor or aerosols, which are hard to derive accurately with sufficient temporal resolution from weather satellites alone.

2.1.2 Spatial resolution

Detailed spatial coverage is an important asset in solar resource assessment. This is particularly the case over complex topography, because of the high spatial variability of DNI under such conditions. DNI data products exist with a variety of spatial resolutions: single spot (ground measurement station), 250 m x 250 m disaggregated satellite pixels, 3 km x 3 km nominal grid cells, large grid boxes like 1°x1° (NASA SSE), or interpolated point data on a variable grid mesh like that of the METEONORM software. Additionally, disaggregation techniques using Digital Elevation Model (DEM) data have been implemented to derive DNI data at much finer resolutions, reaching now about 90 m x 90 m, with detailed account of topographic shading (Ruiz-Arias et al., 2010). Although such a high resolution is desirable because it may result in better accuracy and may offer greater potential for improved CSP design, it is obvious that the concomitant steep increase in data volume might present critical difficulties in the context of a benchmarking exercise.

The higher the spatial resolution of the modeled data set, the higher the need for correct cloud localization. Currently, most satellite models still do not properly map cloud shadows because they do not consider 3D cloud scenes. More sophisticated models (e.g., Wyser et al., 2002) that would be able to take such effects into account have not sufficiently permeated the solar resource modeling community. Usually cloud shadows are mapped in the direct line of sight as seen from the satellite sensor, whereas the actual shadow might be in a completely different position. Unless by chance the Sun is approximately in line with the satellite, a large dislocation between the cloud shadow and the cloud position as seen from the satellite sensor occurs. This parallax error may lead to large deviations between satellite-derived and measured values, especially for time-resolutions of 60 min or finer. Thus, substantial noise is usually observed in scattergrams if cloudy scenes are involved.

In principle the parallax error issue can be solved. It requires a lot of effort, since the cloud bottom and top heights need to be known. Bottom heights can be estimated from Stüve-diagrams, while cloud top heights can be derived from the thermal infrared channels of weather satellites. Therefore, the envisioned benchmarking exercise should be able to accommodate sophisticated high-resolution products.

The spatial resolution issue must also be considered from the standpoint of the requirements imposed by the benchmarking exercise. All modeled data are compared to reference ground observations, which are frequently referred to as "ground truth" after they pass some quality control (QC) tests. This means that spatially-averaged modeled data are compared to spot measured data. If the grid cell area of the modeled data is large and the DNI field inhomogeneous (due to, e.g., complex topography), it is conceivable that the benchmarking results can be distorted. It is concluded that the highest possible spatial resolution in modeled data is desirable for benchmarking, despite the increased computer power needed.

2.1.3 Time resolution

In addition to its spatial variability, DNI has also a strong temporal variability, which needs to be taken into account. CSP applications have inherited the conventional use of 60-min data. In recent years, shorter sampling periods, such as 1-min data for ground measurements, or 15-min data for satellite-derived time series, have been introduced by data providers. Like with spatial resolution, a high temporal resolution is desirable in CSP applications, because of the need to take non-linear, transient processes into account (Hirsch et al., 2010). However, this adds practical computing difficulties in the preparation of meteorological input data because longer time-series also need to be evaluated without gaps. The same kind of difficulty obviously occurs on the CSP modeling side too.

A related issue is that of time coincidence between the modeled and measured data series. A major problem is that the latter nearly always contain missing periods. These can be easily dealt with (i.e., eliminated from the analysis) when using measured data at their original resolution. This becomes critical when time averages (e.g., hourly or monthly means) of measured data must be used, since they can contain a more or less significant fraction of missing data. This, again, can distort benchmarking results. From this standpoint, it appears preferable to benchmark data at the highest temporal resolution possible, despite the increase in computing power needed. However, this may also create some unwanted experimental artifacts. This is related to the fact that DNI reacts extremely rapidly to the passage of small thick clouds. There can be a temporal offset between the actual spot ground measurements and the modeled data averaged over a grid cell area. This offset statistically tends to disappear when the time step is increased. It is concluded that there is currently no well-established "ideal" time resolution to adopt in benchmarking exercises, which means that this issue deserves more up-front scrutiny.

2.2 Performance indicators for single time-series

Benchmarking of solar radiation products can be done in different ways. If high-quality reference (presumably measured) data are available, various modeled data sets can be compared and ranked according to how well they represent the reference data. For site-specific time series there are several possible benchmarking indicators, some more conventional than others. The most conventional indicators are first-order statistics, which include the mean bias *MB,* the root mean square deviation *RMSD,* and the correlation coefficient *R.* These statistics are defined as follows:

$$MB = \frac{\sum_{i=1}^{N}(G_{model} - G_{ref})}{N},$$

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(G_{model} - G_{ref})^2}{N}},$$

$$R = \frac{\sum_{i=1}^{N}(G_{model} - \overline{G_{model}})(G_{ref} - \overline{G_{ref}})}{\sqrt{\left(\sum_{i=1}^{N}(G_{model} - \overline{G_{model}})^2\right)\left(\sum_{i=1}^{N}(G_{ref} - \overline{G_{ref}})^2\right)}}$$

where *G* stands for the irradiance component being evaluated (e.g. DNI or GHI), the index *model* indicates model-derived (data to be benchmarked), whereas *ref* indicates the corresponding reference—typically the measured values. *N* is the number of samples, for which coinciding valid reference and modeled data are available. $\overline{G_{model}}$ symbolizes the average of the modeled irradiance values over all *N* data points. In the expressions above, *MB* and *RMSD* are expressed in irradiance unit (W/m^2). They are alternatively expressed in percent of the average reference value. This way, these statistics can be efficiently compared to the uncertainty of the reference data, and can be easily understood by non-experts in radiometry or solar resource assessment.

RMSD depends primarily on random errors, and is used for that reason, but also depends to a lesser extent on systematic deviations, which are primarily accounted for by *MB*. This quality parameter is thus not characteristic of a single type of errors. To characterize random errors only, the standard deviation, *SD*, of the difference between two data sets appears better than *RMSD*. SD can be calculated by the simple relation

$$SD = \sqrt{RMSD^2 - MB^2}.$$

The two most conventional types of error (bias and random) are addressed in more detail in Sections 2.2.1 and 2.2.2, respectively.

The main goal of the envisioned benchmarking exercise is to give the CSP industry advice on the quality (or "performance") of different DNI data sets. To best support this goal it would be valuable to have performance statistics that are directly, or even linearly, related to the output of CSP systems.

Due to various reasons, there is no direct relation between the incident DNI and the generated power, at least for most CSP designs. For instance, the heat produced by parabolic troughs is more linearly related to an *effective* DNI, defined as DNI weighted with the cosine of the sun zenith angle, Z. Following Meyer et al. (2009), the effective DNI can be expressed precisely as

$$DNI_{eff} = \left(\sqrt{1-\sin^2 Z \cdot \cos^2 A}\right) \cdot DNI \qquad (1)$$

where A is the sun azimuth angle, measured clockwise from North.

By definition, DNI_{eff} is the fraction of the incident irradiance that can actually be converted into usable heat by parabolic troughs. Considering DNI_{eff} rather than DNI as usual has the advantage that it is much closely connected to energy yields (see Fig. 1).

As the thermodynamic response of a parabolic trough CSP plant is fairly linear to DNI_{eff}, the latter seems to be a better irradiance component to be benchmarked—at least for parabolic trough plants that are aligned North-South.

So that the planned CSP benchmarking exercise can equally support all CSP technologies, the quantity DNI_{eff} would need to be calculated separately for each type of technology. Unless analytical or empirical expressions similar to Eq. (1) are proposed in the literature, it appears simpler to use the existing DNI data without correction, while considering that low DNI values have much less interest than higher ones. This is important because it is much more difficult to correctly predict low DNI values than high ones. Since most CSP systems do not operate below some minimum value of DNI, or “threshold”, a non-conventional performance analysis based on non-zero thresholds would be well tailored for CSP applications.

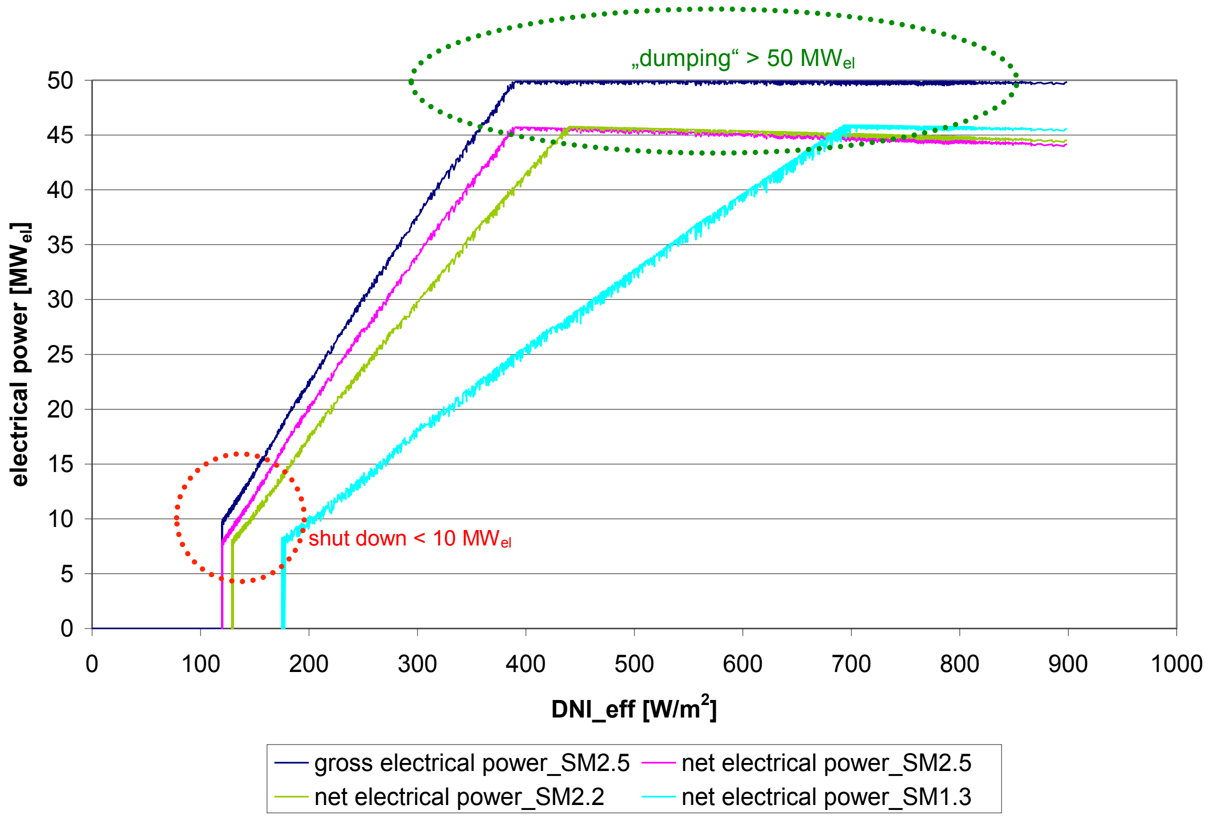


Fig. 1: Transfer function of effective DNI into electric power for various solar multiples (SM) of a parabolic trough plant derived from simulations over a full year (source: Meyer et al., 2009).

Among the remaining questions that would require further investigation, the relationship between cumulative distribution statistics and the accuracy of modeled energy yields of CSP plants appears high on the list. The proper selection of the most critical statistics to evaluate and analyze in a benchmarking exercise is important because the number of quality parameters should be reduced as much as possible to draw clear conclusions. In other words, the statistics with the most information potential and the most powerful leverage on CSP outputs should be defined, and only those should be used to avoid clutter and information glut. For CSP applications, this means that one has to determine what the best performance indicators are to assess the accuracy of DNI data, i.e., indicators that would unambiguously describe how well the predicted CSP yields based on modeled DNI data would match the actual yields—assuming the power plant itself is perfectly well simulated.

Any other specific requirements of CSP on DNI data shall be considered. For this purpose, the input from the industry and from the SolarPACES Task I activity guiSmo for standardizing performance simulation models would be desirable.

2.2.1 Characterizing systematic over- or under-estimations

The role of bias errors in DNI predictions is particularly well understood because they have an obvious effect of similar magnitude on the energy yield of CSP plants. A plant's energy output over some period of time is roughly linearly related to the total DNI during that period. It can be assumed, for instance, that an increase of 1% in annual DNI roughly leads to a similar increase in yield. Therefore, the key importance of a statistic like *MB* is well established. Such a cause-to-effect relationship is not as clear for other statistics.

Many DNI datasets tend to show relatively high deviations from measurement for small DNI values. On an instantaneous basis, low DNI values are typical of either low sun under a clear sky, or of a passing cloud. These conditions are of no practical significance for CSP. However, as the data time step increases, things become more complicated: On a hourly time scale, for instance, there could be one hour composed of 20 minutes of high DNI (e.g., 900 W/m^2) under clear conditions, and 40 minutes of no DNI due to the passage of thick clouds. These 20 "clear" minutes, which result in a hourly average of 300 W/m^2, would be fully usable by a CSP installation. In parallel, the same hourly-mean DNI value could conceptually also be a roughly constant value, if associated with e.g. a continuous thick haze. This hourly period would not be usable if the CSP system's threshold is above 300 W/m^2. This idealized example shows that, even though low hourly (or daily) DNI values may not a priori appear of value for CSP, a deeper understanding is necessary to better characterize the applicability and quality of data series. Considering the higher modeling uncertainties in the presence of clouds (and particularly, broken clouds), and the existence of technology-dependent DNI thresholds under which the accuracy of DNI is irrelevant, a solution to the low-DNI issue would be to analyze data at high frequency, i.e. sub-hourly time steps. This is not always possible, however, since many modeled or measured datasets are only available on a hourly basis.

An explanation for the existence of a technology-specific threshold is that CSP plants have a relatively high parasitic load, roughly in the range of 10%, combined with thermal losses that are approximately independent of DNI, and with a low turbine efficiency under part load. Due to the compact structure of central receiver plants, their thermal losses should be smaller than those of parabolic trough plants. In contrast, linear Fresnel plants should have higher thermal losses than parabolic trough plants, since they need a larger solar field and currently have a less efficient thermal insulation. Parasitic loads are mostly related to the pumping of HTF and to cooling towers, especially air cooled condensers (ACC). Compared to parabolic trough systems, central receiver plants do not have an outspread HTF system, while experiencing higher efficiencies due to higher temperatures. These are other reasons why they can normally utilize lower irradiances than parabolic troughs. From this discussion, it can be anticipated that DNI thresholds should be relatively low for central towers, average for parabolic troughs, and largest for linear Fresnel collectors.

Since different plant designs result in different DNI thresholds, various thresholds would have to be considered in the analysis. However, it seems impractical to consider several different such thresholds because this multiplies the number of performance statistics. A typical DNI threshold of 250 W/m^2 has been mentioned by several authors (e.g., Meyer et al., 2009; Montez et al., 2009). To take the effect of CSP technology on threshold into consideration, it is suggested that two minimum values, such as 200 and 400 W/m^2, be used in the analysis to keep it manageable while providing sufficient insight. The conventional case (no threshold) would still have to be considered (particularly with low-frequency data), because this can help tremendously in assessing the exact effect of the threshold (by interpolation/extrapolation), and in comparing with existing or future validation results, since the proposed threshold analysis is a new methodology.

2.2.2 Characterizing random deviations

SD and *R* show how well data pairs compare with each other at any given moment. They are important if one needs an exact representation of real data, e.g. for continuous temporal evaluations of real operating systems or forecasts of solar radiation components. This exact match is not necessarily important, however, for system design in particular.

2.2.3 Characterizing deviations in frequency distributions

The EU-funded project called MESoR (Management and Exploitation of Solar Resource knowledge; Hoyer-Klick et al., 2009) developed procedures intended for the efficient benchmarking of satellite-derived global horizontal irradiance (GHI) and DNI data. An important (and yet unverified) assumption was that the frequency distribution of irradiance values plays a role for the calculation of energy yields. Therefore, non-conventional statistics, like the Kolmogorov-Smirnov Integral (KSI), were proposed by Espinar et al. (2008) to supplement the classical measures mentioned in sections above.

Although there are several statistical tests and ways of evaluating the performance of a model, the KSI test has the advantage of making no assumption about the data distribution, and is thus a non-parametric, distribution-free test.

The interest here is that the cumulative distribution functions (CDF) of the modeled and the reference data sets can be compared, and their distance objectively evaluated. The CDFs are purposefully binned into *m* intervals, then the local distance *Dn(m)* is computed for each interval. The KSI test ultimately provides a measure of overall conformity between the two CDFs.

Another disadvantage is that KSI depends largely on the chosen bin size to determine it. A small bin size, like steps of 20 W/m^2, tends to be influenced by the inherent noise caused by sampling artifacts, which occurs with some data sets. However this noise may well average out if the bin size is increased to e.g. 100 W/m^2, but this may be too large for precise evaluations of CSP yields. The optimum bin size for such an analysis would have to be determined, using test cases or other means.

For general radiative model validation, where low irradiance values may have their importance, complete CDFs are normally used and compared (e.g., Gueymard, 2012). In the case of CSP applications, however, low irradiance conditions under some fixed or variable threshold are of limited or negligible interest, as discussed earlier. Therefore, it is suggested that the CDFs of modeled datasets be truncated appropriately to ignore DNI data below a specific threshold.

2.3 Summarizing benchmarking results over many sites

The above described quality parameters need to be derived separately for each site. These are also the typical parameters obtained for individual site assessments using model-derived data in combination with measured data. Usually the model-derived historic time-series are more complete than the measurements and cover longer time periods. However, the goal of a DNI benchmarking is to derive summary results for a number of solar resource data products, which might cover different time periods.

Ultimately, what really matters is the quality of a selected solar resource data set at any site where no measurement is available. An uncertainty estimate can be obtained by averaging statistical results over several sites. The relevant result is mainly the average of e.g. the MB over all comparable sites. However, the standard deviation is also valuable because it gives an indication of variability. Ideally, both this variability and the bias should be low, but this is rarely the case. For instance, it might turn out that at all sites a model is found to overestimate by a significant 10 %, whereas its standard deviation is only ≈1 %. In such a case, it would be very reasonable to do a systematic bias correction at all sites in comparable climates. If, conversely, the average MB is only 1 %, but variability is around 10 %, it becomes unlikely that a simple bias correction alone could improve this data set. The high standard deviation indicates that the average is composed of many significant under- and over-estimations, which in this example almost average out over a sufficient time period.

It is recommended to calculate such averaged values and their respective standard deviations for all quality parameters, which are found to be of value:

- mean bias MB,
- standard deviation SD,
- root mean square deviation RMSD,
- KSI or similar for frequency distribution testing,
- CSP-performance-weighted DNI performance statistics.

This should be done over all stations first. Then, the averages of all these performance statistics would be obtained over several sites within a given region. Finally, their standard deviation would provide an indication of uncertainty in each statistic. Since it is expected that the behavior of some (if not all) models is climate-dependent, and even possibly season-dependent, another reasonable segregation of results would be based on climatic region and/or season. Due to the wider pixel sizes and grazing viewing angles towards the edges of a spaceborne sensor's field of view, separate summaries for different categories of satellite viewing angle could also help identify the strengths and weaknesses of some data sets.

The identification of regions for such an exercise can be a somewhat subjective decision. Because of the usual low density of ground observations, larger—rather than smaller—regions will generally have to be defined. In many cases, this could have the adverse consequence of climatic inhomogeneity within the region. This again shows how important it would be to overcome the challenge of obtaining high-quality measured data for a large number of sites.

Additional insight on the relative performance of models vis-à-vis aerosol and cloud extinction might be obtained by separating clear and cloudy situations. Similarly, testing the performance of modeled data over highly reflective surfaces can be informative.

3 Available ground-based DNI-measurements

This chapter reviews the availability of DNI measurements wherever application of CSP makes sense. Since such 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°)
2. A large number of stations must be considered to obtain results of high significance
3. They should be well spread geographically and climatically over all regions of interest
4. Their data quality must be as high as possible, which means they must be obtained with good instrumentation, using high standards for measurement method, calibration, maintenance, quality control, etc.
5. As much as possible, they 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.

The latter criterion 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 any third party. 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 is 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 potential for fulfilling 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. These authors 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. It is possible that some datasets of this nature could be made available to us for a fee, if for example they have been collected at sites that have *not* been selected for any project construction. Access to such data would be great, but this eventuality obviously depends on the budget that will be attributed to such purchases. In any case, two lists of available stations have been prepared for this report: the "public" and the "private" stations. Their location is shown in Figs. 2 and 3, respectively. As of this writing, 185 public stations and 64 private stations have been selected. 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. These figures clearly show 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. This indicates that Criterion 3 will be difficult, if not impossible, to fulfill.

It is hoped that, in the next few years, new countrywide measurement programs (like the World Bank's ESMAP initiative) will be fund the installation of many more stations. This would be important especially over regions like Northern Africa and the Near East or Southeast Asia, where only few DNI stations currently exist.

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 suggests 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. Various QA techniques exist, and are reviewed in Section 5. Since a rigorous QA cannot be completely automated, it is quite time consuming. Due to specific station design, instrumentation and local conditions, the techniques used may vary from site to site. No QA was attempted at this stage of the study due to stringent time and budget limits. Therefore, not all stations pre-selected here may ultimately be used as ground truth.

The type of instrumentation used at a radiometric station has direct bearing on the quality of the measured data that can ultimately used as ground truth. Some instruments perform noticeably better than others, even among the limited group of expensive research-class instruments. The performance of instruments in the field does not always confirm the manufacturer's specifications. There is now specialized literature about the field performance of various types of radiometers (e.g., Gueymard and Myers, 2009; Michalsky et al., 2005, 2011; Myers, 2010; Myers and Wilcox, 2009), which can be used as guidelines to evaluate the likelihood that a given radiometric station can be used as a good source for reference data. Additional resources about the ideal setup of radiometric stations for CSP applications can be found in NREL's *CSP Best Practices Handbook* (Stoffel et al., 2010).

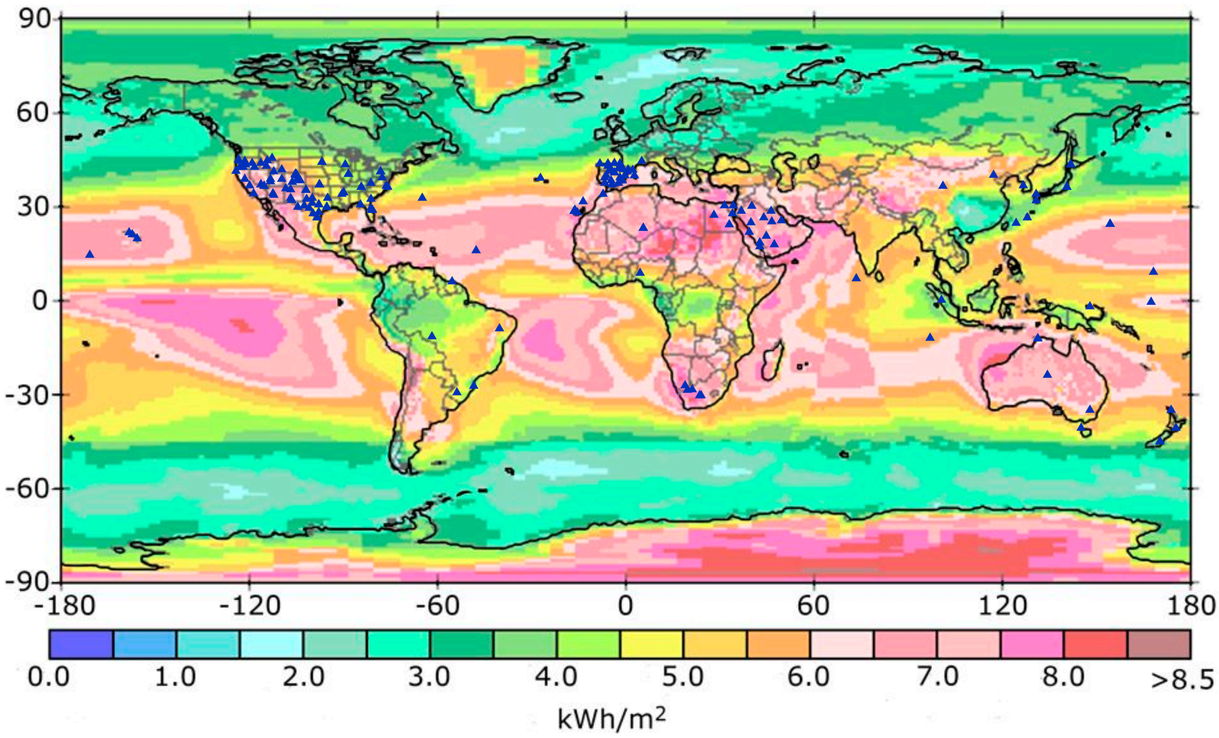


Fig. 2: Location of public stations (blue triangles), superimposed on a map of mean annual DNI based on the NASA-SSE dataset.

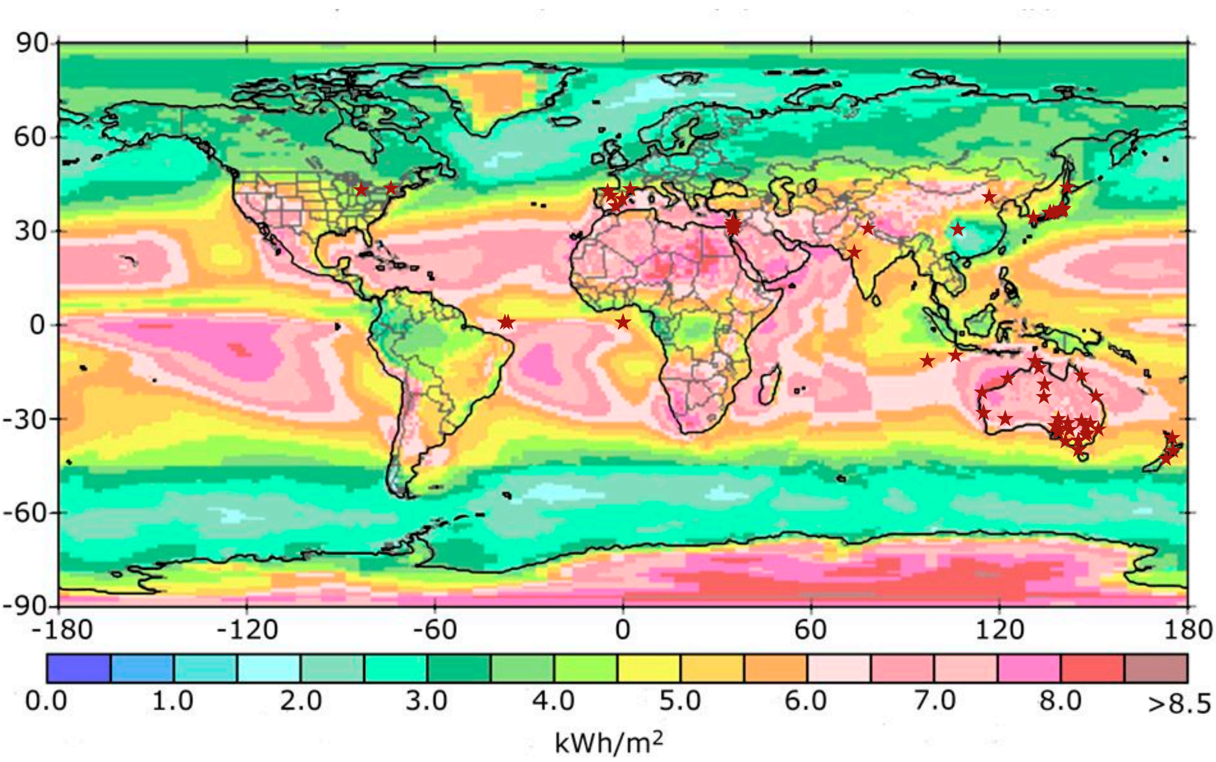


Fig. 3: Same as Fig. 2, but for private or non-public stations.

There are many discussions among experts, and some publications (e.g., Geuder et al., 2006, 2009; Gueymard and Myers, 2008, 2009; Michalsky et al., 2011; Wilbert et al., 2010), 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 unfortunately is often not the case. The downside of such instruments is that they are expensive and have stringent requirements: relatively high power consumption for the trackers and ventilators, and frequent cleaning requirements (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 (Gueymard and Myers, 2009).

To avoid the high costs and potential issues of thermopiles, many recent radiometric stations, particularly those installed by or for CSP developers, include radiometers constructed from solid-state silicon sensors. Such instruments are now called Rotating Shadowband Irradiance sensors (RSI). RSIs are much less expensive, and have much lower power requirement and a much faster response time than thermopiles, but they have drawbacks too: (i) they only measure two radiation components (global and diffuse) independently, from which DNI is calculated through the fundamental closure equation; and (ii) their spectral response is not constant with wavelength, and covers only a part of the solar spectrum. A suite of empirical corrections must therefore be introduced a posteriori to bring their measurement's quality close to that of thermopiles (Geuder et al., 2008, 2009, 2010).

Field experience has shown that RSIs need less frequent cleaning than thermopiles (Geuder and Quaschning, 2006). This can be a decisive advantage for remote stations. The difficult issue that third-party users of data regularly face is that it is virtually impossible to 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 thus often considered that a rotating shadowband instrument with a regular 7-day cleaning cycle should provide better-quality data than a thermopile pyrheliometer with such infrequent cleaning. With all these considerations, we have opted to include stations equipped with both types of instrument.

When thermopile instrumentation is used, an 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). Separate measurements provide redundancy, and most importantly a simple way to test whether the closure equation is respected or not, within some tolerable limits. Not all stations are in this ideal case, unfortunately. When DNI is not measured directly with a pyrheliometer, or indirectly derived from a well calibrated and corrected RSI, it is customary to evaluate it from the direct horizontal irradiance (DHI), calculated as the difference between GHI and DIF (see, e.g., Gueymard 2010, 2011). This indirect method is not as accurate as the normal one, but could be considered at a few stations located in key areas.

Stations being considered for our database must have a time interval of one hour or less. Many research-class stations now have a time increment of one minute, which can be useful to validate those modeled data that exist at 10-min or 15-min intervals, for instance.

All stations have gaps in their data, and this 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, such as 5 or 10% maximum, should be discarded and flagged accordingly. The validation of data series available only with a daily or monthly time increment is dangerous since they 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 not considered here, since most modeled databases only start after that date.

Some details on the largest networks follow.

3.1 BSRN

The Baseline Surface Radiation Network (BSRN) is a component of the Global Energy and Water Cycle Experiment (GEWEX) of the World Meteorological Organization (WMO) World Climate Research Programme (WCRP). The current repository for the measured data is the World Radiation Monitoring Center (http://www.bsrn.awi.de/). Since BSRN's inception in 1992, 52 stations have reported at least some data. A few more stations have the status of "candidate" and have not started reporting data yet. The official BSRN map showing all these stations appears in Fig. 4.

Fig. 4: Map of existing or projected BSRN stations.

It is obvious from Fig. 4 that not all BSRN stations are of interest here, since many are at too high latitudes. For this study, 39 stations fulfill Criterion 1. The main advantage of BSRN stations is that they all must follow very strict protocols to insure data of the best possible quality (Ohmura et al., 1998), thus perfectly fulfilling Criterion 4. They also measure irradiance at rapid intervals, often as frequent as 1-minute. However, this is also probably the most well-known network for quality irradiance data, so that it can be assumed that all modelers and radiation data providers use BSRN data to validate and/or correct their data, as the literature shows. Criterion 5 might therefore not be fulfilled in all cases.

3.2 GAW

The Global Atmospheric Watch is the atmospheric chemistry programme of WMO, which was established in 1992. Only a small number of GAW stations include radiation measurements, which do not always include DNI. The official GAW map for DNI measurements is reproduced in Fig. 5. Information on GAW stations can be found at http://gaw.empa.ch/gawsis/, while the available radiation data are stored as both hourly and daily data series at the World Radiation Data Center, http://wrdc.mgo.rssi.ru/. Some stations are member of both BSRN and GAW, so there is some overlap. Only five distinct GAW stations that both fulfill Criterion 1 and provide DNI data have been identified.

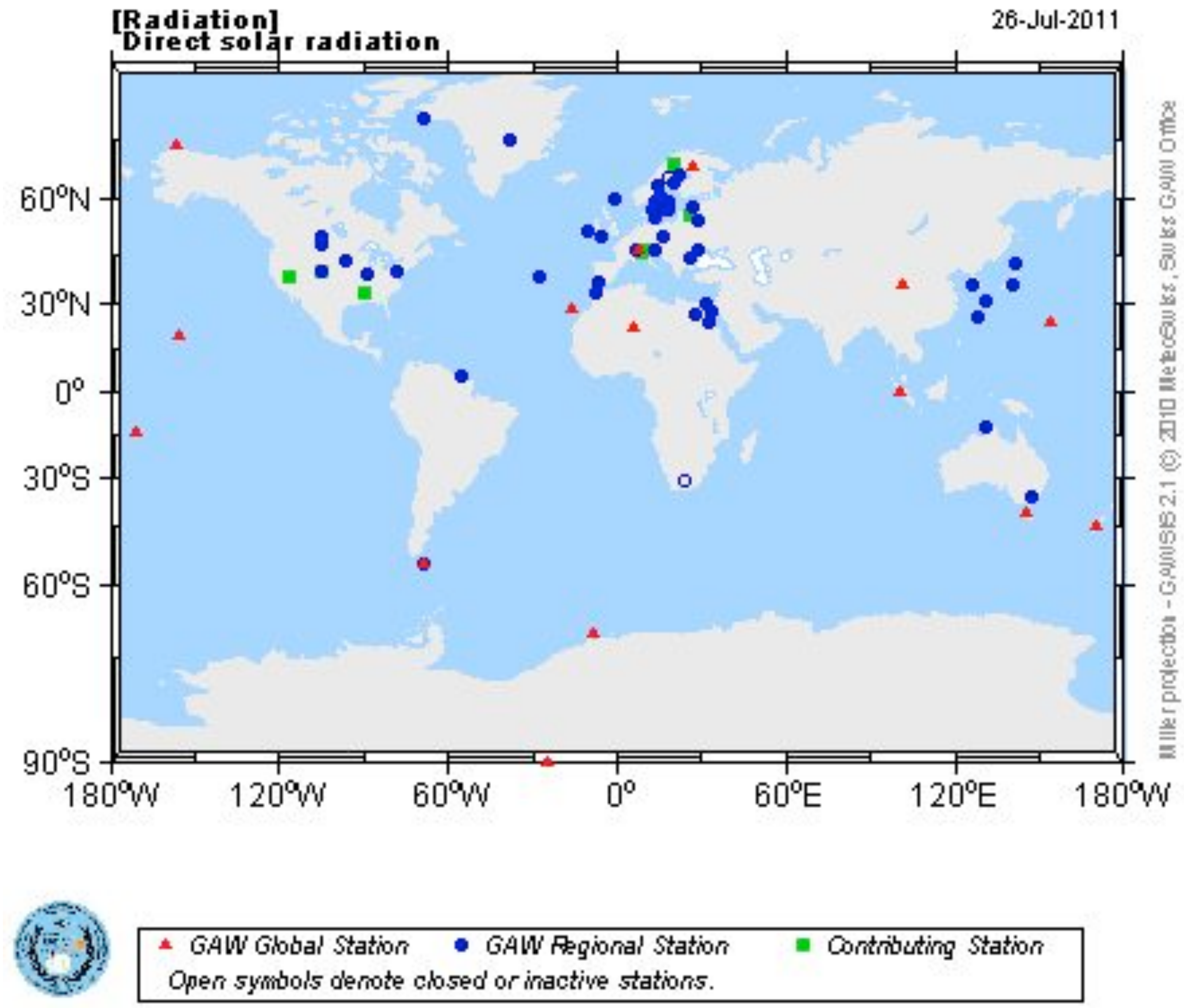


Fig. 5: Map of existing GAW sites with DNI measurements.

3.3 ARM

The Atmospheric Radiation Measurement (ARM) program of the U.S. Department of Energy (http://www.arm.gov) was created in 1989, and has become a key contributor to international research efforts related to global climate change.

The main ARM location is in the U.S. Southern Great Plains, in parts of Kansas and Oklahoma, where operations started in 1992. A cluster of about two dozen sites with three-component radiometric measurements exists there, over an area of 143,000 km^2. Two of these sites are contributing BSRN stations. Considering the relatively short distance between these sites and the good homogeneity of the terrain, including all these sites in the database for actual benchmarking tests might not be highly productive and should not be first priority (especially when considering Criterion 3 and the effort needed to quality control each dataset). Related studies on the spatial variability of DNI over homogeneous terrain could benefit from these datasets, however.

Other temporary or permanent sites were added in various parts of the world, including China and the Pacific. One station in the Pacific is also contributing to BSRN.

Fig. 6: Map of ARM sites.

3.4 NREL

The National Renewable Energy Laboratory has established an experimental measurement station, known as SRRL, at Golden, Colorado in 1981. Over the years, this station has steadily grown from a few radiometers to many dozens, while the NREL staff developed state-of-the-art calibration techniques, now used by other institutions such as ARM or BSRN. The data collection started on an hourly basis, but is now on a 1-min basis (since 1999). In recent years, RSIs were added to the pool of instruments, thus providing the possibility to compare them with conventional thermopiles. All data from that site is available from http://www.nrel.gov/midc/srrl_bms/. Many other sites have been instrumented later, some with thermopiles and some with RSIs. The measurements periods vary by site, but all these observations are available from a central portal: http://www.nrel.gov/midc/.

Datasets from other U.S. sites partnering with NREL are available from http://rredc.nrel.gov/solar/new_data/confrrm/ and http://rredc.nrel.gov/solar/old_data/hbcu/mv/. See also Section 3.6.

3.5 NOAA

The U.S. National Oceanic and Atmospheric Administration maintains two networks, called SURFRAD and ISIS. The former, with high-quality instrumentation and experimental protocols, counts 7 stations, which are all part of BSRN. Their data can be accessed either from BSRN or from http://www.srrb.noaa.gov/surfrad/. The ISIS network is less sophisticated than SURFRAD, and is suspected to have experienced quality problems during periods of budget restrictions. The data is accessible from http://www.srrb.noaa.gov/isis/. Finally, a few additional NOAA stations are not part of these networks. Three of them (Mauna Loa, Hawaii; Trinidad Head, California; and American Samoa, Pacific), are of interest here and their data have been obtained by personal communication.

3.6 Saudi Arabian Solar Radiation Measurement Program

In cooperation with the King Abdulaziz City for Science and Technology (KACST) in 1995 NREL set up 12 precision stations (Fig. 7) in the Kingdom of Saudi Arabia (KSA) (Maxwell, et al., 1999). Wilcox (1996) adapted QC routines to execute regular quality assurance to the 5-min data. However, according to oral communication (Steven Wilcox with Richard Meyer, August 2012) besides the main station at Solar Village near Riyadh, which was also appointed as a BSRN station, frequent cleaning of the radiometers has been an issue. The radiometers are all high-quality thermopile instruments. However the tracking and shading of the pyranometer dedicated to diffuse radiation was realized by a single-axis tracker with clock setting through the AC power connection. Therefore, in addition to soiling problems tracking errors had been frequent.

The measurements stations have good availability and reasonable quality mainly for the years 1998 to 2001. Schillings et al. (2004) used this data to validate DLR's SOLEMI satellite scheme. Except for Solar Village, these datasets have not been widely used outside of NREL, from which they are available (http://rredc.nrel.gov/solar/new_data/Saudi_Arabia/). However, the rapid development of solar projects over the region has recently prompted some data developers to use the NREL data for validation, fine tuning and promotion.

Two new networks are now being developed in KSA. One is developed by K.A. CARE under NREL's supervision, and will ultimately group dozens of stations, with either thermopiles or RSIs. Another network is being developed by Saudi Aramco. The availability of such data for our purposes is not yet known. The main issue here, as with any new or projected network, is that only very short series of data will have (hopefully) been collected at the time of the proposed study.

Fig. 7: Map of historic measurement sites in Saudi Arabia operated by NREL for KACST.

3.7 Solar radiation measurement network of the Spanish AEMET

Spain is currently the country with the largest number of installed CSP power plants in the world. Thus, of course, DNI measurement is of high interest in this country. The Spanish national weather service, Agencia Estatal de Meteorología (AEMET), maintains 35 radiometric stations, 22 of which measure DNI. Almost all (32 stations) also measure diffuse in addition to GHI. Thus, if the AEMET data could be made available for a benchmarking exercise, Spain could well count the largest number of validation sites. However, the distribution of the AEMET stations (Fig. 8) is concentrated in the North, whereas the southern regions are those where most CSP plants are installed. The stations erected close to the Atlantic coast or on the Canary Islands could likely be challenging for satellite-derived procedures, due to the higher cloudiness and often inhomogeneous satellite pixels.

Fig. 8: Radiometric stations with DNI measurements operated by the Spanish weather service AEMET.

Currently, the observations are partly made public with a hourly time resolution, and appear online once a week as part of a rolling 7-day archive. The older data are then removed from AEMET's website. For the contemplated benchmarking activity, AEMET would have to be officially contacted so as to make the data set available and to confirm that the data may be used for validation purposes.

Due to the fact that the data are spooled over the continuous rolling archive it must be assumed that some of the satellite-data provider take advantage and train their models with these measurement data. Then the benchmarking would be biased for those models. It is known that most of the CSP projects operate radiometric stations for qualification and monitoring of their projects. Getting access of these measurements would have the advantage that these are usually not available to satellite data providers. However, most CSP project developers regard these data as confidential.

In addition to AEMET, a few public institutions (universities or research centers) maintain radiometric stations. Such data are normally in the public domain, and could therefore be used without any problem.

3.8 Solar Radiation measurement stations in India

The Indian Meteorological Department (IMD) operates hundreds of automatic weather stations throughout the country. A few of them are equipped with pyrheliometers as can be seen in Fig.

9. Most, if not all, pyrheliometers were of the ancient Ångström model type (for laboratories), an d thus used only a few times a day manually (without a tracker) and weather permitting (clear skies). Availability and quality of this source of pyrheliometric and pyranometric data would have to be better known.

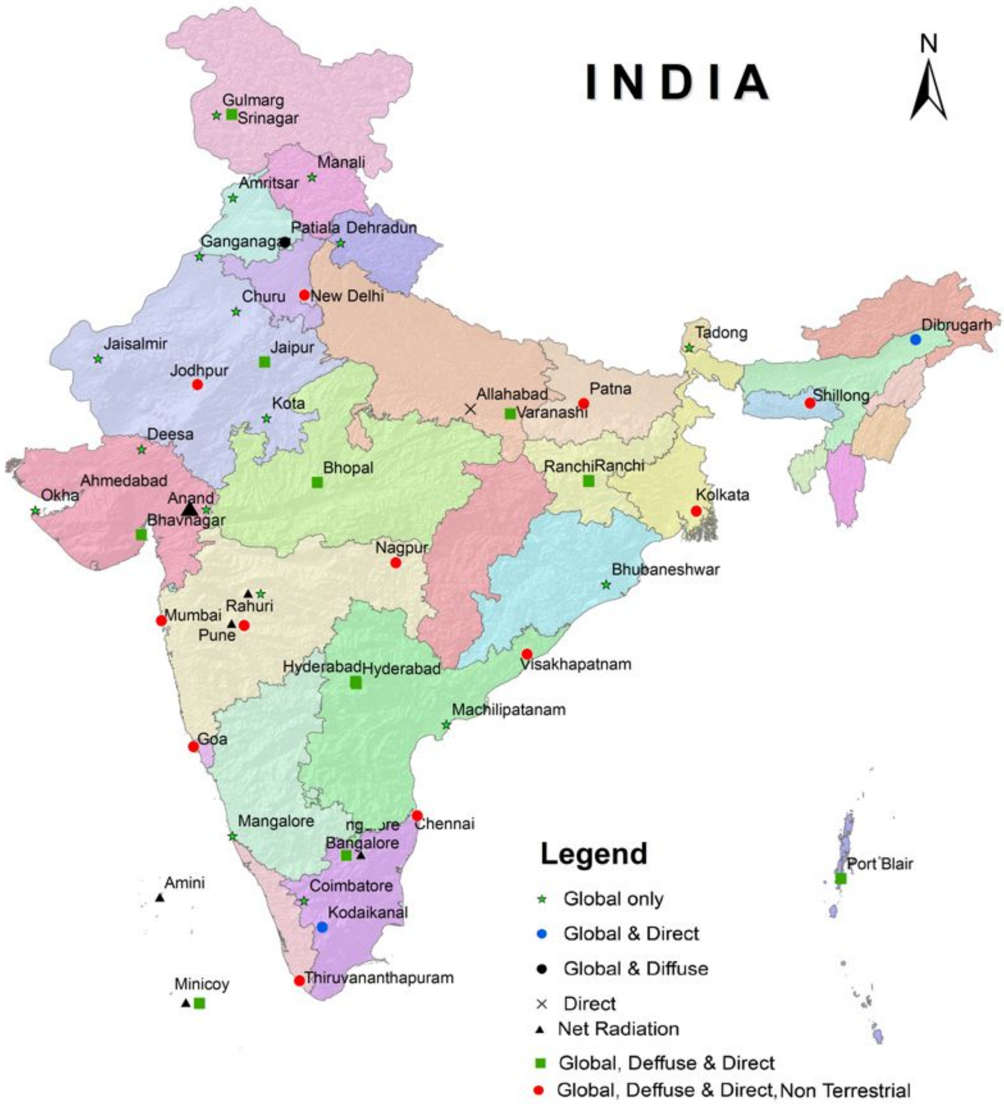


Fig. 9: Solar radiation measurement stations operated by the Indian Meteorological Department (IMD). Some are equipped with pyrheliometers.

In 2011 the Indian Ministry for New and Renewable Energy (MNRE) commissioned the Solar Radiation Resource Assessment (SRRA) system, which consists of 51 precision stations spread over the sunniest Indian states (Kumar et al., 2012). The selection of locations for the 51 stations (Fig. 10) followed the intention of the MNRE to best cover those regions, which are most promising for large-scale solar energy systems. As a result, SRRA provides good coverage of the northwestern states of Rajasthan and Gujarat and of the southern part of the country. Due to the latitude range from 9°N (for the southernmost station) to almost 35°N in the Ladakh Himalaya region, all stations are in the 'Sunbelt' range, which is highly interesting for DNI benchmarking.

All 51 stations are equipped with Eppley thermopile sensors consisting of one pyrheliometer and a shaded and an unshaded pyranometer. Thus, it is expected that this measurement system can provide reliable data. To what extent the normal cleaning and maintenance schedules have been followed rigorously at all sites remains to be verified. At the end of 2013, the second phase of SRRA started with the aim to erect an additional 64 such precision stations.

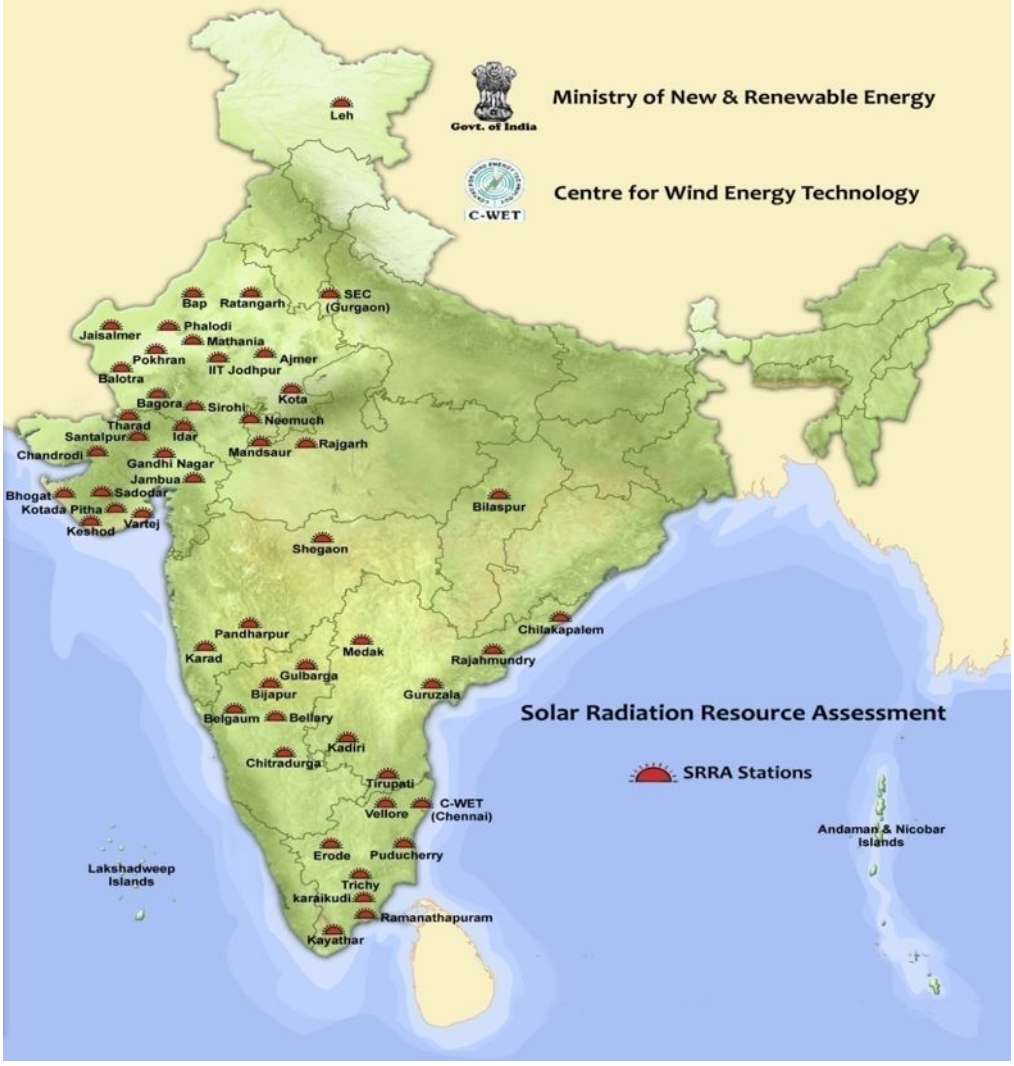


Fig. 10: Map of SRRA measurement sites.

3.9 Negev radiation survey

A network of nine radiometric stations started operation in the Negev desert in the 1980s. The characteristics of these research stations, and some aggregated results have been compiled by Faiman et al. (2004). One of these stations, Sede Boker, is also part of the BSRN network. The hourly data is not directly available, but has been obtained by personal communication with the authors of the aforementioned paper. Since such data are not publicly available, they appear in Fig. 3.

3.10 Measurements from the Australian Bureau of Meteorology (BOM)

The National Australian Bureau of Meteorology (BOM) has started operating around 15 high-quality solar radiation measurement stations in 1993. Due to budget reasons, however, 6 of these stations had to stop operations in mid 2006.

Two of the stations operated by BOM are also BSRN stations (Cocos Island and Alice Springs), and their data is thus publically available. These sites are also equipped with additional instruments, which make them of higher value for developing satellite procedures.

The other datasets are not publically available, and thus appear in Fig. 3, but can be obtained on request for research purposes. Since BOM joined the IEA Task 46 in 2011, it is expected that these datasets could be made available for future benchmarking exercises. The BOM stations are well spread over the continent (Fig. 11), which make them highly valuable for benchmarking. In particular, five stations in Western Australia and on Cocos Island could be used to validate the performance of Meteosat-derived solar radiation products. This part of Australia can be viewed by Meteosat-5, which has been operating on 63°E since mid 1998. All of Australia is also in view from the Japanese GMS geostationary meteorological satellite series. Thus these BOM data sets would be of special value to inter-compare modeled irradiance data series derived from two different satellite systems.

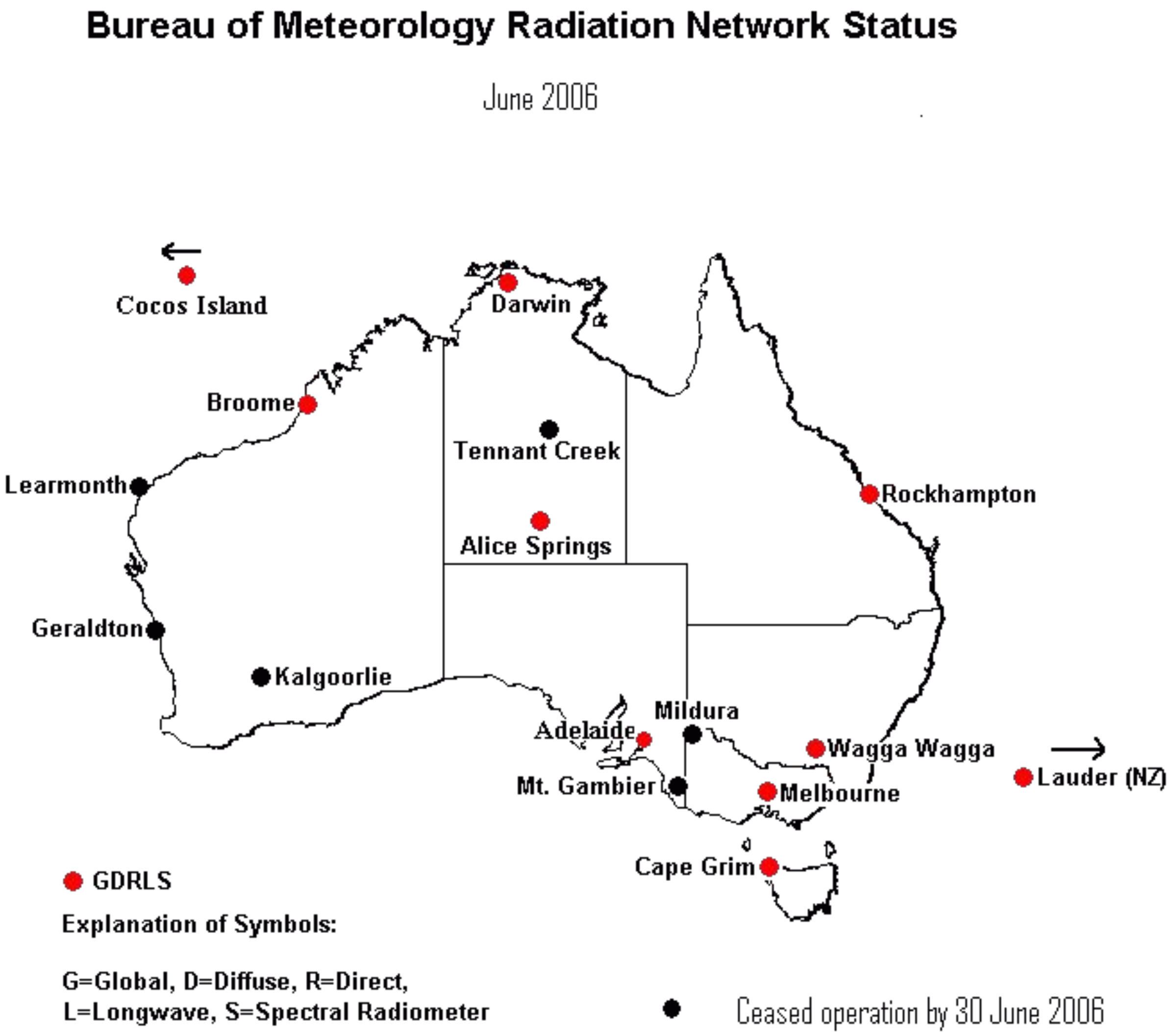


Fig. 11: Network of solar radiation stations of the National Australian Bureau of Meteorology.

3.11 Sites available from non-public stations

Data from a number of radiometric stations have been obtained (or could be soon obtained) by these authors through personal communication with various research groups around the world. Such datasets are not publicly available, but can be used internally for research purposes. The nine stations mentioned in Section 3.9 have been considered in this category.

Data from stations of private solar energy project developers also might be an option. Usually these are very hard to access since developers regard their data as highly confidential. But there are now many stranded CSP projects, whose historic weather/radiometric data are not of commercial value any more. Additionally, existing data from bankrupt companies might be obtained through personal contacts. However, private measurement campaigns are not always of high quality standard. Therefore, before accepting such data for validation purposes they would have to be screened very carefully.

4 Quality control of the reference measurements

Valid reference data are the pre-condition for any benchmarking. For the purpose of benchmarking of model-derived DNI products, mainly DNI measurements from state-of-the-art solar radiation instruments should be considered. Since measuring DNI is a difficult exercise, experience tells that many stations—despite being equipped with good instruments and operated by established organizations—show serious errors. Therefore, thorough quality control (QC) or quality assessment (QA) is essential before starting any validation process. All doubtful reference data must be clearly flagged and may not be used for benchmarking.

Of course, even thoroughly checked high-quality solar radiation measurements have uncertainties. The goal of this QC task is to assign individual values of uncertainty to each reference dataset, or partial dataset. This may be a single average uncertainty value, like a fixed percentage or amount in W/m^2, depending on the type of instrument used and the maintenance they receive. The actual uncertainty of irradiance measurements at any given site may be quite variable over time, which complicates the issue. Uncertainty may change with sun elevation, cleaning conditions or time elapsed since the last re-calibration. Only a rough estimate of the uncertainty can be made without a lot of details on the instrument model, calibration history, maintenance, cleaning schedule, etc. In most cases, only the instrument model is known, so that a generic uncertainty flag based on judgment might have to be used.

Sensor calibration is the key point for precise measurements in the field of solar radiation. Before the beginning of the acquisition period, all radiation sensors should be calibrated by comparison against a sub-standard of the World Radiometric Reference (WRR) kept at the PMOD/WRC (Physikalisch-Meteorologisches Observatorium Davos / World Radiation Center). Then, at least every year or every other year, a fresh calibration should be performed against a sub-standard instrument. Due to a variety of possible errors, a precise calibration correction (or *virtual* calibration, based only 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.

It is important to note that no *a posteriori* QC/QA procedure can detect all acquisition problems that could have happened. Trying to correct erroneous data based on such QC tests is often too uncertain to be useful. Therefore, keeping the best operating conditions during the complete measurement period is by far the best option (Stoffel et al., 2010). Unfortunately, it is our experience that this ideal situation is the exception rather than the rule.

To help with automated QC/QA during the measurement period, or later on as the preliminary step of a benchmarking or validation exercise, appropriate software is extremely useful (e.g., Anderberg, 1999; Anderberg and Stoffel, 1998; Anderberg et al., 2002; Long and Dutton, 2002; McArthur, 2005; NREL, 1993; Wilcox, 1996: Wilcox and McCormack, 2011).

The main quality elements to be assessed in historic datasets are:

- Closure equation: how close the three radiation components respect it if they are measured separately;
- Measurement's time stamp (needed to compute the solar geometry);

- Sensor calibration coefficients used to convert the acquired data into physical values;
- Consistency between the direct and global irradiance components.

4.1 Closure equation

If the three irradiance components—DNI, GHI and DIF—are measured with separate thermopile radiometers, a fundamental consistency test can be applied, based on the closure equation that should link them:

$$G = G_b \sin(h) + G_d \tag{2}$$

where G stands for GHI, G_b stands for DNI, G_d stands for DIF, and h is the solar elevation. The left-hand side of this equation represents what an unshaded pyranometer senses. The right-hand side represents the combination of readings from a pyrheliometer and a shaded pyranometer, which is actually the most accurate way to measure GHI (Gueymard and Myers, 2009). The conventional way of measuring GHI with a pyranometer has been shown to present various shortcomings, so that this measurement should be used only for quality tests and redundancy (in case DNI or DIF data points are missing). This technique is applied at research-class sites, but assumes that (i) the shading of the pyranometer sensing DIF is done with a tracking shade (rather than a fixed shading band); and (ii) this pyranometer is as insensitive as possible to the thermal imbalance effect (Gueymard and Myers, 2009; Michalsky et al., 2005).

A temporal analysis of the deviation from its ideal value (1.0) of the ratio $G / [G_b \sin(h) + G_d]$ is extremely instructive. This analysis can be done during clear days to detect tracking, shading or leveling issues, any type of transient problems, or to evaluate above what sun elevation the setup's idiosyncrasies make the data less reliable. An example of results pertaining to a research-class station with top-notch caretaking appears in Fig. 12, using two different (but collocated) setups of instrument combinations with 1-min records. Setup #1 is composed of instruments considered among the best in their category. Setup #2 consists of popular instruments that equip a large fraction of stations worldwide. It is obvious that that the percent closure error is more pronounced with Setup #2 than with Setup #1, even though both datasets are of the highest intrinsic quality. This indicates unequal instrument performance (Gueymard and Myers, 2009). For instance, a ±5% closure error when $h > 5°$ can be considered acceptable. A series of finer tests of this kind is implemented in the BSRN QC procedure (Roesch et al., 2011). If the observed closure does not respect the promulgated error limits, the measurement data and consequently the referring model-derived DNI values must be eliminated.

The same type of analysis should be repeated on a daily basis, for all clear days included in the measurement record. This is helpful to detect trends in cleaning or calibration deficiencies, in addition to all other problems just mentioned. The interest of this detection technique is illustrated in Fig. 13 for a station in the mountains of Saudi Arabia, which was part of the network described in Section 3.6, and was equipped with Setup #2. In contrast to Fig. 12, a series of problems occurred during the 6-day period illustrated here, leading to closure errors frequently outside of the ±5% limits.

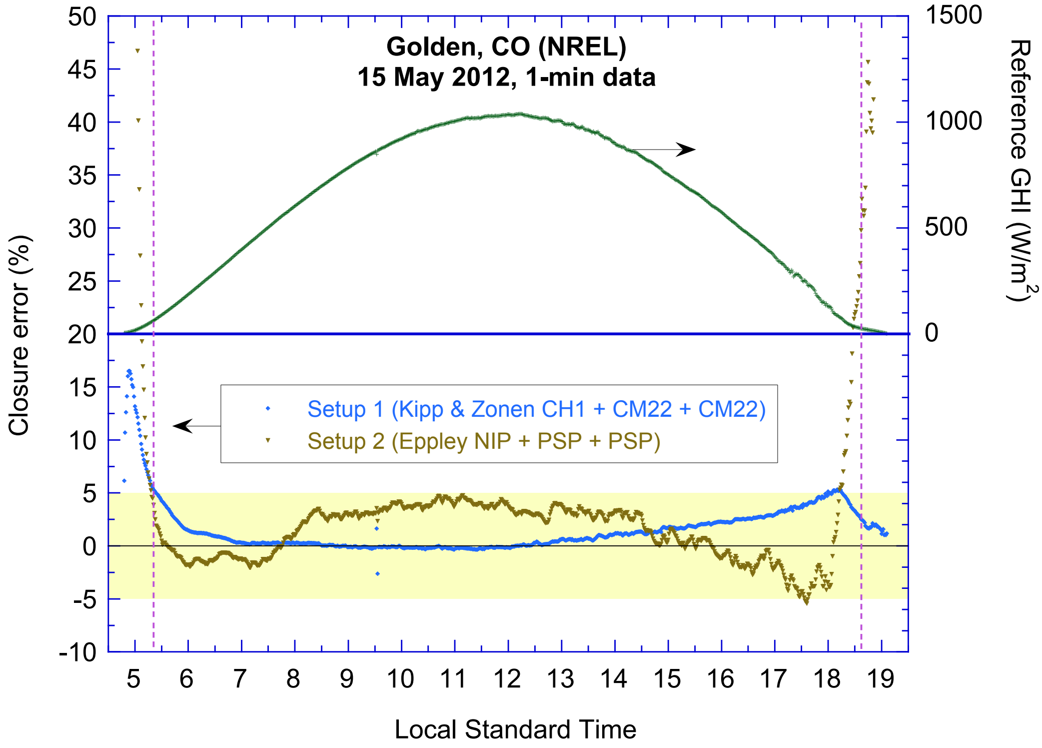


Fig. 12: Closure error during one cloudless day at Golden, CO.
The yellow rectangle indicates the acceptable ±5% limits for the closure error.
The vertical dashed lines indicate when the sun elevation was 5°.

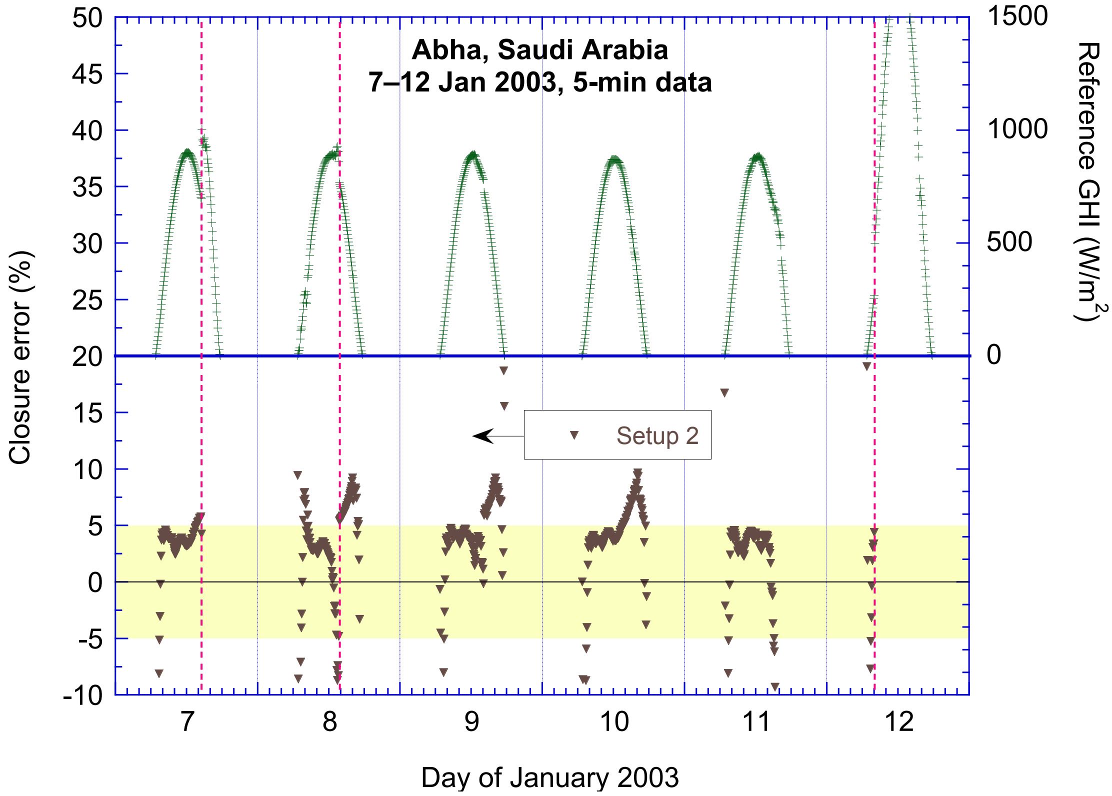


Fig. 13: Closure error during six cloudless days at Abha, Saudi Arabia. The vertical red lines indicate obvious experimental problems

4.2 Time stamp

For reliable benchmarking of satellite-derived DNI data it is highly important to fix any time errors in the measured data. Otherwise, benchmarking results referring to high time-resolution data would be falsely reporting low performance of the satellite data. In principle, satellite-derived data show much less time-stamp errors because the reference time used for satellite operations is in almost all cases strictly UTC.

For any measured data point, a correct time stamp is essential to evaluate the solar position accurately, and to properly assign the measured values. Time shifts are frequently observed in measured data, and can result from

1. Wrong clock settings,
2. Acquisition system clock drifts,
3. Usage of civil time instead of local standard time or UTC (this often causes trouble in countries where daylight savings time is enforced),
4. Inconsistencies in irradiance integration over time. Following WMO standards, hourly data summarizing or averaging should be done over the 60 minutes of any specific hour, and the corresponding hourly result should be assigned to the full hour with a time stamp set at the end of the considered hourly period. Not all stations follow this rule, however. For instance, the BSRN time stamp is set at the onset of a record period, whereas GAW is assigning hourly values to the middle of the hour.

To detect any possible time shift in the data, the clear-sky irradiance symmetry (with respect to solar noon) should be checked, as automatically as possible. The first possible method for checking the time stamp intrinsically from the measurements itself is to derive the time of the maximum irradiance on clear-sky days. Since even on perfect clear sky days there are small DNI fluctuations due to turbulence or undetected thin cirrus clouds, the time of the absolute maximum irradiance should be derived from a curve fit, which averages out these fluctuations. The method can be applied to both GHI and DNI. Application to DIF is not recommended because it is even more sensitive to fluctuations. DNI and GHI derived solar noon should lead to very close solar noon times, if both readings are recorded on the same data logger, which is the case for most of the stations.

The second method for finding the time of solar noon is to plot GHI and DNI values versus the sine of solar elevation angle *h* for clear-sky days. If the time stamp is correct, the afternoon curve should lay over the morning curve, as visualized in Fig. 14. 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 systematically and significantly lower in the afternoon than in the morning.

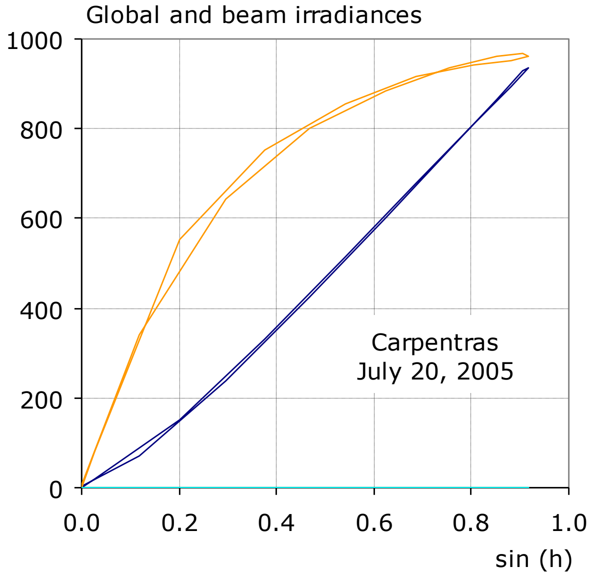


Fig. 14: G and G_b versus the sine of the solar elevation angle for a clear day at Carpentras, France.

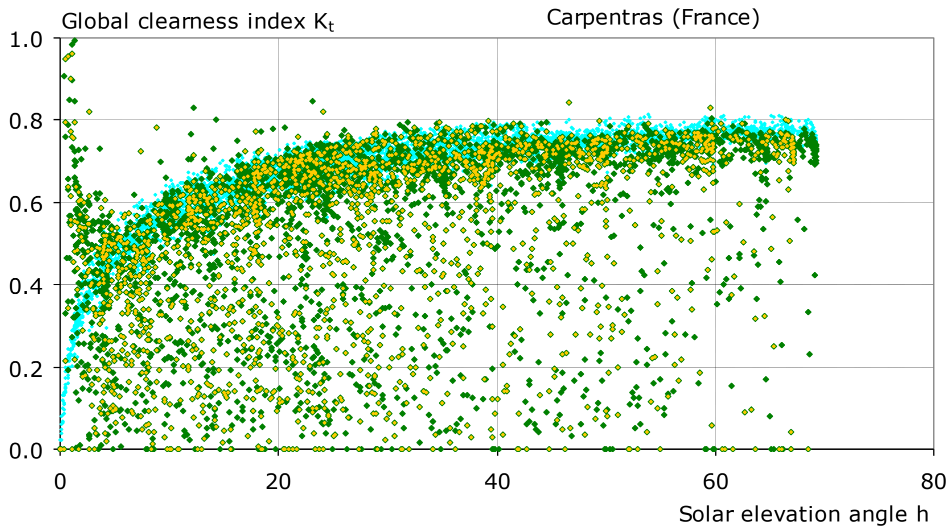
The third possible verification can be done with the help of the clearness indices K_t and K_b, which are respectively defined from:

$$K_t = \frac{G}{I_o \cdot \sin(h)} \quad (3)$$

$$K_b = \frac{G_b}{I_o} \quad (4)$$

where 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 after-noon periods separately. The upper envelope of the morning and afternoon data is representa-tive of clear-sky conditions. An example appears in Fig. 15, based on one year of GHI data ac-quired at Carpentras (France). Ideal hourly clear-sky values, calculated with the Solis model (Mueller et al., 2004; Ineichen, 2008) are plotted in blue on the same graph.

When these two conditions (symmetry around solar noon and consistency of envelope) are ful-filled, 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 conduct to a visible asymmetry.



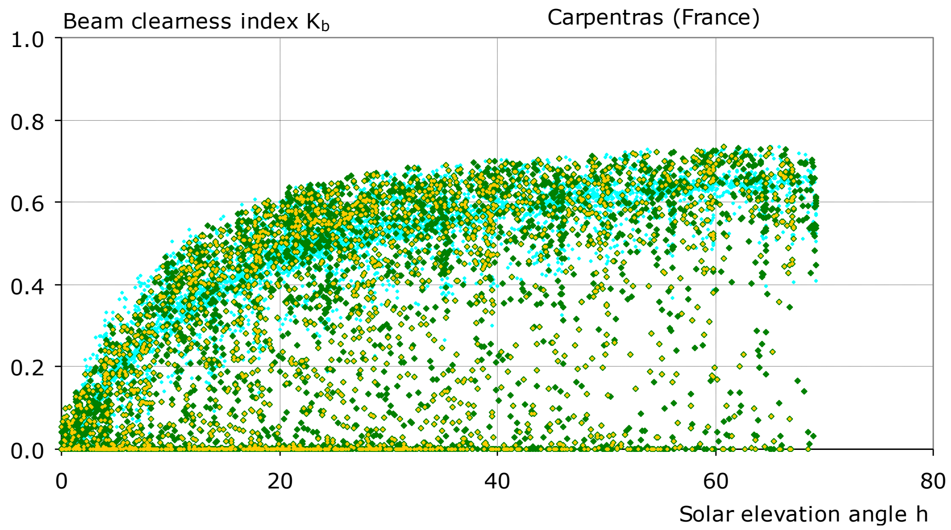


Fig. 15: K_t (top plot) and K_b (bottom plot) represented separately for the morning (green) and the afternoon (yellow) data, versus solar elevation angle for one year of hourly values at Carpentras. Corresponding clear-sky modeled data are represented in light blue.

The first method described above is normally the easiest one to automate. Thus, it is proposed to use this method to derive the maximum irradiance as an automatic check and method of correction. The two other methods can be used for double checking of the results. Finally, the time stamp of all time-corrected time-series need to be adjusted from solar time to UTC. Overall, the effort needed for these tests and corrections is significant. Experience proves that many problems of this nature exist, even with datasets from reportedly high-quality measurements, which makes this QC step all the more important.

4.3 Sensor calibration and cloud enhancement

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 measured time series and plotted against the day of the year. These points are normally representative of the clearest conditions. Under partly cloudy, high-sun conditions, GHI is frequently higher than theoretical clear-sky values, and sometimes even higher than the extraterrestrial value during brief moments. Such situations occur when broken clouds are present, and are usually referred to as cloud enhancement. This is usually caused by intense downward scattering from nearby bright clouds of high vertical extension in addition to high direct irradiance (high sun "trapped" between clouds). Another reason can be intense backscattering, if the albedo of the surrounding surface is high (e.g. snow on ground) and bright clouds around the station but not masking the sun.

Such cloud enhancement effects are temporary, and depend on cloud type, geometry, surrounding surface albedo, and topography. Their frequency depends on wind speed and cloud dynamics. They are obvious when analyzing high-frequency (e.g., 1-min) data, but can still be noticed in hourly data. Figure 16 shows the DNI and GHI traces recorded at very high frequency (3-second) by an RSI at Oahu, Hawaii, during a summer day with scattered cumulus clouds (a typical sky condition there). Around noon, GHI reached 1654 W/m^2, i.e., 25 % more than the extraterrestrial value at that moment (marked as ETHI on the graph), or 68 % over the clear-sky DNI value. Cloud enhancement can also affect DNI, but to a much more limited extent, since only the additional scatter from cloud sides within a 5°-cone can increase DNI. This effect on DNI might not be noticeable on a hourly basis. For instance, for the conditions of Fig. 16, the peak DNI decreases from 1053 W/m^2 when using 3-sec data to 913 W/m^2 with 15-min data, and 896 W/m^2 for hourly data. This averaging effect is much more pronounced in the case of GHI: its peak decreases from 1654 W/m^2 with 3-sec data to 1051 W/m^2 for 15-min data and 904 W/m^2 for hourly data. Another consequence of the cloud enhancement effect is that the usual statistical relationship between K_b and K_t can be significantly distorted in high-frequency data.

GHI data points affected by cloud enhancement may appear incorrect, even though they are actually real, and thus should not be identified as erroneous. To avoid incorrect flagging of such data points, a broken-cloud situation flag needs to be developed, using the standard deviation or rate of change of GHI over ±2 hours, for instance. If this broken-cloud parameter indicates such situations, the threshold value (beyond which an error flag is produced automatically) should be increased for both GHI and DIF.

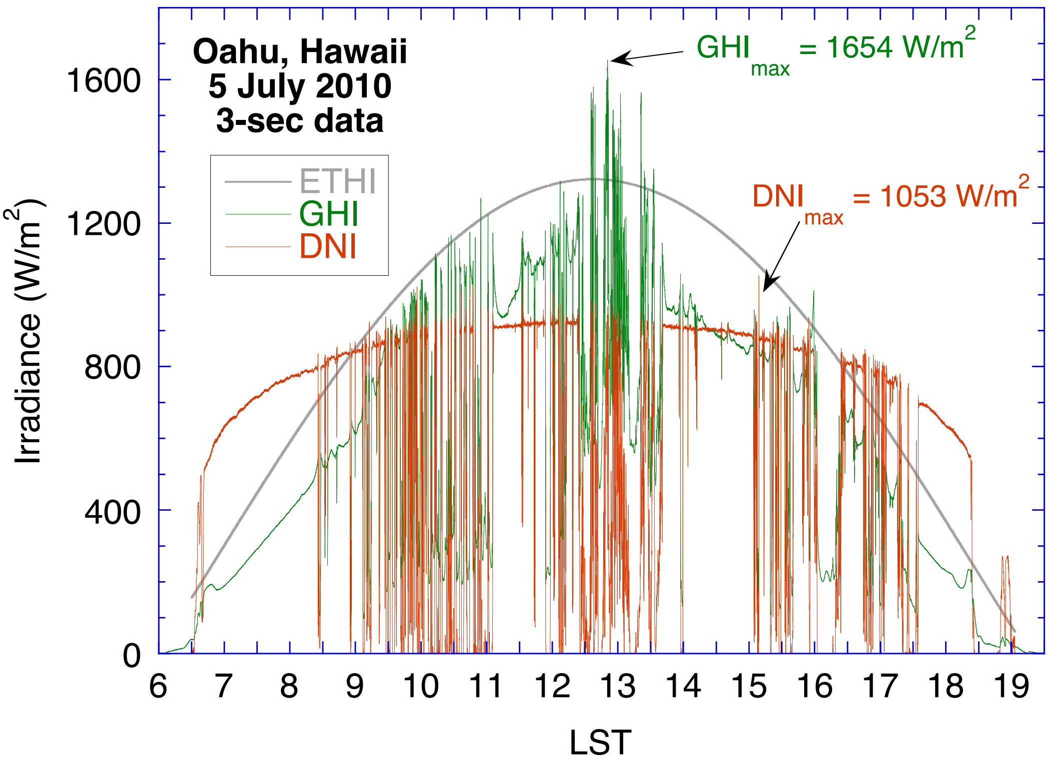


Fig. 16: 3-second records of GHI and DNI at Oahu, Hawaii during a partly cloudy sunny day. ETHI is the calculated extraterrestrial horizontal irradiance for that day.

For proper sensor calibration tests, only those days identified as unaffected by strong cloud enhancement (based on the broken-cloud parameter just mentioned) should be used. Clear-sky data from different sites or from different years for the same site can then be compared on a common basis. Additional care must be devoted to verify that no enhanced aerosol loads, e.g. due to volcanic eruptions or dust storms, is biasing the data during specific years.

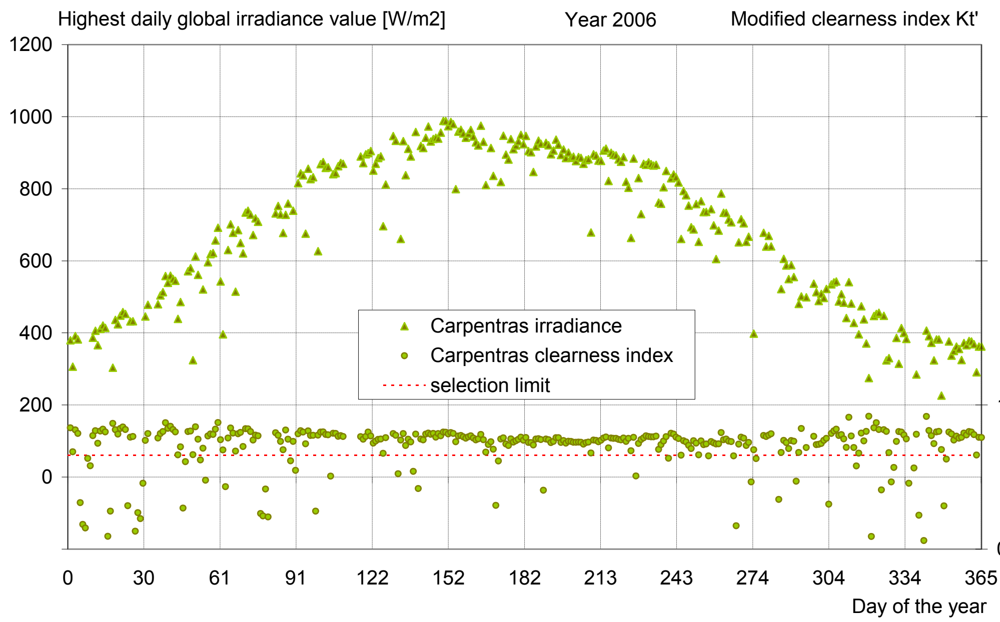
The GHI graphs can be augmented by superimposing the modified clearness index K_t', which was defined by Perez et al. (1990) as:

$$K_t' = \frac{K_t}{\left(1.031 \cdot \exp\left(-1.4 / \left(0.9 + 9.4 / AM\right)\right) + 0.1\right)} \tag{5}$$

where *AM* is the optical air mass (a pure function of *h*) as defined by Kasten (1980). This modified clearness index has the advantage of being relatively more independent from *h* than K_t. Therefore, it is possible to delineate three K'_t zones to characterize the sky condition (Ineichen et al., 2009):

- Clear-sky conditions $0.65 < K'_t \leq 1.00$
- Broken cloud conditions $0.30 < K'_t \leq 0.65$
- Dense cloud conditions $0.00 < K'_t \leq 0.30$.

Time series of the different clearness indices at Carpentras are shown in Fig. 17, using only values for which $K_t' > 0.65$.



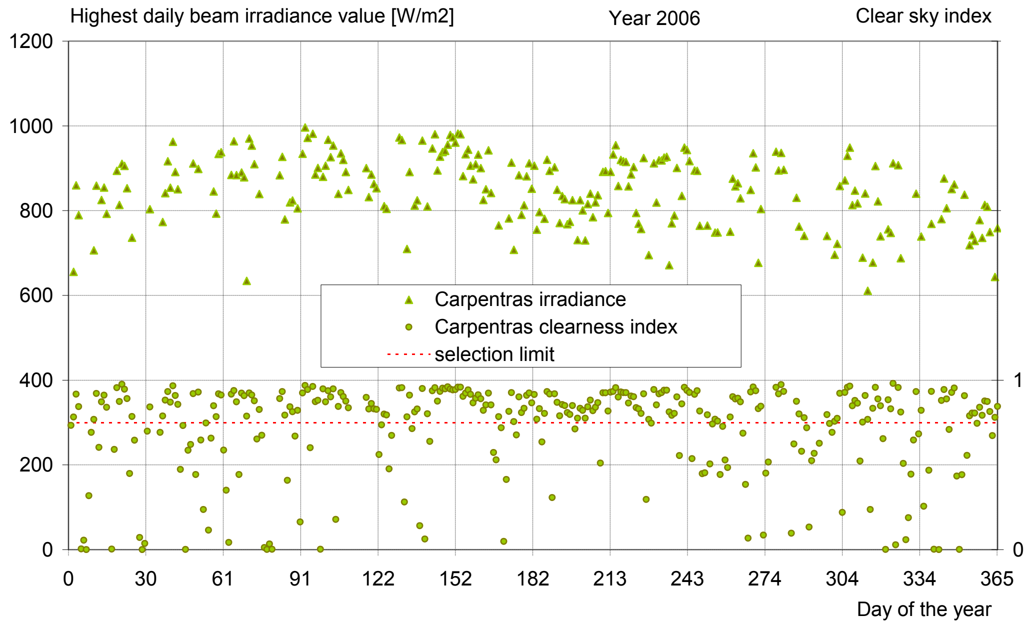


Fig. 17: Highest daily value of GHI (top) and DNI (bottom) versus the day of the year at Carpentras. The corresponding modified clearness index and clear-sky index are also represented.

For DNI, the clear-sky index, K_{bc}, is defined as:

$$K_{bc} = \frac{G_b}{I_o\, e^{-AM\cdot(\delta_{cda}+\delta_w)}} \tag{6}$$

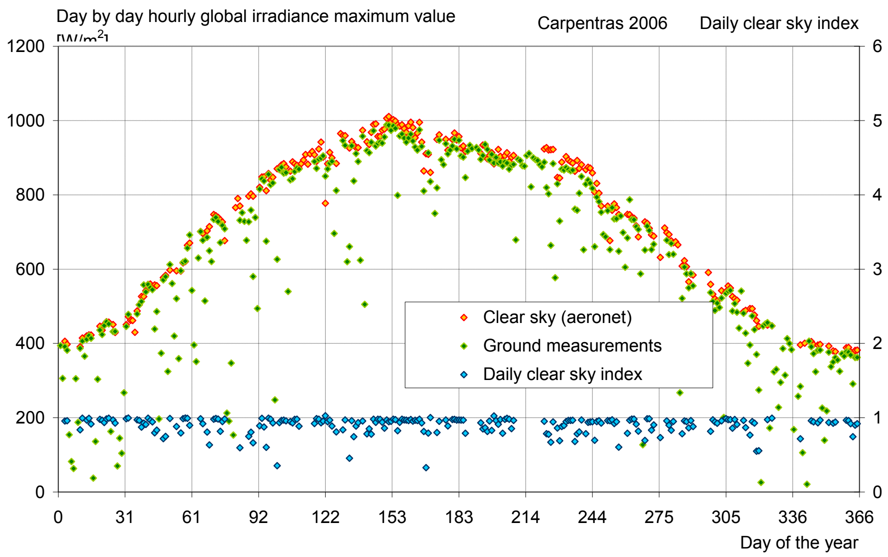
where δ_{cda} is the broadband clean-and-dry atmospheric optical depth, and δ_w is the water vapor optical depth. These two broadband optical depths can be evaluated with simplified expressions, following Molineaux et al. (1998):

$$\delta_{cda} = -0.101 + 0.235 \cdot AM^{-0.16} \tag{7}$$

$$\delta_w = 0.112 \cdot AM^{-0.55} \cdot w^{0.34} \tag{8}$$

where w is precipitable water (cm).

The denominator of K_{bc} in Eq. (6) is representative of the beam irradiance transmitted by a clean atmosphere. K_{bc} is also represented in Fig. 18.



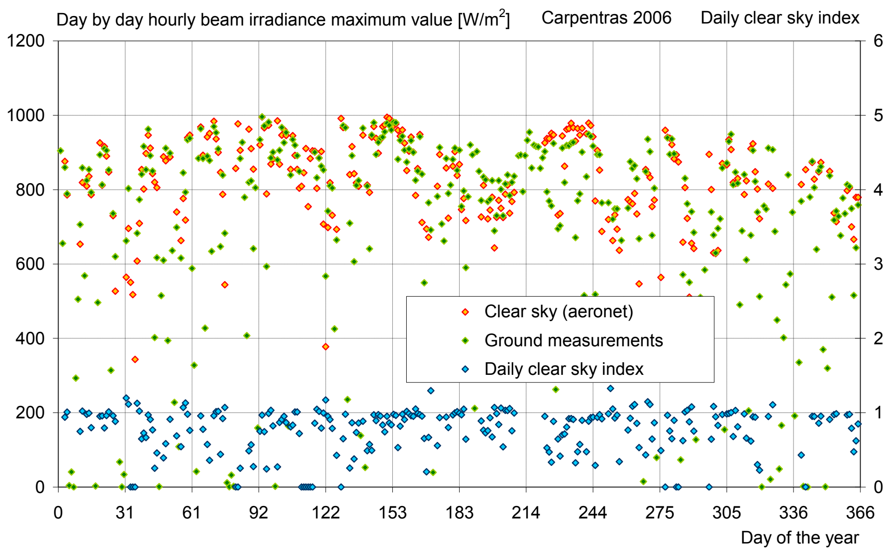


Fig. 18: Highest daily values of GHI (top plot) and DNI (bottom plot) versus the day of the year at Carpentras, for the irradiance (left Y-axis) and the corresponding modified clearness index (right Y-axis).

Another technique to delineate cloudless periods is to use the algorithm developed by Long and Ackerman (2000). This is more precise than just relying on K'_t, but requires separate measurement of the three components at relatively high frequency (time step of 6 minutes or less).

Properly testing the calibration correctness of radiometers is a difficult proposition because various factors or compensation of errors can lead to overcorrections. For instance, lack of regular cleaning of a thermopile radiometer in a dusty environment can lead to a rapid degradation of the instrument's response, which can be confused with an incorrect calibration. Whereas an error due to mis-calibration will remain constant for many months (if not years), sensitivity degradations due to soiling are short lived and fluctuate on a daily basis.

The calibration uncertainty of the tested instruments must also be considered. It is typically 1–2% for pyrheliometers and 3–5% for pyranometers. These uncertainties must be borne in mind, so that no correction should be attempted if the observed differences with reference data are within these uncertainty limits.

The calibration correctness can be assessed by direct comparison with data from a nearby site, if 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 an appropriate clear-sky radiative model, if the local atmospheric aerosol optical depth (AOD) and precipitable water (PW) are known. The upper limits of the compared plots should then coincide. Alternatively, AOD and PW data may be retrieved from a nearby sunphotometer. For instance, the stations of the Aeronet network (http://aeronet.gsfc.nasa.gov/) automatically retrieve AOD and PW at 15-minute intervals. Aeronet's Level-2 data (cloud screened and quality controlled) should be used whenever possible.

In case PW is not measured in the vicinity, it can be evaluated from ground ambient temperature (Ta) and relative humidity (RH) through the use of an appropriately selected empirical model. Many such simple models have been proposed in the literature (e.g., Garrison and Adler, 1990; Gueymard, 1994; Smith, 1966). Because of their empirical nature, their validity must first be verified over the area of study. At most sites, such models are known to introduce large random errors in PW at the hourly time scale. A better alternative is to obtain PW from satellite data (such as MODIS) or reanalysis data (such as NCEP). These alternate methods are still imperfect, but spatially extrapolating actual measurements also introduces errors, so that there is no perfect solution. These AOD and PW values are then used as inputs to a high-performance clear-sky radiative model (e.g., Gueymard, 2008; Ineichen, 2008) to evaluate the clear-sky hourly G and G_b values.

To quantify the correctness of the calibration factor, the selected clear-sky hourly values between the modeled and measured data points are linearly regressed. This is illustrated in Fig. 18 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. This method assumes that there is no error (and particularly no bias) in the modeled clear-sky irradiances, which is virtually impossible. It is thus assumed that calculated slopes between 0.97 and 1.03 are indicative of a good calibration for the period under scrutiny. Beyond these limits, a correction factor (equal to the calculated slope) should be applied to the data.

4.4 Direct/Global consistency

The consistency test between the global horizontal and direct normal components can be verified with the help of the global and beam clearness indices.

The hourly beam clearness index is plotted versus the corresponding global index as illustrated for the site of Carpentras in Fig. 19. On the same graph, the clear-sky predictions from the Solis radiative model are represented for four different a priori values of AOD. The corresponding Linke turbidity coefficient T_{Lam2} is then calculated from the DNI thus obtained:

$$G_b = I_o \, e^{-AM \cdot (\delta_{cda} + \delta_w)} \tag{9}$$

T_{Lam2} is evaluated for AM = 2 and its correspondence with δ_a is also indicated on the graph. Any important deviation between the predicted and measured clear-sky values indicates calibration uncertainties, pyrheliometer misalignment, soiled or shaded sensors, or misdetection of clear-sky conditions.

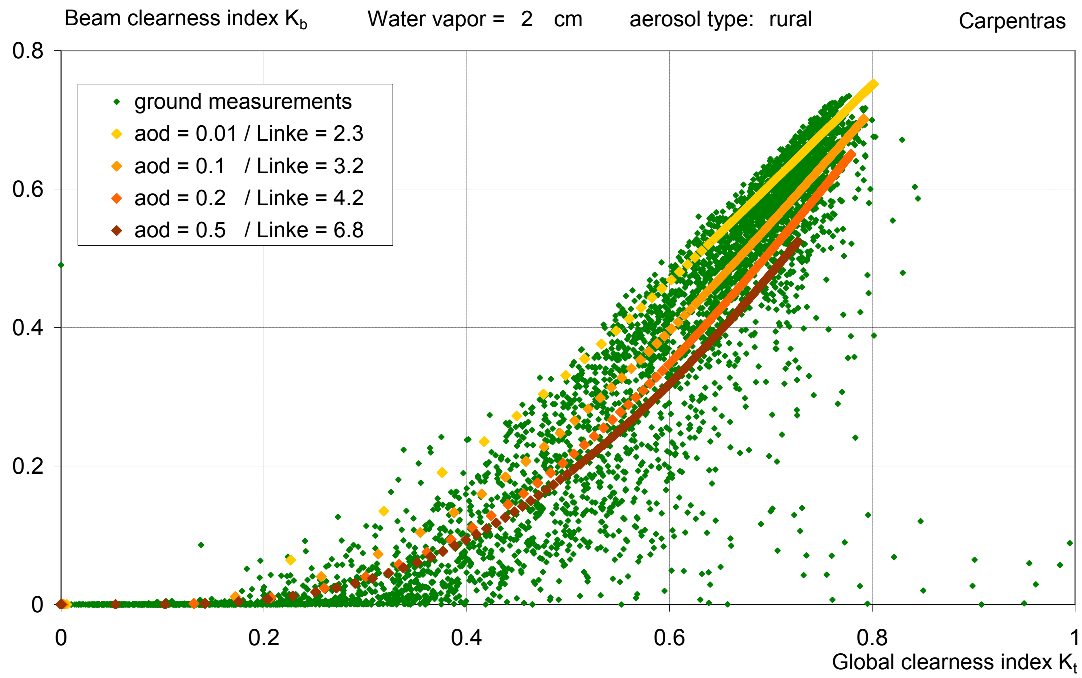
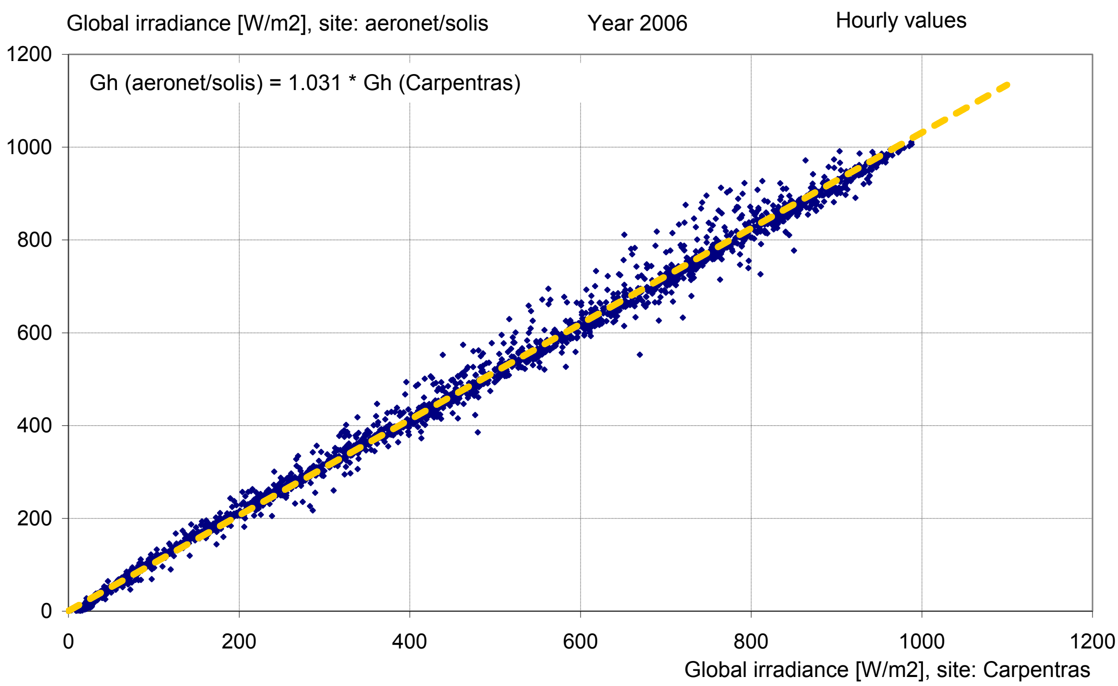


Fig. 19: Beam clearness index K_b as a function of the global clearness index K_t. The modelled clear-sky values are also represented for 4 different aerosol loads.



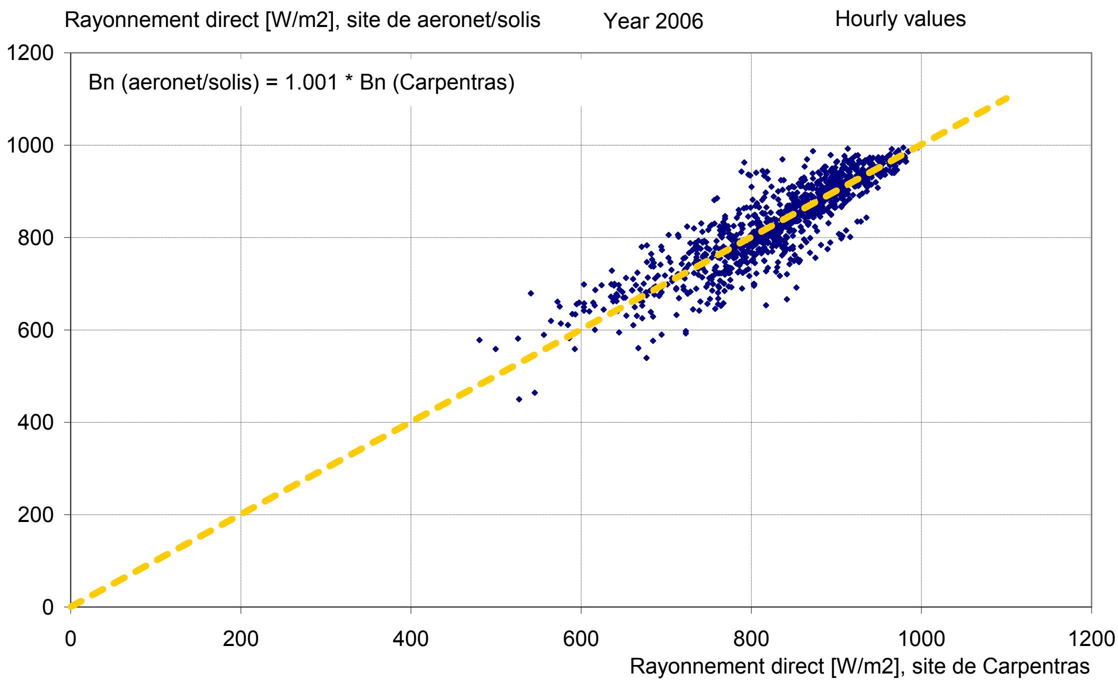


Fig. 20: Scatter plots of the modelled vs. measured GHI (top) and DNI (bottom) at Carpentras. The calculated slope is an estimate of the calibration coefficient correction factor.

5 Proposal for a benchmarking of DNI products

This section describes the next steps needed before the execution of a worldwide benchmarking of model-derived data can proceed. It outlines what would be beneficial to do and how it should be done in principle. Some preparatory activities are also proposed in order to facilitate and speed up the benchmarking process.

5.1 Objectives of a worldwide DNI benchmarking

The benchmarking should be open to all providers of model-derived DNI data products. These can be public entities entitled to providing such data, or commercial data providers who want to qualify their data by such an independent evaluation.

The objectives of this project are to:

- cover all world regions that are relevant for CSP,
- derive results pertaining to all regions, with hopefully sufficient geographical coverage of stations to derive overall regional statistics,
- involve as many data providers as possible,
- differentiate results of all assessments using
    - all available stations,
    - only public stations (which the model developers might have used for training)
    - only non-public stations,
- allow multiple data sets from the same provider, e.g.
    - products at various time resolutions, or
    - related to various spatial resolutions, or
    - successive versions or updates of products,
- provide a set of performance statistics for each delivered data product:
    - mean bias MB,
    - standard deviation SD,
    - root mean square deviation RMSD,
    - KSI or similar for frequency distribution testing,
    - CSP-performance-weighted DNI performance statistics,
- calculate the averages of all these performance statistics over many sites within a given region, and derive their standard deviation as an indication of uncertainty of each statistic,
- repeat these comparisons over long time periods to reach stable and significant results and derive precise uncertainties.

All the regional results should be ultimately combined to derive world-average statistics. Since it is expected that the behavior of some models is climate-dependent, and even possibly season-dependent, another reasonable segregation of results would be according to climatic regions and/or season. Due to the wider pixel sizes and grazing viewing angles towards the edges of a spaceborne sensor's field of view, separate summaries for different categories of satellite viewing angle could also be help identify the strengths and weaknesses of some data sets.

Additional insight on the relative performance of models vis-à-vis aerosol and cloud extinction might be obtained by separating clear and cloudy situations. Similarly, testing the performance of modeled data over highly reflective surfaces can be informative.

In summary, such a detailed quality assessment would help data providers improve their products further. The envisioned results would allow them to direct their development effort optimally, so as to improve the weak points of their model or input data. In parallel, irradiance data users would obviously be empowered by the detailed results, since they would have the necessary information to select the best source of data for any specific region. Technical and financial analysts would also benefit, by using performance results as inputs to their estimate overall uncertainties, for instance when analyzing the commercial viability of CSP projects.

The International Renewable Energy Agency (IRENA), in partnership with the Multilateral Working Group on Solar and Wind Energy (MWGSW) of the Clean Energy Ministerial (CEM), is just developing a Global Atlas for Solar and Wind Energy. The goal is to provide a web-based platform, which provides homogenous solar and wind datasets with global coverage, supporting policy formulation, as well as planning and stimulating investments in pre-feasibility studies for wind or solar projects. IRENA's goal of *homogenous* data sets is hard to reach, however. The planned benchmarking activity would certainly be of benefit to their endeavor. Ideally, the maps of actual variables such as DNI's long-term mean would be supplemented by maps indicating the corresponding uncertainty (Nicolas Fichaux, pers. comm. with Richard Meyer, Feb. 2012). The effort needed to obtain such uncertainty maps is large, but the goal is attainable if enough independent and qualified measured datasets are available.

A professional DNI benchmarking analysis of worldwide scope can only be successful if many conditions are met, such as:

- a sufficient number of available ground measurements, which are proven to qualify as reference,
- the willingness of data providers to supply data free of charge for a sizeable number of sites,
- informing each data provider independently about the preliminary results, showing them their own results and the results of others in an anonymized form,
- leave the providers the choice to opt in or out for publication of the results or disclosure,
- involve at least two independent experts or teams to perform the actual analysis independently, so as to avoid errors in the programming, interpretation of data quality, statistical assessments, etc.,
- have a reputable international organization, such as SolarPACES, for supervision, editing and dissemination of the reports,
- sufficient funding for the work to be done by the independent experts.

The first benchmarking exercise should concentrate on the evaluation of historical data, which is most important for the design, thermal simulation, and ultimately the overall qualification of CSP projects from a financial standpoint.. An ulterior phase of the project could be devoted to the validation of DNI forecasting products.

In the more long-term future, such the benchmarking exercise could be repeated to allow the evaluation of updated products form existing providers, or new data providers altogether. Once the processes are established and automated, the effort will become lower for new editions of the analysis. Given the current pace at which new satellite-derived products are developed, it would be worthwhile to repeat the benchmarking process every other year.

5.2 Proposed execution procedure

A worldwide DNI benchmarking meeting the above objectives is a major effort, which needs to be well structured. It is expected that measured data from about 100–200 sites could be made available worldwide. Of these we assume ≈50% could be of good enough quality for benchmarking. This means the benchmarking analysis could use an estimated 50–100 sites worldwide.

For each of these sites we expect that 5 to 8 data suppliers could deliver satellite-derived DNI data series. Some of these suppliers could supply different data sets related to as many time-steps, like 60-min, 30-min, 15-min, or even 10-min intervals. Some suppliers may also want to send their data relatively to various spatial resolutions, like averaged over several satellite pixels to reduce noise, or disaggregated to 250 m or even 90 m resolution, as sharp as the underlying digital elevation model allows to derive proper local shading conditions. Thus, it is expected that for each site 10 to 20 different data products could be supplied. In total, this leads to an analysis using 500 to 2 000 data sets. Such a vast amount of data sets can be assessed with reasonable effort only if the process is automated as thoroughly as possible.

The DNI benchmarking exercise could be organized like this:

1. Define standards for DNI-time-series products
2. Detailed project organization and secure funding
3. Prepare a detailed project organization and secure funding
4. Prepare a workshop to inform stakeholders
5. Gather measured datasets and check their quality
6. Ask data providers for satellite-derived DNI data products
7. Check compliance of DNI data products with the agreed-upon file format
8. Perform the statistical analysis
9. Prepare a final report and some technical or scientific publications.

These steps are described further in the following sections.

5.2.1 Standards for DNI time-series data products

For an easy processing of approximately 500 to 2000 DNI data products, all these time series will need to be archived in a database. To reduce the effort of handling and reading into the database, the data providers will be asked to provide their data products in a specific file format, which will be determined in the next phase of the study.

A more chaotic handling problem will occur for the data originating from the many measurement sites, because each institution or caretaker of these stations has adopted a specific format. Such datasets are often kept in a myriad of small files, e.g. 1 file per day or month. This makes the reading of such data files a hassle, especially if they do not have descriptive headers.

From this first phase of the DNI benchmarking process the following recommendations are issued:

The goal is to define the radiation data formats that are most suitable to support the computerized performance simulation of CSP plants. Specific requirements of CSP on DNI data shall be considered. For this purpose industry and the SolarPACES Task I activity for standardizing performance simulation models shall be contacted.

The data formats to be chosen should be capable of various time resolutions, as well as sub-hourly and variable start and end dates (i.e., not just fixed years). The formats should follow established open standards, as far as possible. Handling of the data should be easy to read for the users. For the data suppliers creation of the data sets also should be possible with low effort. The size of data files should be reasonable—especially when high time resolutions need to be maintained over long periods. For adding younger data it would be an advantage, if the data files are flexible enough to allow additions or concatenations.

1. A common file format for numerical data is Microsoft-Excel (here indicated as xls or xslx). It is of everyday use in science and engineering because it imports directly into essentially all spreadsheet software currently available for all computer platforms. Files formatted in xls/xlsx can be converted to ASCII, and vice versa. However this format can hide many flaws, like formulas rather than numbers, incorrect formatting, etc. Excel files can be inadvertently mis-arranged or mis-handled, so that an automated read might fail. With the original Excel format the file size is limited to around 64 000 rows. This does not allow storage of more than 7 years of hourly data in one table, or a single year for a time resolution better than 10 min. This, as well as the large file size typical of this format, would not be acceptable in an intensive benchmarking analysis. The newer Excel format has a limit of 1 048 576 lines, but this can still be exceeded if more than 2 years of 1-min data has to be stored in a single file. This newer format seems to store data in a compressed way, which makes file sizes now acceptable, and comparable to that of ASCII files. Experience with Excel files across various operating systems shows that the end of line / carriage return marks are not defined consistently when exporting to ASCII. Thus, for instance, some weather input files created in Excel under Mac OS X would not be usable in NREL's SAM tool. Due to these inconveniences, it is concluded that the Excel format is inadequate for the present purpose.
2. The ubiquitous ASCII (text) format is still, by far, the easiest file format to use. One shortcoming of ASCII files is that they do not adjust the digital separator sign to the user's preferences like in Excel. There are many ASCII data files where the comma "," is used instead of the digital point ".", which must be fixed before the actual processing can take place. Columns are often separated by inserting many blank spaces instead of the "tab" character. Moreover, floating point numbers with some digits occupy many bytes—basically one per digit plus the digital separation sign, plus the column separator. Thus, if more than 2 digits are used to indicate a floating-point-number, 4 bytes per floating point-values are used. Since usually more than just 2 digits are provided, ASCII files are notoriously inefficient in storage (leading to large file sizes), and often truncate details. The problem of large file sizes can be easily circumvented by using compression. Usual compression engines (zip, rar or gz) are extremely efficient, often saving 90% or more space. Text files can also be easily read and manipulated with simple codes written in Fortran, C, or other programming languages. ASCII formats can also be read directly by basically all plotting software on the market, which adds convenience.
3. Often it is observed that ASCII data files do not have a proper description of meta-data because the format itself is completely free and has no structure, which forces the user to implement meta-data. In the course of the MESoR-project a ASCII-based data format was defined, which allows a well-defined, but flexible header structure. If meta-data are filled in properly, that format would allow fully automated reads into a data base. Consequently, this MESoR format is a good candidate for an acceptable file format.
4. If uncompressed, ASCII files of high-resolution meteorological data can easily reach 100 MByte, which makes them difficult to handle. With many radiometric stations recording at 1-min resolution, and new DNI products (forecasting) appearing with fast refresh cycles, this question of file size will become critical in the near future. Therefore, the adoption of a storage-efficient file format is desirable.
5. Established binary formats, like HDF or netCDF, are well designed to carry meta-data and multi-dimensional arrays of large size, and are routinely used by meteorological or space organizations (such as NASA or NOAA) to distribute satellite or numerical weather prediction data, for instance. They appear to be well designed as end products, because of their ability to contain dense spatial data that need to be visualized. In contrast, their suitability to store irradiance data at single sites remains to be established. Moreover, their handling (read/write) requires special software tools and additional steps,

which makes their manipulation difficult, and ultimately makes them unattractive to most users in the solar community. If netCDF or other binary formats were to be established in this sector, the potential users would certainly need to be educated and given some basic tools, for instance to easily convert such files into ASCII, make them readable by common performance simulation tools, analysis programs such as Excel, or plotting programs. Who would offer to develop such tools rapidly and in an open-source environment remains to be seen.

5.2.2 Project organization and funding

As pointed out above, the execution of a professional DNI benchmarking assessment is a major effort. It requires the set-up of databases and implementation of customized routines for quality control, and analysis of the data according to the proposed statistical method. Assessing and interpreting the statistical results obtained from the estimated 500 to 2000 data sets will also require a substantial amount of work. As a preliminary estimate, and depending on the level of detail and number of sites, the benchmarking exercise would take at least 12 months of full-time work from at least 4 solar radiation experts.

Currently there is no funding available to these authors that would allows them to undertake this effort. One of the objectives of this report is to help clarify what remaining preliminary steps need to be taken. Therefore, it is anticipate that serious funding opportunities will materialize next year, which would be enough to launch a basic analysis at least.

5.2.3 Organization of a workshop

A SolarPACES Conference could be the perfect venue for an experts workshop that would publicize this effort. The workshop would cover the following topics:

- Importance of independent benchmarking to improve the credibility of data suppliers
- Proposal on how to conduct a fair benchmarking
- Discussion about the benchmarking criteria—scientific vs. commercial considerations
- Ways of dealing with possibly confidential data, or the publication of results
- Statistical tests—what the CSP industry wants to know
- Previous experiences with measured data from various sites—what interpretation errors to avoid
- Definition of a common data format for the modeled data sets to be benchmarked.

The data providers would benefit from this workshop by gleaning valuable information from other experts in the field, and detect ways to improve their products or make them more appealing to the CSP industry.

5.2.4 Collecting reference measured data

Many power plant projects have unfortunately failed during project development. In several cases the developers had already installed a weather/radiometric station and obtained measurements during a year or more before the project eventually failed and could not be realized. In such cases, a budget could be set aside to pay at least 2000 € per year for qualified and nearly complete measured datasets from interesting private sites situated at low-latitude. Since the up-front cost for a year of sound DNI measurements is about 15 to 40 k€ (depending on equipment), this financial offer could be a sufficient incentive to push project developers into providing their data, thus ultimately helping the whole industry.

5.2.5 Invite potential DNI time-series suppliers

As soon as the necessary budget to conduct the actual benchmarking process is secured, the potential satellite data providers would be contacted to inform them of the project, discuss the proposed procedures, file formats and performance indicators, etc. Those would be the result of the next phase of this project. Agreements with providers of satellite data and ground-based measurements would need to be signed, after a general consensus is reached about confidentiality issues. These agreements would be discussed, and possibly modified, during the workshop. Signatures would be expected soon after that from all interested parties.

The list of test sites would then be sent along with instructions on allowed data formats, etc. A processing period of up to 8 weeks would be offered to all data suppliers.

5.2.6 Data compliance

Only modeled data files delivered in the accepted format(s) would be considered for evaluation. Preliminary tests would have to determine whether the supplied files respect the format options. Additionally, a basic test could be done to spot check the data validity. If the submitted files do not pass these initial tests, they would be rejected. Their supplier would be given, e.g., one week for resubmission.

5.2.7 Data processing and analysis

The actual analysis would then start. The goal would be to obtain the various statistics indicated in Section 5.1. This requires the development of many automated algorithms. The achievable level of detail of the benchmarking assessments would have to be adjusted as a function of the available budgets, timelines, and unexpected difficulties encountered.

5.2.8 Reporting

The benchmarking process and results should be well documented. After the independent experts complete their individual analysis, each of them would write a separate internal report. The content—mainly tables and figures—should be agreed upon beforehand, so as to perform the analysis as fast, transparently, and error-free as possible. If results differ, the sources of error or disagreement need to be identified, so that only correct results are handed out to data providers and other stakeholders.

Separate reports for each DNI data product supplier would then be prepared. In these series of reports, only the results pertaining to the report's recipient would be identified, whereas all results from other suppliers would be anonymized. This step allows the data suppliers to evaluate their own products, identify the necessary modeling improvements that would be needed to keep up with the competition, and authorize which of their data products may be published with their names or shall stay anonymous.

Based on this feedback a summary report would be written and ultimately officially published. This report should be peer-reviewed to ensure high quality of the process, guarantee professional documentation, and provide expert advice on improvements.

The level of detail of these reports will strongly depend on the available budgets and timelines. Ideally, an additional series of reports could be prepared for those who offered measured data, or institutions (like the World Bank) having the mandate to help fund the development of CSP projects in various countries. These reports would summarize all benchmarked products for a site or a region. As with the public summary report, these specific reports would only disclose the names of those providers who have agreed to be identified in such publications.

6 References

Anderberg, M.H.L. (1999) A review of SIRS data quality at the ARM Southern Great Plain site. Proc. 9th ARM Science Team Meeting, San Antonio, TX.

Anderberg, M.H.L. and Stoffel, T.L. (1998) Solar radiometric data quality analysis for ARM Southern Great Plain site. Proc. 8th ARM Science Team Meeting, Tucson, AR.

Anderberg, M.H.L., Rael, F.P. and Stoffel, T.L. (2002) Procedure for generating data quality reports for SIRS radiometric measurements. Proc. 12th ARM Science Team Meeting, St. Petersburg, FL.

Beyer, H.G., Polo Martinez, J. , Šúri, M., Luis Torres, J., Lorenz, E., Müller, S.C., Hoyer-Klick, C., Ineichen, P. (2009): Report on Benchmarking of Radiation Products. Report of the MESoR-project. Deliverable D 1.1.3.

Cebecauer, T., Šúri, M., and Gueymard, C.A. (2011) Uncertainty sources in satellite-derived direct normal irradiance: How can prediction accuracy be improved globally? Proc. SolarPACES Symposium, Granada, Spain

Chhatbar K. and Meyer, R. (2011) The influence of meteorological parameters on the energy yield of solar thermal plants. Proc. SolarPACES Symp. Granada, Spain, Sep. 20-24 2011, 8 p.

Eck M., Barroso H., Blanco M., Burgaleta J.-I., Dersch J., Feldhoff J.F., Garcia Barberena J., Gonzales L., Hirsch T., Ho C.K., Kolb G., Neises T., Serrano J.A., Tenz D., Wagner M., Zhu G. (2011): *guiSmo: Guidelines for CSP Performance Modeling – Present Status of the SolarPACES Task I Project*,. Proc. SolarPACES Symposium. Granada, Spain.

Espinar, B. et al., (2006) Analysis of different comparison parameters applied to solar-radiation data from satellite and German radiometric stations. *Solar Energ,* **83**, 118-125, 2009.

Faiman, D. et al. (2004) The Negev radiation survey. *J. Solar Energy Eng. Trans. ASME* **126**, 906-914.

Geuder, N. and Quaschning, V. (2006) Soiling of irradiation sensors and methods for soiling correction. *Solar Energy,* **80**, 1402-1409.

Geuder, N., et al. (2006) Examination of different irradiation sensors: Operation experiences and comparative study. Proc. 4th International Conference on Experiences with Automatic Weather Stations, Lisbon, Portugal.

Geuder, N., et al. (2008) Corrections for Rotating Shadowband Pyranometers for solar resource assessment. Proc. Solar Energy + Applications Conf., SPIE #7046, San Diego, USA.

Geuder N., et al. (2009) Precise measurements of solar beam irradiance through improved sensor calibration. Proc. SolarPACES Symposium, Berlin, Germany.

Geuder, N., et al. (2010) Validation of direct beam irradiance measurements from rotating shadowband pyranometers in a different climate. Proc. SolarPACES Symposium, Perpignan, France.

Garrison, J.D. and Adler, G.P. (1990) Estimation of precipitable water over the United States for application of the division of solar radiation into its direct and diffuse components. *Solar Energy,* **44**, 225-241.

Gueymard, C.A. (1994) Analysis of monthly average atmospheric precipitable water and turbidity in Canada and northern United States. *Solar Energy,* **53**, 57-71.

Gueymard, C.A. (2008) REST2: High performance solar radiation model for cloudless-sky irradiance, illuminance and photosynthetically active radiation—Validation with a benchmark dataset. *Solar Energy,* **82**, 272-285.

Gueymard, C.A. (2010) Variability on direct irradiance around the Sahara: Are the modeled datasets of bankable quality? Proc. SolarPACES Symposium, Perpignan, France.

Gueymard, C.A. (2011) Uncertainties in modeled direct irradiance around the Sahara as affected by aerosols: Are current datasets of bankable quality? *J. Solar Energy Eng. Trans. ASME,* **133**, 031024.

Gueymard, C.A. (2012) Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Solar Energy,* **86**, 2145-2169.

Gueymard C.A. and Myers D.R. (2008) *Solar radiation measurement: Progress in radiometry for improved modeling*, in *Modeling Solar Radiation at the Earth Surface*, V. Badescu, Editor, Springer.

Gueymard C.A. and Myers D.R. (2009) Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling*. Solar Energy,* **83**, 171-185.

Hoyer-Klick, C., Hustig, F., Schwandt, M., Meyer, R. (2009) Characteristic meteorological years from ground and satellite data. Proc. SolarPACES Symposium, Perpignan, France.

Hoyer-Klick, C., *et al.* (2009) MESoR—Management and exploitation of solar resource knowledge. Proc. SolarPACES Symposium, Berlin, Germany, Sep. 2009.

Ineichen, P. (2008) A broadband simplified version of the Solis clear sky model. *Solar Energy,* **82**, 758–762.

Ineichen, P. (2011) Five satellite products deriving beam and global irradiance validation on data from 23 ground stations. Univ. of Geneva. Available from <http://www.cuepe.ch/html/biblio/pdf/ineichen_2011_Five_satellite_products_deriving(iea).pdf>.

Ineichen, P., Barroso, C., Geiger, B., Hollmann, R., Marsouin, A., Mueller, R. (2009) Satellite Application Facilities irradiance products: hourly time step comparison and validation over Europe. *Int. J. Remote Sens.,* **30**, 5549–5571.

Kasten F. (1980) A simple parameterization of two pyrheliometric formulae for determining the Linke turbidity factor. *Meteorol. Rundsch.,* **33**, 124–127.

Kumar, A., Giridhar, G., Mitra, I., Vashistha, R., Meyer, R., Schwandt, M., Chhatbar, K. (2012) Solar Radiation Resource Assessment and Mapping of India. Proc. SolarPACES Symposium, Marrakech, Morocco.

Long, C.N. and Ackerman, T.P. (2000) Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.,* **105D**, 15609–15626.

Long, C.N. and Dutton, E.G. (2002) BSRN Global Network recommended QC tests, V2.0. Technical Note to WMO/BSRN.

Maxwell, E.L., Wilcox, S., Cornwall, C., Marion, B., Alawaji, S.H., Mahfoodh, M., Al-Amoudi, A. (1999) Progress Report for Annex II – Assessment of Solar Radiation Resources in Saudi Arabia 1993–1997. Rep. NREL/TP-560-25374, National Renewable Energy Laboratory, Golden, CO.

McArthur, L.J.B. (2005) World climate research programme. Baseline Surface Radiation Network (BSRN). Operations Manual. Version 2.1. April 2005, Technical Report WCRP-121, WMO/TD-No. 1274.

Meyer, R. (2010) Recommendations for bankable meteorological site assessments for solar thermal power plants. Proc. SolarPACES Symposium, Perpignan, France.

Meyer, R., Beyer, H. G., Fanslau, J., Geuder, N., Hammer, A., Hirsch, T., Hoyer-Klick, C., Schmidt, N., Schwandt, M. (2009) Towards standardization of CSP yield assessments. Proc. SolarPACES Symposium, Berlin, Germany.

Michalsky, J.J., et al. (2005). Toward the development of a diffuse horizontal shortwave irradiance working standard. *J. Geophys. Res.*, **110D**, doi:10.1029/2004JD005265.

Michalsky, J., *et al.* (2011) An extensive comparison of commercial pyrheliometers under a wide range of routine observing conditions. *J. Atmos. Ocean. Tech.,* **28**, 752-766.

Molineaux, B., Ineichen, P., O'Neill, N.T (1998) Equivalence of pyrheliometric and aerosol optical depth at a single wavelength. *Appl. Opt.*, **37**, 7008-7018.

Mueller, R.W. *et al.*, (2004) Rethinking satellite based solar irradiance modeling – The SOLIS clear sky module. *Remote Sens. Environ.*, **91**, 160-174.

Myers, D.R. (2010) Comparison of direct normal irradiance derived from silicon and thermopile global hemispherical radiation detectors. Proc. Solar Energy + Applications Conf., SPIE #7773. San Diego, CA.

Myers, D. and Wilcox, S. (2009) Relative accuracy of 1-minute and daily total solar radiation data for 12 global and 4 direct beam solar radiometers. Proc. Solar 2009 Conf., Buffalo, NY, American Solar Energy Soc. <http://www.nrel.gov/docs/fy09osti/45374.pdf>.

NREL (1993) User's Manual for SERI_QC Software – Assessing the Quality of Solar Radiation Data, National Renewable Energy Laboratory, Tech. Rep. NREL/TP-463-5608.

Ohmura, A., *et al.*, (1998) Baseline Surface Radiation Network (BSRN/WCRP): New precision radiometry for climate research. *Bull. Amer. Met. Soc.,* **79**, 2115-2136, 1998.

Perez, R., Ineichen, P., Seals, R., Zelenka, A. (1990) Making full use of the clearness index for parameterizing hourly insolation conditions. *Solar Energy,* **45**, 111-114.

Roesch, A., Wild, M., Ohmura, A., Dutton, E.G., Long, C.N. and Zhang, T. (2011). Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech.*, **4**, 339–354.

Ruiz-Arias, J.A., Cebecauer, T., Tovar-Pescador, J. and Suri, M. (2010) Spatial disaggregation of satellite-derived irradiance using a high-resolution digital elevation model. *Solar Energy*, **84**, 1644-1657.

Schillings, C., Meyer, R., Mannstein, H. (2004) Validation of a method for deriving high resolution direct normal irradiance from satellite data and application for the Arabian peninsula. *Solar Energy,* **76**, 485–497.

Smith, W. L. (1966) Note on the relationship between total precipitable water and surface dewpoint. *J. Appl. Meteor.*, **5**, 726-727

Stoffel, T., *et al.* (2010) Concentrating solar power—Best practices handbook for the collection and use of solar resource data. Tech. rep. NREL/TP-550-47465, National Renewable Energy Lab., Golden, CO <http://www.nrel.gov/docs/fy10osti/47465.pdf>.

Šúri, M., Remund, J., Cebecauer, T. , Hoyer-Klick, C., Dumortier, Huld, T., Stackhouse, Jr., P.W., Ineichen, P. (2009): Comparison of direct normal irradiation maps of Europe. Proc. SolarPACES Symp., Berlin (Germany), Sep. 15-18, 2009.

Wilbert, S.*, et al.* (2010) Reduced uncertainties of field pyrheliometers through improved sensor calibration. Proc. SolarPACES Conf.*,* Perpignan, France.

Wilcox, S. (1996) A Visual Quality Assessment Tool for Solar Radiation Data. Proc. Annual Conf. of the American Solar Energy Society, Boulder, Colorado, April 1996.

Wilcox, S. M. and McCormack, P. (2011) Implementing Best Practices for Data Quality Assessment of the National Renewable Energy Laboratory's Solar Resource and Meteorological Assessment Project. Proc. SOLAR 2011 Conf., Raleigh, North Carolina, NREL/CP-5500-50897 <http://www.nrel.gov/docs/fy11osti/50897.pdf>.

Wilcox S. et al., (2007) National Solar Radiation Database 1991–2005 Update: User's Manual. Tech. Rep. NREL/TP-581-41364 <http://www.nrel.gov/docs/fy07osti/41364.pdf>.

Wyser, K. et al. (2002) Remote sensing of surface solar irradiance with corrections for 3-D cloud effects. *Remote Sensing Environ.*, **80**, 272-284.