Principles and Practical Methods for Estimating Uncertainty and Evaluating Solar Irradiance Data

Aron Habte, Christian Gueymard, Stefan Wilbert, Elke Lorenz, Balenzategui Manzanares, Manajit Sengupta, Daryl Myers, Thomas Stoffel, Frank Vignola, Jaemo Yang, et al.

10 Principles and Practical Methods for Estimating Uncertainty and Evaluating Solar Irradiance Data

Aron Habte,[1] Christian Gueymard,[2] Stefan Wilbert,[3] Elke Lorenz,[4] José Lorenzo Balenzategui Manzanares,[5] Manajit Sengupta,[1] Daryl Myers,[6] Thomas Stoffel,[7] Frank Vignola,[8†] Jaemo Yang,[1] Philippe Lauret,[9] and Mathieu David[9]

[1] National Renewable Energy Laboratory (NREL), USA
[2] Solar Consulting Services, USA
[4] German Aerospace Center (DLR), Germany
[4] Fraunhofer Institute for Solar Energy Systems, Germany
[5] PVLab-CIEMAT, Spain
[6] NREL (retired), USA
[7] Solar Resource Solutions, LLC, USA
[8] University of Oregon, USA
† *Deceased*
[9] University of La Réunion, France

Executive Summary

The rising investment in solar energy projects has necessitated the development of improved methods to quantify and assess the uncertainty of solar resource data. The practical challenges in this area stem from the instrument used to monitor the measurand, the model and input data used to predict and forecast the measurand, and their interactions. By maintaining the proper traceability, these sources of uncertainty amplify or compensate each other as they propagate from the reference data to prediction and forecast, for example. This propagation of uncertainty has significant impacts on the prediction and forecast data, which subsequently affects the project's financing, as well as the levelized cost of energy (LCOE) and decision making at various steps.

This chapter discusses the uncertainties associated with various forms of solar resource data and how these data impact the predictions of physical or empirical models that use such data. For the purposes of this chapter, solar resource data can be classified into three different categories: experimental data, as those measured at ground stations; modeled data, estimated for past periods using physical, semiempirical, or other radiative models; and forecast data, which use current experimental data and models to estimate the future irradiance for a particular area, season, and time. The latter can be distinguished between short-term forecasting (intrahour, intraday, and days-ahead) and long-term predictions for the next decades.

Accurate measurement, prediction, or forecasting of the solar resource is complicated by the rapidity with which the solar irradiance can change, both in magnitude and spectral distribution, and the varied environmental conditions experienced during measurements.

In the case of predicted and forecasted datasets, it is essential to understand the factors that impact their accuracy relative to ground measurements because of, for example, error propagation. In parallel, the quality of these measured data is key for confidence in the determination of uncertainty in predicted and forecasted datasets. Additional factors can be considered, such as the interannual variability if using only a short dataset of, for example, 12 months or less. Thus, the overall uncertainty of a modeled dataset should include an estimate of the uncertainty of the modeled solar resource, of the ground measurements, and that resulting

from the probable interannual variability. As a general rule, even with improved instrumentation and radiation models, both the measurement and modeling of the incident irradiance can have uncertainties, depending on various factors that cannot be neglected and should be properly taken into consideration.

Sections 10.1 and 10.2 summarize the basic concepts and methods of uncertainty in datasets. Section 10.3 discusses the measurement uncertainty, mainly using the Guide to the Expression of Uncertainty in Measurements (GUM) methodology for quantifying the uncertainty for measured irradiance. Afterward, the uncertainty of modeled and forecasted data is discussed in Sections 10.4 to 10.6. Section 10.7 illustrates some available diagnostic algorithms and tools. Section 10.3 on measurements comes first, because the uncertainty in the modeled data is typically obtained by comparison with reference measurements.

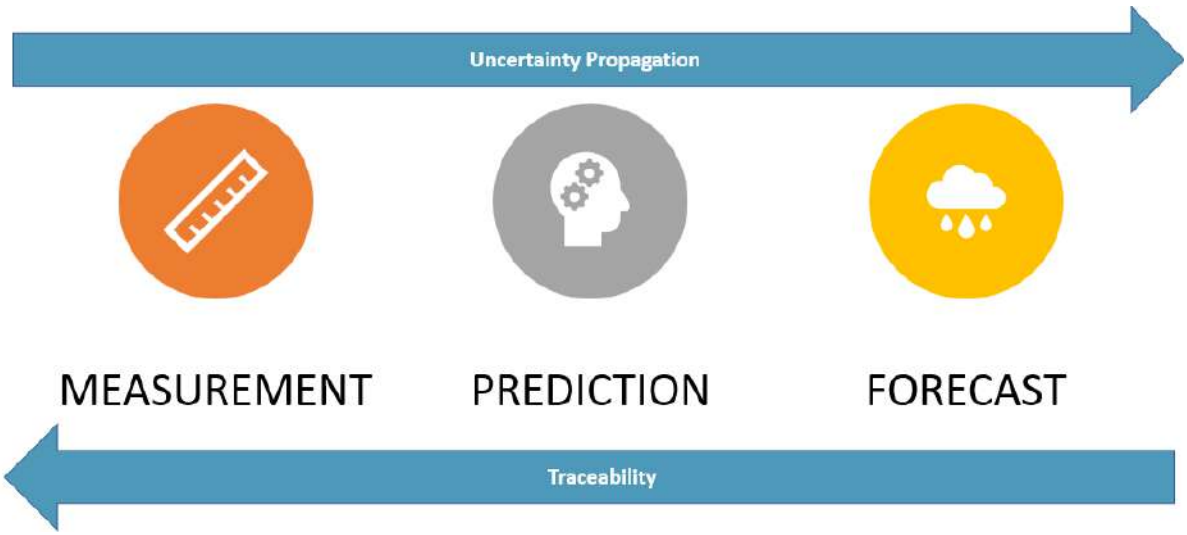

Figure ES 10-1. Traceability and uncertainty propagation for various sources of solar resource data

Image by NREL

10.1 Introductory Outline

Solar irradiance can be measured, modeled, or forecasted with various methods, as described in other chapters. These solar irradiance data are imperfect and thus uncertain.

The actual uncertainty of a dataset strongly depends on the measurement and/or modeling approach, as well as on the considered spatial and temporal scales. Measurements normally serve as a reference baseline for validating modeled data because the latter are expected to have a relatively larger uncertainty. In what follows, the main types of models under scrutiny are those that provide either satellite-derived irradiance estimates or forecasts using numerical weather prediction (NWP).

The following list summarizes a number of common notions and perceptions that are encountered in the solar industry's current practice and provides some initial recommendations based on experience.

- Measurements:
    - Measurements are the data source many users consider “ground truth”; however, measurement uncertainty is frequently not, or not sufficiently, taken into account.
    - Measurement uncertainty strongly depends on instrument specification (e.g., Instrument Class, according to ISO 9060:2018 (ISO 9060 2018)) and quality control (QC); it is essential to evaluate and further consider this uncertainty whenever using measured data.
    - Measurements obtained with Class-A instruments and subject to thorough maintenance program and QC procedures have the lowest possible uncertainty for the site and period under scrutiny. Therefore, they are a suitable reference among other tasks, such as evaluating modeled or forecasted data.
    - Because high-quality irradiance measurements can be sparse, especially for long-term periods, modeled data (of greater uncertainty) are needed in most cases.
- Modeled predictions (satellite- or NWP-based):
    - Definition: Irradiance values for past periods that are inferred using various inputs related to the same site and time through a radiative transfer model or a suite of functions, such as GHI(t)= f(M1(t), M2(t)), where the inputs are retrieved from satellite imagery or meteorological measurements.
    - Uncertainty strongly depends on the time scale of evaluation and decreases with increasing time scales (e.g., hourly values have a higher uncertainty than monthly values or, even more so, long-term mean annual values).
    - The actual uncertainty of predictions strongly depends on the performance of the model itself (or type of model) and time scale; generally, satellite-based modeled data have lower uncertainty than NWP-based irradiance data.
    - The most common way to evaluate the quality of modeled datasets is to compare them to high-quality irradiance measurements for the same location and period; the differences are summarized by some usual statistical indicators of deviation, such as root mean square deviation (RMSD), mean bias deviation (MBD), or mean absolute deviation (MAD) (see Section 10.2.3). The final letter D stands here for “deviation” or “difference” and is used in lieu of “error” to reflect that both datasets have uncertainty, which might be of similar magnitude.
    - The uncertainty of modeled data can be estimated based on statistical metric evaluations. Here, measurement uncertainty is often not properly taken into account because the differences depicted by the metrics between the modeled and measured datasets could move up and down depending on the uncertainty of the measured dataset, and may or may not be significant within these uncertainty intervals; this critical aspect is discussed further below. Note that, in the case of modeled data, accuracy and uncertainty are first a consequence of the quality of the experimental data used for validation and the amount of data available.
    - High-quality satellite-based irradiance predictions are generally considered the next best thing whenever high-quality measurements are not available for a given location and period. In such cases, they might constitute a suitable reference for

the evaluation of other modeled or forecasted data, at least for global horizontal irradiance (GHI).

- Short-term forecasts (intrahour, intraday, and days-ahead):
    - Definition: Future irradiance values that are inferred using present or past inputs: $GHI(t+\Delta t) = f(M1(t), M2(t), N1(t-\Delta t), N2(t-\Delta t))$ Here, short-term forecasting refers to forecasting up to days ahead, in contrast to long term-predictions, as described below.
    - The forecast uncertainty increases with increasing forecast horizon, and also strongly depends on the spatial and temporal scale of evaluation. It decreases with increasing time scales (e.g., minute values have a higher uncertainty than hourly or daily values). It also decreases with increasing spatial resolution (e.g., site-specific or local forecasts have higher uncertainty than regionally averaged or aggregated forecasts).
    - Forecast uncertainty is typically greater than the uncertainty of high-quality measurements or modeled data (suitable for the given spatial and temporal scale).
    - The most common way to evaluate the quality of forecast datasets is by assessing their similarity to high-quality irradiance measurements (or, if not available, high-quality modeled data) and by calculating the usual statistical indicators, such as root mean square error (RMSE), mean bias error (MBE), or mean absolute error (MAE) (see Section 10.2.3), with E denoting "error," which is common practice in meteorological forecasting because the reference measurements are assumed to be of negligible uncertainty compared to that of forecasts.
    - Forecast uncertainty is described by probabilistic forecasts consisting of probability distributions or scenarios of possible future values, depending on the current meteorological situation. As before, the reliability and resolution of these probabilistic forecasts is evaluated against high-quality irradiance measurements (or high-quality modeled data).
- Long-term predictions:
    - Long-term predictions aim to describe irradiance conditions for the next decades, typically 20–30 years for yield assessments, that is, the expected conditions in 2030, 2040, or 2050 toward the preparation of future energy scenarios.
    - Currently, long-term predictions are based on long-term historical measured or modeled data time series, a combination of both, or alternatively on typical meteorological year (TMY) datasets.
    - The accuracy of these long-term predictions is typically estimated from a statistical evaluation of modeled and/or measured data for a specific period in the past. Sometimes, interannual variability and irradiance trends are also taken into account. There is currently no consensus methodology to characterize estimates of the future resource with a precise uncertainty value and a specific confidence interval, contrary to the situation for measurements (see Chapter 6).
    - Long-term trends caused by climate change or regional air-quality measures should additionally be accounted for to estimate the uncertainty of long-term

irradiance predictions. This is a very complex task that requires more research to adequately address (see Chapter 6).

10.2 Basic Uncertainty Concepts

A clear statement of uncertainty should accompany any comprehensive solar radiation dataset to provide the necessary context for understanding the reliability of the data for various solar energy applications. For example, estimation of uncertainty provides a basis to assess confidence in the predicted output of a planned solar conversion system and is thus a key factor when determining the bankability of the project. Uncertainty is a way to specify confidence in the data.

In this preliminary approach, it is worth mentioning two important concepts associated with uncertainty that are sometimes misunderstood or mixed up. The first one is the accuracy of an experimental sensor or of a mathematical model. The accuracy is related to the ability of the sensor/model to determine the correct value of a magnitude, its true value, and how much the readings or estimations are deviated or separated from that value. The uncertainty gives an idea of how much the values of the magnitude provided by the sensor/model can be spread or dispersed around its readings/estimations in repeated measurements or calculations. The second important concept is the calibration (for an experimental sensor) or the evaluation (for an algorithm), which is the procedure or the means by which these deviations are determined and stated. In this step, it is necessary to rely on a reference value given by a standard instrument (for a sensing device) or provided by a set of real data or a function (for a model). A contribution for the overall uncertainty of the value of the magnitude given by the sensor or model is generated during this calibration or evaluation step.

It is important to determine the uncertainty using a standard methodology to provide authoritative results that can be relied on for analysis and comparisons. The GUM (ISO/IEC 2008) is an example of how to determine the uncertainty in measurements. GUM has been formalized by several organizations, including the International Bureau of Weights and Measures (French acronym: BIPM) and published by the International Standards Organization (ISO).

In the case of experimental data, the GUM terminology refers to a quantity (i.e., the value of a physical magnitude) as the measurand. To characterize this measurand, it is necessary to provide a measure of the measurand and its unit. Still, this characterization of the measurand is incomplete without supplying the associated uncertainty. This uncertainty provides an estimate of how well the value of the measurand is known and provides a range of values that would result from equivalent measurements taken under similar circumstances with similar instruments. In general, the measurand has four general sources of uncertainty: the act of measurement, the instrument doing the measurement, the device recording the measurement, and the environment in which the measurements take place. These factors follow a basic metrology's principle, in which the accuracy of the measurement is ensured with some confidence by using common standards.

Therefore, any measurement only approximates the quantity being measured, and it is incomplete without a quantitative statement of uncertainty. Each element of a measurement system contributes to the final uncertainty of the data. For example, the accuracy of solar radiation measurements made at ground stations depends on the radiometer specifications, proper installation and maintenance, data acquisition and accuracy, calibration method and frequency, location and environmental conditions, and possible real-time or a-posteriori adjustment to the data.

A large portion of this overview of uncertainty in measurements of solar radiation made at ground stations is based on Gueymard and Myers (2008); Habte et al. (2014); Habte et al. (2016); Myers et al. (2002); Reda (2011); Stoffel et al. (2000); and Wilcox and Myers (2008).

Similarly, it is desirable that predicted and forecasted datasets be qualified with a specific uncertainty, just like measurements. In this case, the different sources of error/deviation can be classified into six categories: (i) the "imperfections" of the model itself in its mathematical description of the actual physical processes; (ii) the uncertainty in its input data; (iii) the magnitude of the error propagation from input to output, depending on the specific model, location, and period; (iv) the uncertainty in the ground-truth measurements used to validate the modeled predictions; (v) the interannual variability of the predicted dataset, if the model is validated over only a short historical period; and (vi) the possible long-term trends that will affect the predictions in the future, with the latter being most important for long-term predictions. At this point, no general procedure provides a precise quantification of each of these sources of error, let alone a reproducible determination of overall uncertainty, as discussed further in Sections 10.4–10.6. One important reason for this situation is that the first three causes of error are typically site-dependent, which considerably complicates the issue.

10.2.1 Traceability

As with any other quantity or measurand, solar irradiance requires a standard reference value and physical units, which all the measuring instruments can be compared and referred to. In the International System of Units (SI) the units of solar irradiance are those of the radiant power (in W) received by the surface of a device within a given area (in m^2). Although the modern philosophy of SI is to materialize fundamental physical units by means of universal constants, there are still many derived magnitudes based on prototypes or artifacts (a given reference instrument or a given sample of reference material), which realize and implement in practice the real values of these units. This is particularly the case of solar irradiance, which is currently referred by international consensus to a set of reference instruments of high accuracy realizing the unit of W m^{-2}, namely, the World Radiometric Reference (WRR). (As discussed in Section 3.2.3, there is progress toward using the SI-based electric W in radiometry, which will eventually make WRR obsolete.)

According to the World Meteorological Organization (WMO) guide, solar resource measurements are traceable to WRR, which is maintained by the World Radiation Center (WRC) in Davos, Switzerland (see Chapter 3). The International Pyrheliometer Comparison (IPC), which is carried out there every 5 years, is used to maintain the WRR by intercomparing the World Standard Group (WSG) radiometers and evaluating their long-term stability. The WSG is now equipped with a few absolute cavity radiometers, which constitute the group of standard instruments realizing the reference unit of solar irradiance. Their average WRR reduction factor is used to transfer the scale to other participating radiometers. Moreover, during the intervening 5 years, other agencies, such as the National Renewable Energy Laboratory (NREL), organize annual regional intercomparisons to verify and maintain the WRR factor transferred through the IPC. The transfer of calibrations from the WRR to national standards results in an expanded uncertainty for these measurement standards of ±0.45% (Reda et al. 2013).

Various methods and standards are used to transfer the WRR values to field pyrheliometers and pyranometers. The calibration and assessment of calibration and field uncertainties for

pyrheliometers and pyranometers are described in detail in national and international standards, including ASTM E824 (2018); ASTM G167 (2023); ASTM G213 (2017); ISO 9059 (1990); ISO 9060 (2018); ISO 9846 (1993); ISO 9847 (2023); and ISO/TR 9901 (2021). Figure 10-1 illustrates the process of calibration transfer from WRR to any field radiometer, along with the uncertainty added at each step.

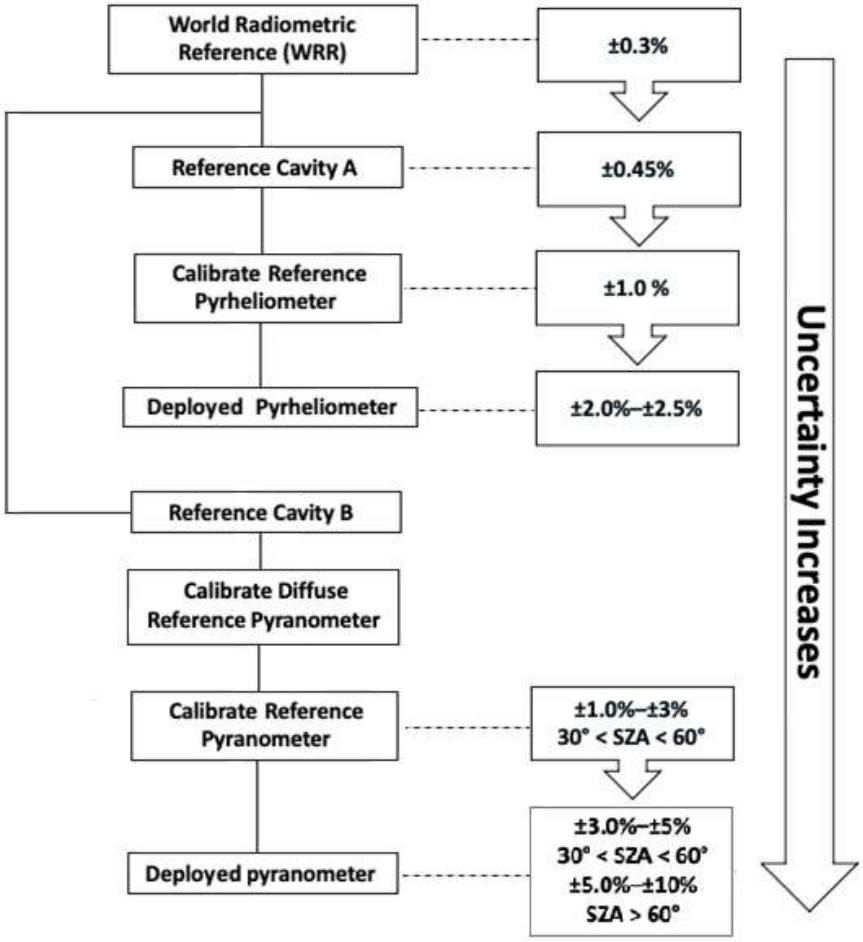

Figure 10-1. Measurement traceability and accumulation of measurement uncertainty for pyrheliometers and pyranometers (coverage factor *k* = 2)

Note: SZA stands for solar zenith angle.

Image by NREL

Similarly, spectral irradiance measurements are traceable to a national metrological laboratory (e.g., National Institute of Standards and Technology [NIST]) that has participated in intercomparisons of standards of spectral irradiance. The traceability chain and associated methods and uncertainties are exemplified in Figure 10-2. This chapter does not cover the uncertainty of spectral data, but this topic may be included in future editions of the handbook.

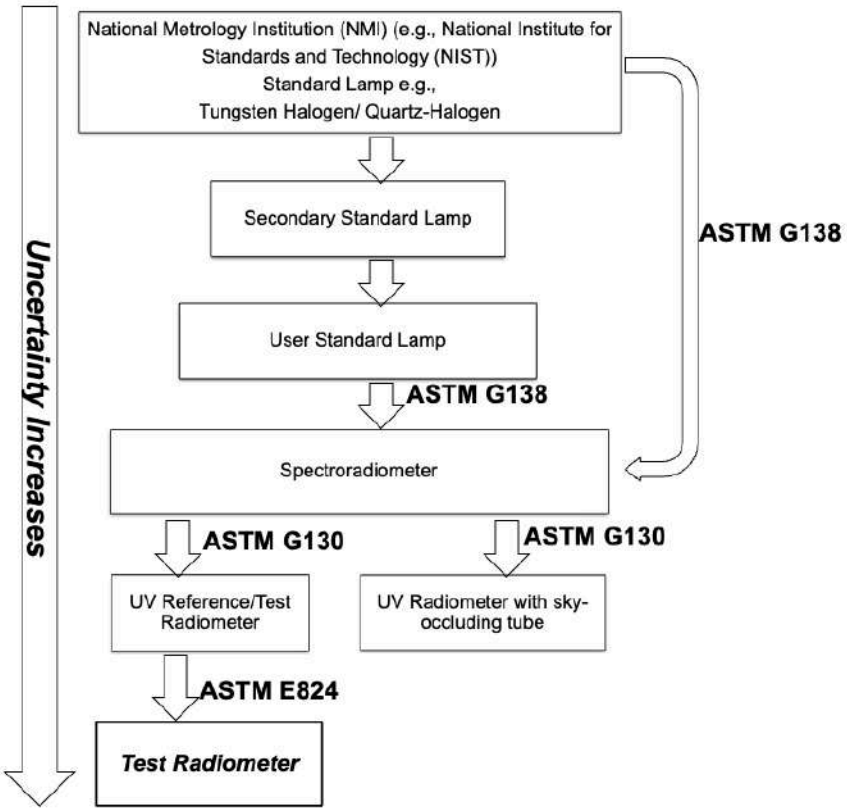

Figure 10-2. Measurement traceability for spectral irradiance

Note: UV stands for ultraviolet.

Image by NREL

Compared to the situation with irradiance measurements just described, the traceability of modeled datasets is not as straightforward. Traceability is either derived from remote-sensing (e.g., satellite or sunphotometer) information, NWP calculations, and/or statistical means. Overall, traceability is maintained through evaluation with ground measurement, as well as ongoing evaluation of uncertainty analysis and error propagation analysis. As noted above, this process is neither rigorous nor standardized yet.

10.2.2 Uncertainty vs. Error/Deviation

The difference between error and uncertainty is well documented. According to GUM (ISO/IEC 2008), the concept of uncertainty analysis supersedes the notion of error analysis. The term uncertainty describes the degree of not knowing the "true value"; it is defined as "a parameter, associated with the result of a measurement, that characterizes the dispersion of the values that could reasonably be attributed to the measurand." This concept can also be applied to predicted and forecast data as well as long-term predictions.

Conversely, error is a signed difference that is the degree of deviation of a measurement from the true value. For example, in GUM (ISO/IEC 2008), the term "error" is described as imperfections in the data that can be reduced. Error can also be explained by random and systematic errors.

Random sources are related to the standard deviation or variance of measured datasets. Biases are estimates of deviations from a "true value," primarily based on engineering judgments of the measurement system performance.

The terms "error" and "deviation" in the context of evaluating any kind of modeled data (including forecasts) typically denote deviation from "reference values" (with uncertainty as low as possible) rather than deviation from the true value. Deviations from the true value can only be approximated by this. Most often, irradiance measurements are used as reference data because they are expected to have lower uncertainty than other data. Despite this uncertainty (see Section 10.3), these data are often referred to as "ground-truth data" if their quality has been confirmed after assessment and possible improvement through appropriate QC processes (see Chapter 4). Alternatively, satellite-retrieved irradiance values, especially in the context of forecasting, or the output of a detailed physical model, might serve as a reference. Depending on the context, either "error" or "deviation" is used in the literature to characterize the difference between the modeled value and the reference value. In this chapter, the term "deviation" is used in the context of modeled historical data. It is used to emphasize the uncertainty in the reference values, for example, ground measurements.

When comparing modeled data to ground measurements, the difference can be considered statistically significant if it is greater than the uncertainty of the latter. In many cases, the uncertainty in ground measurements is not known precisely. That is why reporting the statistical metrics in terms of difference or deviation rather than error is preferred in this context.

To evaluate forecasts, on the other hand, it is preferred to report errors instead of deviations because this is the terminology generally applied in meteorological forecasting. Extrapolating in time to forecast future values adds another source of uncertainty, which leads to deviations from the true value or reference value that increase with longer forecast horizons, making the term "error" a suitable choice in this context.

10.2.3 Statistical Terms and Metrics Used to Define Accuracy

This handbook covers multiple statistical metrics used to define any deviation from the true or reference values. Here, we present the most-used error and deviation measures based on first-order statistics. A more general review is provided by Gueymard (2014) for modeled solar radiation data. A similar overview, but for forecast evaluations, is given in Jolliffe and Stephenson (2011).

10.2.3.1 Deviation of Measurements from True Value

The expected deviation of a measurement from the true value can be estimated by taking a sufficiently large number of measurements. For example, bias provides a measure of the mean overall deviation from the true value. This can be described by the MBE:

$$\mathrm{MBE} = \frac{1}{N}\sum_{N=1}^{N}(X_i - \overline{X}) \tag{10-1}$$

where N is the number of measurements, X_j is the measured value, and $\overline{X}$ is the average of the measured values. Bias can be the result of a systematic error (roughly constant over time) or of different errors that change over time but do not completely compensate for each other in the

long term. Similar to bias, the standard deviation provides a measure of the mean individual deviation from the true value.

10.2.3.2 Deviation of Modeled Historical or Forecasted Data From Reference Values

A time series of N solar radiation values X_i (measured, modeled, or forecasted) at instant i can be characterized by its average:

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{N} X_i \tag{10-2}$$

and by its standard deviation, which provides a measure of the mean individual deviation from its average:

$$\sigma(X) = \sqrt{\frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}} \tag{10-3}$$

For the comparison of modeled historical or forecasted solar radiation values X_i (e.g., GHI or direct normal irradiance [DNI]) to reference values X_i^{ref} (e.g., measured GHI or DNI), the following metrics are applied. Here the metrics are introduced using deviation, which is recommended for the evaluation of modeled historical data, as noted above. The corresponding notation using error for forecast data is given between brackets.

The deviation (error) δ_i between a single modeled (forecasted) value and the corresponding reference value (e.g., measurement) is simply:

$$\delta_i = X_i - X_i^{ref} \tag{10-4}$$

To evaluate the agreement between modeled (forecasted) data and a reference value, the RMSD (RMSE) is commonly used:

$$\text{RMSD} = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N}\delta_i^2} \tag{10-5}$$

MSD = RMSD^2 (MSE = RMSE^2), is also commonly used to characterize modeled data (forecasts).

Typically, only daytime values are considered for evaluations. Relative errors for modeled historical or forecast irradiance are generally derived by normalization with respect to the mean measured irradiance over a given time interval. In contrast, relative errors of photovoltaic (PV) power forecasts for utility applications are often normalized to the installed power rather than the mean measured value, for example, in Lorenz et al. (2011).

The RMSD (RMSE) metric can be split into two components: (1) systematic, related to the MBD (MBE); and (2) stochastic, related to the standard deviation of the deviations (errors) of single

values. The MBD (MBE) is the difference between the mean of the modeled (forecasted) data and the mean of the reference values (systematic error):

$$\text{MBD} = \overline{X_i} - \overline{X_i^{ref}} = \bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}\delta_i \tag{10-6}$$

A positive MBD (MBE) means the modeled (forecasted) values exceed the reference values on average (overestimation), while a negative MBD (MBE) corresponds to an average underestimation by the modeled (forecasted) data.

The standard deviation of the deviations (errors), SD (SE), is defined as:

$$\text{SD} = \sigma(\delta) = \frac{1}{\sqrt{N}}\sqrt{\sum_{i=1}^{N}(\delta_i - \bar{\delta})^2} \tag{10-7}$$

The SD (SE) metric provides information on the spread of the deviations (errors) around their mean value.

Ultimately, the decomposition of RMSD (RMSE) yields:

$$\text{RMSD}^2 = \text{MBD}^2 + \text{SD}^2 \tag{10-8}$$

For more detailed analyses, the SD (SE) metric might be further decomposed into one part related to the difference between the standard deviation of the modeled (forecasted) time series, $\sigma(X)$, and that of the reference time series, $\sigma(X^{ref})$, and another part related to the correlation coefficient, *r*, of the time series, which is defined as:

$$r = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(X_i^{ref} - \overline{X_i^{ref}})}{\sigma(X)\sigma(X^{ref})} \tag{10-9}$$

Overall, the complete decomposition of RMSD (RSME) yields:

$$\text{RMSD}^2 = \text{MBD}^2 + (\sigma(X) - \sigma(X^{ref}))^2 + 2\sigma(X)\sigma(X^{ref})(1 - r) \tag{10-10}$$

Another common measure to assess the accuracy of modeled (forecasted) data is the mean absolute deviation (MAD):

$$\text{MAD} = \frac{1}{N}\sum_{i=1}^{N}|\delta_i| \tag{10-11}$$

The metrics reviewed above do not have the same importance, depending on application. In solar resource assessments, long-term yield predictions, and various other aspects of solar energy utilization, bias (MBD) can be considered “enemy number one” because it translates into systematic underestimations or overestimations of the solar power plant’s output. In turn, this can have damaging consequences on the plant’s financing, profitability, or viability. A bias in the estimated solar resource of just a few percentage points can make a big difference at the design and financing stages. To characterize random errors, the use of RMSD is more frequent than that of MAD. Because the solar resource is characterized over long periods (1 year to a few decades), and because random errors tend to decrease rapidly as the time integration increases, RMSD

tends to a limit close to |MBD| by virtue of Eq. 10-8, which is another reason why its role is secondary.

The situation is different for the application of short-term forecasting of solar irradiance and power for grid management, energy management, or marketing. Here, the errors of every single forecast value matter because balance between demand and supply must be maintained at all times. Both negative and positive forecast errors have negative consequences, like penalties or need for balancing power that do not balance out over time. This makes MAE and RMSE the most crucial metrics for deterministic forecasting.

MAE is recommended by Hoff et al. (2013) as a preferred measure in solar forecasting, in particular for reporting relative errors. In contrast, RMSE is recommended by Yang et al. (2020). An in-depth discussion on the respective merits of MAE and RMSE is given by Hodson (2022). From a user's point of view, the choice of the most suitable error measure(s) should be based on the impact of the prediction or forecast errors on their application. MAE is appropriate for applications with linear cost functions (i.e., when the costs caused by inaccurate results are proportional to the prediction error). RMSE is more sensitive to large forecast errors and is therefore suitable when small errors are tolerable and larger errors cause disproportionately high costs, which is the case for many applications in the energy market and for grid management issues.

10.2.4 Visual Diagnostics

In addition to the computation of statistical error measures, creating effective figures that demonstrate visual analysis is strongly recommended. These figures are interesting to show a direct comparison between measurements and modeled or forecast data, with the aim of developing a better understanding of the performance. Many scientific papers such as Badosa et al. (2014); Forstinger et al. (2023); Gueymard (2014); Habte et al. (2016); Markovics and Mayer (2022); Sengupta et al. (2018); Vuilleumier et al. (2014); Yang (2020); and Yang et al. (2020) demonstrate effective figures to communicate the result of solar resource and/or solar energy research. Yang (2020) illustrates violin plots that are hybrids of boxplots and kernel density plots and explain summary statistics as well as the density of the variables.

As an example, Figure 10-3 illustrates the combination of a scatterplot with a boxplot. This arrangement explains the complex relationship between the measured and modeled data. The boxplot provides information about the median, lower and upper quartile, and interquartile ranges. Obviously, the scatterplot provides a comprehensive overview of the datasets and illustrates the strength of their relationships. In parallel, the time series plot shows a comparison of successive time intervals, constituting an excellent way to visualize the possible features of a long-term dataset by providing detailed and intuitive information on patterns or relationships caused by, for example, sky conditions, seasonality, or outliers. However, such time-series plots illustrating specific patterns show only a small part of the data. Moreover, the spatial and temporal visualization of aggregated statistical metrics is important to identify any spatial and temporal pattern in relation with geographic or climate signature, for example.

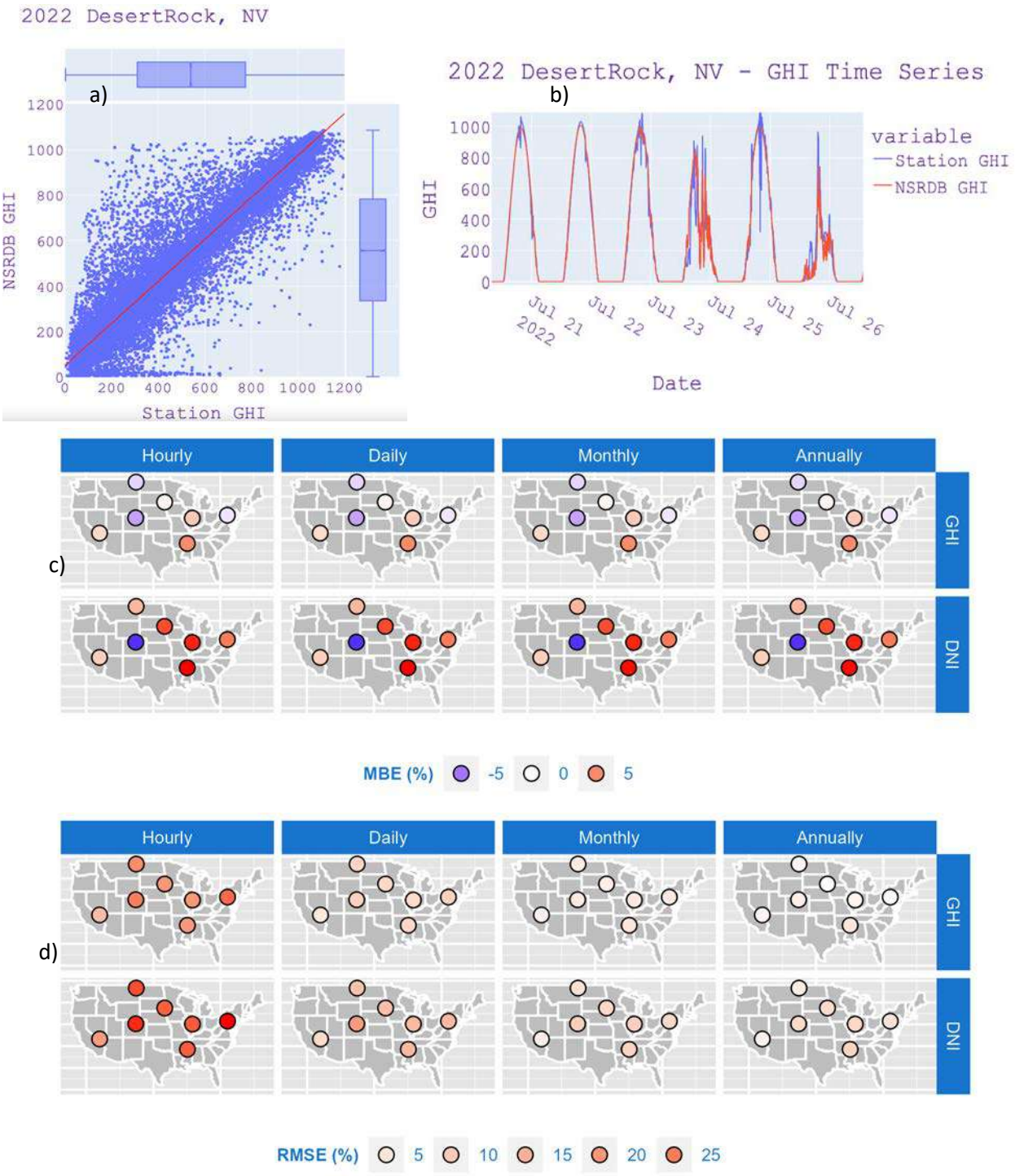

Figure 10-3. Example plots showing various comparisons of National Solar Radiation Database (NSRDB) predictions vs. ground irradiance measurements in the United States: (a) scatterplot and (b) time series plot for GHI at Desert Rock, Nevada; and plots depicting the temporal and spatial distribution of GHI and DNI for seven National Oceanic and Atmospheric Administration (NOAA) Surface Radiation Budget (SURFRAD) locations at different time scales, considering both MBE (c) and RMSE (d).

The 2022 NSRDB data based on the Physical Solar Model (PSM) V4 model was used to generate all plots. Image by NREL

10.3 Estimating Measurement Uncertainty

Uncertainty in solar radiance measurement is dependent on the type of radiometer and on the irradiance component to be measured, such as DNI, GHI, global titled irradiance (GTI), or diffuse horizontal irradiance (DHI). For each component, the measurement equation can be different, and the evaluation of the uncertainty can vary, even when following the same derivation principles. In general, the uncertainty will be estimated as a function of several contributions: calibration of the sensor, its characteristics or specifications, measurement of the output signals, working conditions, etc. Every contribution can have a different weight or relevance, so that not all of them are equally important.

It is important to note here that the uncertainty of the solar irradiance measured by a radiometer is always greater than the uncertainty of its calibration. For instance, in the case of well-maintained, high-quality thermopile radiometers deployed in the field, factors such as accuracy of solar tracking and/or leveling, data logger accuracy, cleanliness of the windows, and frequency of recalibration could contribute more sources of uncertainty. Detailed uncertainty analyses for high-quality field pyrheliometers can be found in ASTM G213 (2017); Balenzategui et al. (2022a; 2022b); Habte et al. (2016; 2017); and Michalsky et al. (2011). Similarly, the study by Vuilleumier et al. (2014) includes field pyranometers, though the term "field instrument" might be misleading here, because most of these studies refer to instruments located at research-class stations and operated under quasi-laboratory conditions (i.e., with optimal calibration and maintenance). In the practice of solar resource assessments, particularly those involving temporary stations under harsh conditions, instruments are typically maintained on a more sporadic schedule, implying that additional uncertainties would apply. In any case, the aforementioned studies show that the uncertainty of calibration is one of the most important contributions to the overall uncertainty for well-maintained high-quality instruments. Calibration of a radiometer usually consists of the determination of its responsivity, R, or the relationship between its output signal (current or voltage) and the incident solar irradiance. Calibration methods depend on the type of radiometer under test and on the type of radiometer used as reference instrument; they are normally specified by international standards (see Chapter 3).

The calibration stability of the present commercially available pyrheliometers and pyranometers is generally satisfactory, as revealed by only a slight change in their R value—typically less than ±0.1% and ±0.2% per year, respectively. When finally deployed in the field, factors such as accuracy of solar tracking and/or proper leveling and orientation, data logger accuracy, cleanliness of the windows and domes, and frequency of recalibration could contribute more sources of uncertainty. Even if these effects are kept low by following measurement and maintenance best practices, expanded uncertainties of ±2.0%–±2.5% in DNI measurements and ±3.0%–±5.0% in GHI measurements have been found from a high-quality measurement system (Reda 2011). As mentioned above, field instruments deployed at solar resource assessment stations in harsh environments can be expected to have greater uncertainties, particularly in the absence of a stringent maintenance program.

Moreover, the ASTM G213 (2017) standard provides guidance and recommended practices for evaluating uncertainties when calibrating and performing irradiance measurements with pyranometers and pyrheliometers. The standard follows the GUM method and attempts to quantify the uncertainty in measuring irradiance. Further, the standard aims to maintain the

measurement traceability through WRR with respect to SI, which ensures that the uncertainty quoted for radiometric measurements can be intercompared, based on documented methods of derivation. Figure 10-4 shows an example of the contribution of uncertainty from each source, expressed either in absolute values or percentages. Some sources of uncertainty contribute more than others, but also, the relative importance of the uncertainty budget varies during the day, with a total uncertainty that increases significantly with SZA, for reasons explained next.

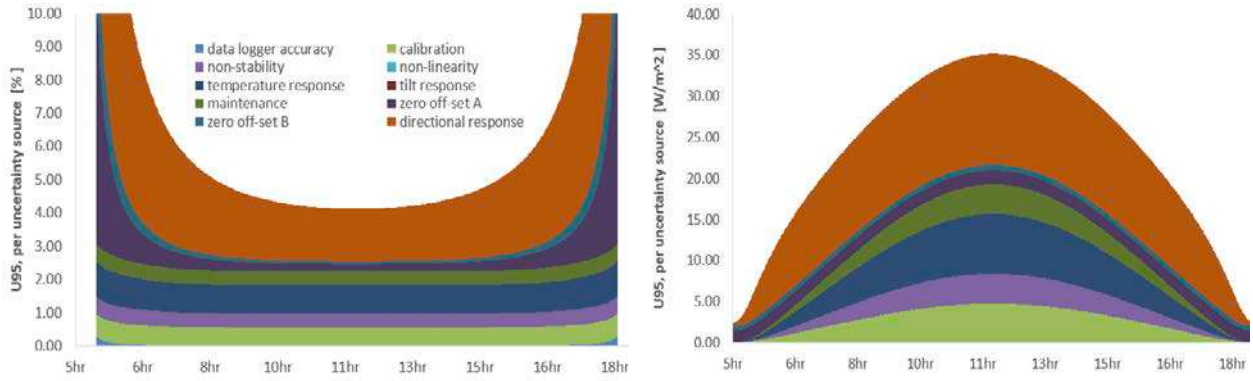

Figure 10-4. Example of expanded uncertainty as a function of time and source of uncertainty contribution, based on ASTM G213 (2017); left image, expressed in percentages, and right, expressed in W m^{-2}

Image by NREL

In addition to being sensitive to the magnitude of GHI, the responsivity of most pyranometers depends on multiple factors such as the angle of incidence of the beam irradiance component. This explains why the typical shade/unshade calibration uncertainty (see, e.g., ISO 9846 [1993]; WMO [2018]) of any thermopile pyranometer placed in a horizontal position with respect to a WRR reference cavity radiometer is ≈0.5% at any very narrow range (±2°–±5°) of SZA (Reda et al. 2008). Typically, during calibration, *R* is selected as an average responsivity for a specified SZA over one or more days. In the field, however, the monitored irradiance is sensed over a wide range of SZAs (up to 0–90°), and the measurement uncertainty over the whole range is larger. As mentioned in Chapter 3, for some pyranometers, *R* can vary by ±3–±10% or even more over this zenith angle interval. These effects then need to be combined with all other potential sources of error in the field (e.g., pyranometer installation, data logger accuracy, cleanliness, spectral dependency, or temperature sensitivity).

If only one *R* is used for a wide range of solar angles, that value is often derived for relatively low SZAs (high solar elevation), thus making the highest irradiance values (on average) associated with the lowest possible uncertainty. Another option, which is the standard procedure at NREL, for instance, is to report *R* for a fixed SZA=45° because that geometry is more representative of all daily situations. The variation of responsivity with SZA and azimuth angles is typically greater for high SZAs; thus, large uncertainties usually occur at high SZAs. These high-SZA-related uncertainties occur throughout parts of the day (early morning and late afternoon) when the available solar resource is much smaller than typical midday values and/or when SZAs are smaller. Because the minimum SZAs vary significantly throughout the year (except close to the equator), the uncertainty in hemispherical radiation data will vary as well. This effect is

especially important for latitudes beyond ±45°, when SZA is rarely smaller than or equal to the SZA at which the responsivity of the pyranometer was determined.

Even when good measurement conditions exist, such as near midday under clear-sky conditions, the uncertainty in hemispherical global or diffuse measurements is typically two to three times that of direct-beam measurements, or ±3–±5% throughout a year, primarily because of seasonal variations in uncertainty. Better instrumentation design and careful applications of correction factors as a function of SZA are ways to improve (reduce) the uncertainty in GHI measurements. The alternative is to use high-quality DNI and DHI measurements using a tracking device (e.g., a disk or a ball) to derive GHI from the closure equation (Michalsky et al. 1999). The expanded uncertainty for this calculated GHI then approaches that of DNI (±2%) for clear-sky measurements. One limitation of this method, however, is that it assumes "perfect" operating conditions, such as correct tracking for both DNI and DHI. Slight misalignments of tracking and complete tracker failures do happen in practice, resulting in large errors in all three components, unless the errors are properly and rapidly detected during the QC procedure, which is difficult in practice.

Figure 10-1 describes the calibration traceability for pyrheliometers used to measure DNI and for pyranometers used to measure GHI or DHI. The figure indicates how uncertainties accumulate from calibration to field deployment. Broad arrow boxes show the accumulated expanded uncertainty at each phase of the process. The resulting uncertainty in field deployment for pyrheliometers is ±2.0–±2.5% in this example, assuming regular and high-quality maintenance. Measurement uncertainties for pyranometers used to measure GHI in the field range from ±3.0–±5% for SZAs between 30° and 60°, but are higher for angles greater than 60°, again assuming regular and high-quality maintenance.

Calibrations of pyranometers can be performed horizontally (for GHI) or at tilt (for GTI). More specifically, pyranometers measuring GHI are calibrated horizontally using either GHI or combined DNI and DHI measurements as a reference. Calibration for a pyranometer intended to sense GTI is done using a reference pyranometer installed on the same exact tilt (ASTM E824 2018). Tilting a pyranometer for GTI measurements can slightly alter its responsivity compared to its horizontal position because of, for example, changes in convection patterns inside the dome or changes in thermal offset. This typically affects the calibration uncertainty of GTI measurements. Some thermopile pyranometers are not designed for tilted measurements, and at certain times of the day, direct sunlight can strike their unshaded bodies, affecting measurements. Shielding can reduce or eliminate this problem. Calibrating a tilted pyranometer with a reference instrument of a different type (or make and model) might also introduce additional uncertainty. To help evaluate the uncertainty in GTI data, the metadata of such datasets should include shielding information.

This caveat also holds for the measurement of upwelling irradiance using a down-facing pyranometer. (This measurement is necessary to obtain the surface albedo by dividing it by GHI, see Chapter 3.)

Digital radiometers have been recently introduced on the market and are now deployed in many solar energy projects. This also brings new challenges in terms of uncertainty quantification. Some digital radiometers, for example, include a built-in temperature compensation feature,

among other things. In such a case, the uncertainty of the calibrated internal sensitivity with temperature coefficients must be applied, and the contribution of the coefficients to the uncertainty should be clearly identified. More research is needed to understand the propagation and relationships of various sources of uncertainty related to digital radiometers.

For rotating shadowband irradiometers (RSIs) and photodiode pyranometers, which are typically used in Tier-2 stations (see Chapter 3), one of the most crucial impacts on uncertainty is the spectral irradiance error. This is because silicon photodiode sensors detect only visible and infrared radiation in the range ≈300–1100 nm and have a spectral response that varies strongly within this wavelength interval (see Chapter 3). Further, the role of using algorithms to reduce systematic effects and the uncertainty introduced by imperfect shading must be considered. A more detailed uncertainty analysis for RSIs following GUM can be found in Wilbert et al. (2016). The study defines a method for the derivation of the spectral error and spectral uncertainties, and presents quantitative values of the spectral and overall uncertainties. The results of this detailed analysis and other studies such as Wilcox and Myers (2008) indicate lower overall uncertainties than those presented in Table 10-2 for silicon photodiode pyranometers because the results in the table assume that no rigorous correction is applied. The expanded measurement uncertainty for subhourly DNI measurements is approximately ±7% for a photodiode RSI radiometer with state-of-the-art correction functions for systematic errors. Similarly, RSI-based GHI measurements are found to be affected by slightly lower uncertainties than DNI (6%, $k = 2$, after application of advanced adjustment functions; see Chapter 3). Moreover, advanced adjustment functions can significantly reduce the uncertainty in both GHI and DNI. In parallel, considering the lower incidence of soiling effects on RSIs than on thermopile pyranometers, the use of advanced adjustment functions can bring RSI measurements at resource assessment stations almost on par with those from reference instruments (Al-Rasheedi et al. 2018).

The average uncertainty of an irradiance time series is expected to vary from one station to another and even from time stamp to time stamp for a specific station. An individual uncertainty analysis per station and time interval is complex and also depends on the applied QC. After detailed QC and the rejection of suspicious data, a significant part of the variation of the uncertainty with station and time is removed. As a simplification, the uncertainty of DNI and GHI at either Tier-1 or Tier-2 stations can be estimated, based on the methods described in this section. In problematic cases and for the most relevant zenith angles, the QC procedure for stations with an independent measurement of the 3 components (GHI, DHI, and DNI) flags data with deviations larger than 8% between calculated (using the closure equation) and measured GHI. Assuming good maintenance, the uncertainty of the used data is expected to be lower than this QC-based limit and close to the mentioned uncertainties.

As indicated above, the standard uncertainty for well-maintained Tier-2 stations is estimated as 7% and 6% for DNI and GHI, respectively. As the three-component test cannot be performed at such stations, the upper uncertainty limit that is related to the QC procedure for these stations is higher than in the case of the Tier-1 stations with independent measurements of DNI, GHI, and DHI, and can be expected to be ≈10%. At Tier-1 stations, the uncertainty is expected to be lower than the QC-related limit and close to the estimations for well-maintained stations.

10.3.1 Method for Quantifying Uncertainty: The GUM Method

The method for estimating uncertainty has changed significantly over the last few decades. The general adaptation to the current methodology takes time, so some outdated terminology and methods still appear in the literature and might be in use by the industry. Even though the use of outdated methodologies is discouraged, short descriptions are provided to help users understand and correctly use uncertainty data based on older methodologies.

10.3.2 Practical Examples

GUM is currently the accepted guide for measurement uncertainty (ISO/IEC 2008). The method provides the expanded uncertainty for a 95% confidence interval by multiplying the combined uncertainty by the coverage factor k (k = 1.96 for a Gaussian distribution for infinite degrees of freedom; it is often approximated as 2, which is also alternatively used in this chapter). GUM defines Type-A uncertainty contributions as derived from statistical methods and Type-B sources as evaluated by other means, such as scientific judgment, experience, specifications, comparisons, and calibration data. GUM defines the concept of a standard uncertainty (u_{std}) for each uncertainty type, which is an estimate of an equivalent standard deviation (of a specified distribution) of the source of uncertainty. To appropriately combine the various uncertainties, the GUM methodology uses a sensitivity coefficient (c) that is calculated from the measurement equation using partial derivatives with respect to each input variable in the equation. GUM removes the historical factor of 2 and introduces the coverage factor k (whose value depends on the known or assumed statistical distribution of uncertainties),[61] which is applied to compute the expanded uncertainty (U_E) as:

$$U_E = k \cdot u_c \tag{10-12}$$

where:

u_C is the combined standard uncertainty, per all the steps of the GUM summary below (points 1–6).

As shown in Figure 10-5, the GUM procedure can be summarized in six steps (Konings and Habte 2016; Reda 2011):

1. ***Define the measurement equation for the calibration and/or measurement system.*** This consists of a mathematical description of the relation between sensor voltage and any other independent variables and the desired output (calibration response or engineering units for measurements). The example equations used to quantify radiometric measurement are:

$$E = \frac{(V - R_{net} \bullet W_{net})}{R} \tag{10-13a}$$

$$E = \text{DNI} \bullet \cos(Z) + \text{DHI} \tag{10-13b}$$

[61] k is 1.96 for a Gaussian distribution for a 95% confidence level. Generally, a 95% confidence level means that 95% of the values will be within the statistical limits defined by the uncertainty.

where:

- E = irradiance, in W m^{-2} (GHI, GTI, DHI, or DNI); in particular, when E stands for DNI, DHI = 0 in Eq. 10-13b and $R_{net} \approx 0$, resulting in a simplified version of Eq. 10-13a: $E = V / R$
- R = responsivity of the radiometer in µV/ (W m^{-2})
- V = sensor output signal (e.g., voltage or current) of the radiometer (e.g., µV, mA)
- R_{net} = net infrared responsivity of the radiometer in µV/(W m^{-2})
- W_{net} = effective net infrared irradiance measured by a collocated pyrgeometer in W m^{-2}.

In the case of GHI, the closure equation (10-13b) applies, in which:

- DNI = beam irradiance measured by a primary or standard reference standard pyrheliometer in W m^{-2}
- Z = SZA, in degrees or radians
- DHI = diffuse horizontal irradiance, measured by a shaded pyranometer (W m^{-2}).

2. ***Determine the sources of uncertainty.*** Most sources of uncertainty are obtained from statistical calculations, specifications from manufacturers, and previously published reports on radiometric data uncertainty or professional experience. Some common sources of uncertainty are associated with the cosine response, spectral response, nonlinearity, temperature response, thermal loss, data logger accuracy, soiling, and calibration, including the drift of the calibration constant(s).

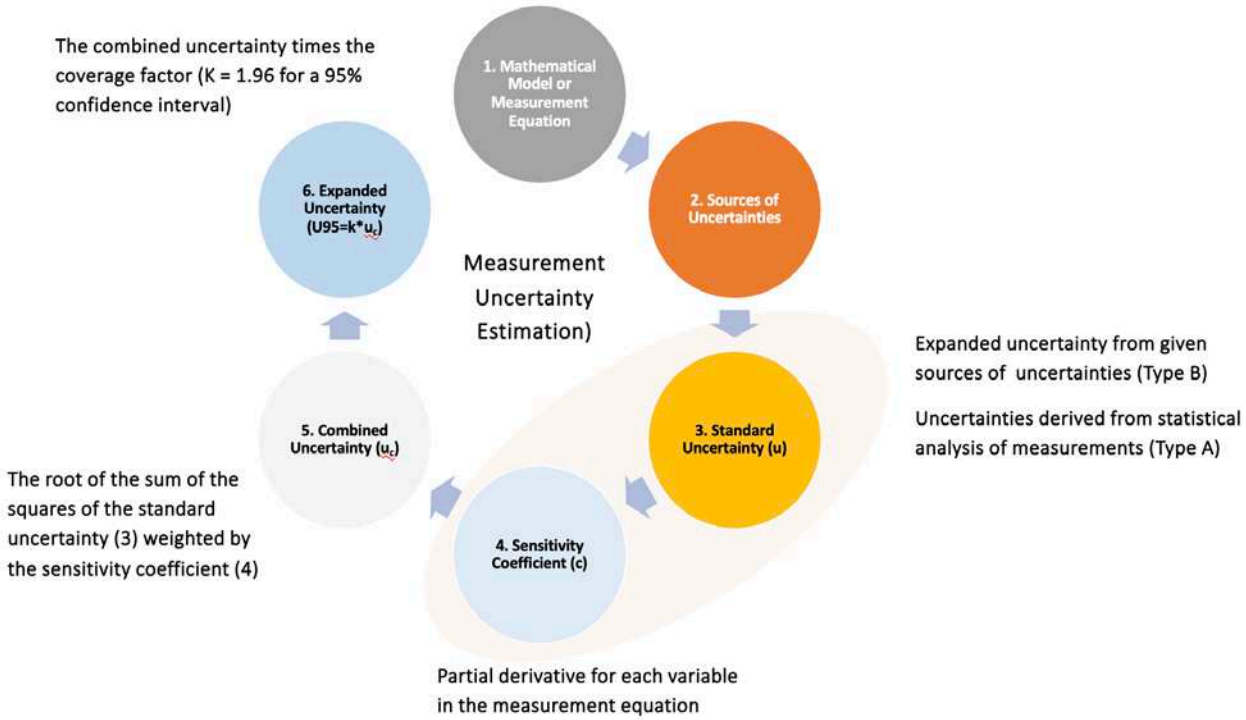

Figure 10-5. Measurement uncertainty estimation flowchart

Image modified from Habte et al. (2016)

3. ***Calculate the standard uncertainty, u.*** In this step, an individual *u* for each variable in the measurement equation is calculated using either a statistical method (Type-A uncertainty component) and/or other methods (Type-B uncertainty component). In the GUM method, the standard uncertainties are calculated by dividing the expanded uncertainty of each source by a factor corresponding to the specific statistical distribution of the experimental data (ASTM G213 2017).

   A. Type-A uncertainty:

      Type-A standard uncertainty (u) is calculated when taking repeated measurements of the input quantity value, from which the sample mean and sample standard deviation (SD or σ) can be calculated, resulting in Eq. 10-13:

$$u^2 = \frac{\sigma^2}{n} \qquad \text{where:} \qquad \sigma^2 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1} \tag{10-14}$$

   B. Type-B uncertainty:

      Type-B uncertainties are often provided (e.g., in calibration certificates) as an expanded uncertainty (U). To be consistent with Type-A uncertainties, the standard Type-B uncertainties, u, are calculated from the expanded uncertainties, U, using one of the three following methods:

      i. Equation for unknown statistical distribution (common assumption: rectangular distribution): $u = U/\sqrt{3}$, where U is the expanded uncertainty of a variable
      ii. Normal distribution: $u = U/k$, where k is a coverage factor of 2 or, more exactly, 1.96 (ISO/IEC 2008)
      iii. For other statistical distributions, the applicable values for k are used.

4. ***Compute the sensitivity coefficient, c.*** To appropriately combine the various uncertainties in the next step, the uncertainties must be weighted. According to the GUM method, this is done by calculating the sensitivity coefficients (c) of the variables in a measurement equation. These coefficients affect the contribution of each input factor to the combined uncertainty of the irradiance value. Therefore, the sensitivity coefficient for each input is calculated by partial differentiation with respect to each input variable in the measurement equation. Table 10-1 shows those sensitivity coefficients applicable to radiation measurements.

   The sensitivity equations given in Table 10-1 are for two distinct situations. The calibration sensitivity is for calibrations when the reference GHI is calculated from reference DNI and DHI measurements. The second column is for GHI measurements in the field. The calibration sensitivities are related to the inverse of the GHI value, whereas the field sensitivities are related to the inverse of the responsivity.

Table 10-1. Example of Computing Sensitivity Coefficients for GHI Pyranometer Calibration and Measurement Using Partial Derivatives

Calibration Sensitivity Equations	Field Measurement Sensitivity Equations
$c_V = \frac{\partial R}{\partial V} = \frac{1}{\mathrm{DNI}\cos(Z) + \mathrm{DHI}}$	$c_R = \frac{\partial \mathrm{GHI}}{\partial \mathrm{R}} = \frac{-(V - R_{net}\, W_{net})}{R^2}$
$c_{Rnet} = \frac{\partial R}{\partial R_{net}} = \frac{-W_{net}}{\mathrm{DNI}\cos(Z) + \mathrm{DHI}}$	$c_{R_{net}} = \frac{\partial \mathrm{GHI}}{\partial R_{net}} = \frac{-W_{net}}{R}$
$c_{Wnet} = \frac{\partial R}{\partial W_{net}} = \frac{-\mathrm{R_{net}}}{\mathrm{DNI}\cos(Z) + \mathrm{DHI}}$	$c_{W_{net}} = \frac{\partial \mathrm{GHI}}{\partial W_{net}} = \frac{-R_{net}}{R}$
$c_{DNI} = \frac{\partial R}{\partial \mathrm{DNI}} = \frac{-(V - R_{net}\, W_{net})\cos(Z)}{(\,\mathrm{DNI}\cos(Z) + \mathrm{DHI})^2}$	$c_V = \frac{\partial \mathrm{GHI}}{\partial V} = \frac{1}{R}$
$c_{SZA} = \frac{\partial R}{\partial Z} = \frac{\mathrm{DNI}\sin(Z)\,(V - R_{net}\, W_{net})}{(\,\mathrm{DNI}\cos(Z) + \mathrm{DHI})^2}$	
$c_D = \frac{\partial R}{\partial \mathrm{DHI}} = \frac{-(V - R_{net}\, W_{net})}{(\,\mathrm{DNI}\cos(Z) + \mathrm{DHI})^2}$	

5. ***Calculate the combined standard uncertainty, u_c.*** This is the combined standard uncertainty using the propagation of errors formula and quadrature (square root sum of squares) method. It is applicable to both Type-A and Type-B sources of uncertainty. Standard uncertainties (u) in Step 3 multiplied by their sensitivity factors (c) in Step 4 are combined in quadrature to give the combined standard uncertainty, u_c:

$$u_c = \sqrt{\sum_{j=0}^{n} (u_j \bullet c_j)^2} \tag{10-15}$$

   where n is the number of uncertain variables that are used to calculate the combined uncertainty.

6. ***Calculate the expanded uncertainty (U_{95}).*** The expanded uncertainty is calculated by multiplying the combined standard uncertainty by the coverage factor, typically by applying the student's t-analysis to determine the appropriate value of k (typically 1.96 for 95% and 3 for 98% confidence, respectively, for large datasets assuming a Gaussian distribution):

$$U_{95} = k \bullet u_c \tag{10-16}$$

These six steps, also described in Figure 10-5, demonstrate that the uncertainty quantification is a cycle. This means that one can use the expanded uncertainty in Step 6 as an input to a measurement equation. This would be the case, for example, in calculations of the performance ratio of solar conversion systems: to calculate the ratio of system output/solar input, the expanded uncertainty in Step 6 is used as an input to evaluate the denominator (solar input), and the cycle continues to ultimately quantify the expanded uncertainty of the performance ratio.

Further, these steps are applicable to the quantification of the uncertainty in both calibration and field measurements. Uncertainty in measurements begins with the uncertainty in calibration references, calibration processes, and sensor design characteristics. For example, for thermopile sensors, a calibration constant is required to convert the output voltage to the required irradiance, as discussed in Chapter 3. The resulting uncertainty in calibration factors must then be combined with the influence of additional sources of uncertainty in the field measurement instrumentation, installation methods, data acquisition, and operations and maintenance processes (Reda 2011). Overall, estimates of uncertainties for the magnitudes of values (e.g., voltage, R_{net}) need some (documented) experimental, theoretical, or other (specifications) sources. These sources of uncertainty are the magnitudes adjusted in these steps—for example, in the sensitivity coefficients calculation. Such example data are presented in several references (ASTM G213 2017; Habte et al. 2014; Konings and Habte 2016; Reda 2011).

Users must pay close attention to the sources of uncertainty. For instance, the SZA uncertainty includes sources of error such as accuracy in latitude and longitude, air pressure (for refraction corrections), or timekeeping (clock accuracy). The units of these variables must be treated carefully and consistently, whether they are percentages (such as of full scale or reading) or absolute units (such as volts, degrees, or W m^{-2}). Additionally, it is essential to consider the symmetry of the sources of uncertainty. In this section, all sources of uncertainty are considered symmetrical (±); however, some other sources could be asymmetrical or one-sided. For example, Konings and Habte (2016) considered non-stability and zero offset of Type-A as one-sided sources of uncertainty.

Applying the GUM procedure to the case of pyrheliometer and pyranometer calibration, Table 10-2 summarizes the estimated uncertainties that are typically found in practice. In addition, the table identifies the typical sources of uncertainty considered for the overall uncertainty analysis of irradiance measurements from two types of radiometers: radiometers with thermopile detectors and photodiode radiometers with silicon detector (before the application of correction functions for systematic errors). Note that the contribution to uncertainty caused by insufficient maintenance (alignment, leveling, cleaning, etc.) can be much greater than the combined uncertainties for well-maintained instruments. As explained in Chapter 3, instruments with clear optics (such as most thermopile radiometers) are more strongly affected by soiling; therefore, the uncertainty related to their operation in the field directly depends on the regularity and quality of their maintenance over time.

Table 10-2. Example of Estimated Expanded Uncertainties at 95% confidence interval of Responsivities of Field Pyranometers and Pyrheliometers

Modified from Reda (2011)

Type-B Uncertainty Source	Thermopile Pyranometer (%)	Photodiode Pyranometer (%)	Thermopile Pyrheliometer (%)	Photodiode Pyrheliometer (%)
Calibration[a]	3	5	2	3
Zenith Response[b]	2	2	0.5	1
Azimuth Response	1	1	0	0
Spectral Response	1	5	1.5	8
Tilt[c]	0.2	0.2	0	0
Nonlinearity	0.5	1	0.5	1
Temperature Response	1	1	1	1
Aging per Year	0.2	0.5	0.1	0.5
U_{95}	**4.1**	**8.0**	**2.7**	**8.9**

[a] Includes zenith angle responses from 30° to 60°.
[b] Includes zenith angle responses from 0° to 30° and from 60° to 90°.
[c] This uncertainty is set to zero for untilted radiometers.

10.4 Estimating the Uncertainty of Modeled/Predicted Datasets

Solar radiation can be modeled in many ways, depending on the available inputs, origin (ground-based, satellite-based, or NWP-based), application requirements (e.g., clear-sky or all-sky conditions), and degree of detail (broadband or spectral irradiance).

Satellite-based models used to estimate solar radiation can use a physics-based approach relying on radiative transfer modeling, an empirical or semiempirical approach relating the reflected radiance sensed by the satellite sensor directly to surface radiation, or a mix of both (see Chapter 7).

Models developed using empirical or semiempirical correlations between ground-based irradiance measurements and reflected radiance observations from satellite sensors inherently carry the uncertainty of these measurements. This uncertainty is embedded in the ultimate model accuracy, along with the uncertainties associated with the satellite sensors and the modeling process. Models empirically based on ground-based irradiance measurements with 2%, 5%, or 10% uncertainty cannot have a lower uncertainty than the data used to derive and/or validate the model. Similarly, models based on first principles of physics and radiation transfer cannot be validated or verified to a level of accuracy greater than that of the ground-based irradiance measurements. A thoroughly documented uncertainty analysis of these measurements (Gueymard and Myers 2008; 2009; Habte et al. 2016; Vuilleumier et al. 2014) is necessary to ascertain the validity of model accuracy claims. The effect of biases on ground-based irradiance measurements should be part of any model analysis.

An understanding of the differences between the perspectives of satellite-derived irradiance estimates and ground-based measurements is essential when the latter are used to derive and validate satellite-derived irradiance values. Observations of a specific pixel (or grid cell) by a spaceborne radiometer ultimately provide (after substantial modeling) an estimate of surface radiation based on the estimated properties of those clouds and other atmospheric constituents spread throughout that pixel or a larger area. In contrast, surface irradiance observations are made by an instrument viewing the sky from a specific point. If the satellite pixel size is small enough, parallax errors enter into the comparison. Conversely, if it is too large, the radiation field over the pixel might not be homogenous enough for a correct comparison. Terrain effects could also influence a comparison in which cloudiness, elevation, and/or topographic shading could vary within a short distance. Often the data available for satellite modeling lack the exactitude for differentiating fine variations seen by ground-based measurements. Another intricate situation results from the fact that the clear-sky part of a satellite-based radiation model typically has a much coarser spatial resolution than the cloudy part. Whereas the latter is up to 1 km in most modern products, the former is, for example, 0.5 by 0.625° (or ≈65 km) when the atmospheric input data for the clear-sky radiation model are extracted from Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2). Thus, the actual resolution of the irradiance product ultimately depends on the cloudiness conditions. These issues can be compounded by the fact that ground measurements represent an average irradiance value calculated over a fixed time interval (e.g., 1 minute or 10 minutes), whereas satellite-based model predictions solely rely on instantaneous snapshots taken at different intervals (e.g., every 10 minutes).

10.4.1 Statistical Metrics

To alleviate the absence of any standardized method for accuracy assessment and uncertainty calculation, many possible statistical metrics used in the literature have been reviewed (Gueymard 2014). Still, most authors report only the RMSD and MBD (or RMSE and MBE), that is, randomness and bias (absolute or relative). As an example, the model of Darnell et al. (1988) was used to evaluate surface radiation using cloud information from the International Satellite Cloud Climatology Project C1 cloud database. The results were then compared to surface observations collected by the World Radiation Data Center in Darnell et al. (1992). The RMSD from this comparison was ≈16 W m^{-2}, and the MBD was ≈4 W m^{-2}. Note that the interpretation of the reported sources of uncertainty depends on the spatial and temporal resolution of the data being compared (random errors tend to decrease rapidly with increasing averaging period) and that the relative uncertainties in the modeled DNI are always greater than in GHI—opposite to what occurs with high-quality measurements.

According to Perez et al. (1987), satellite-based retrievals of DNI were accurate to 10–12%. Later, Renné et al. (1999) and Zelenka et al. (1999) found that the target-specific comparison to ground-based observations had a relative RMSD of at least 20%; the time-specific pixel-wide accuracy was 10–12% on an hourly basis at the sites under scrutiny. Most accuracy results contain values that are proportional to the measured values (percentage), given that the measured values are within a certain range and specifications are related to a fixed value in W m^{-2}. The validation of satellite-based irradiance predictions is sometimes performed on a daily (instead of hourly or subhourly) timescale. This might not always be appropriate, however, particularly in areas where strong morning/afternoon cloudiness asymmetries exist (Salazar et al. 2020).

From an application or use of statistical metrics perspective, using both MAD and RMSD is not necessary and can lead to misinterpretations. This stems from considerations of statistical consistency that become important when verifying the accuracy of modeled data—most particularly in the case of forecasts (Section 10.6). In summary, if a model is optimized by minimizing the squared error, RMSD is consistent, but MAD is not (Yang et al. 2024). More in-depth theoretical details are provided by Gneiting (2011).

From a solar resource standpoint, the most important error measure is MBD by far, because any bias in the predicted resource leads to a similar bias in the estimated power production over the long term, which can put the whole project at risk, either at the financing stage or later if the actual production does not meet expectations. Although random errors in modeled irradiance estimates can be large when considering short time intervals (e.g., a few minutes), they decrease rapidly when integrating over time. They normally reach a low value when averaging over one or more years, which is the typical time frame used for resource assessments.

Moreover, bias is actually part of RMSD, per Eq. 10-8. It indicates that, even if a long averaging period is considered, such as ≈15 years, RMSD can approach the absolute value of MBD, |MBD|, but can never be lower. In contrast, good measurements have relatively small random errors, so that their total uncertainty does not change much over time. Thus, when comparing modeled irradiance estimates to reference ground-truth measurements, the RMSD of the former decreases over the averaging time and tends toward a limit that is either the uncertainty of the reference measurements or the |MBD| of the modeled results, whichever is greater. This is exemplified in Figure 10-6, using modeled data from the NSRDB.

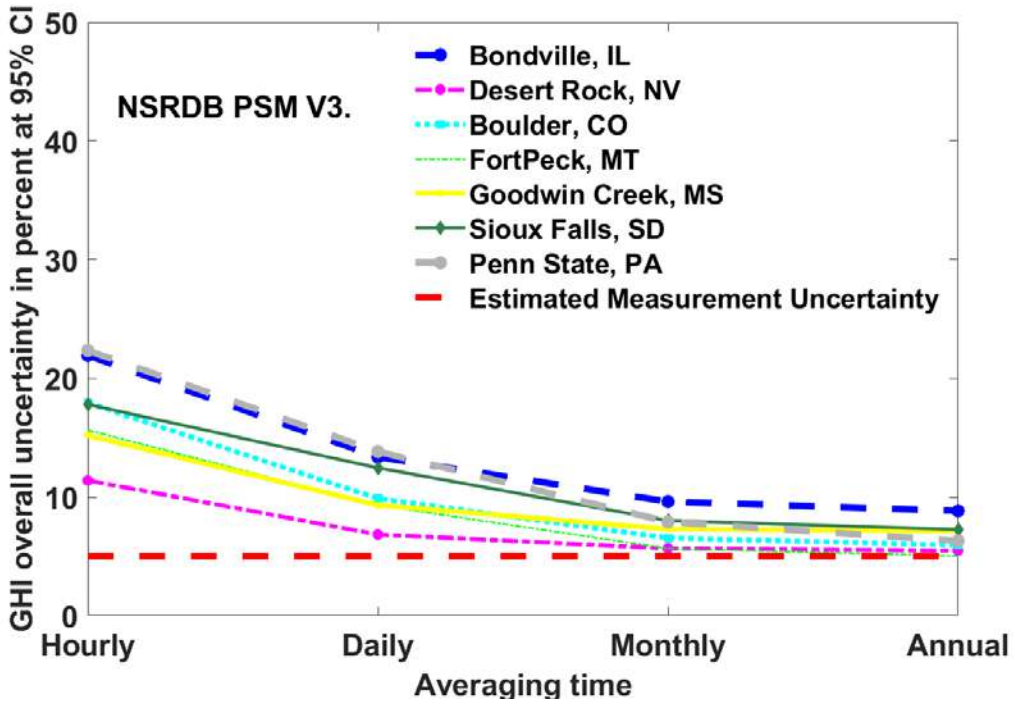

Figure 10-6. Example of decreasing trend of the RMSD of NSRDB-modeled GHI (1998–2018 PSM V3) with averaging time at various U.S. stations in comparison with measurement uncertainty. Y-axis shows the overall uncertainty with 95% confidence interval.

Image by NREL

10.4.2 Practical Examples

To improve modeled data integrity, a comprehensive representation of the model uncertainty method is desirable. As discussed in Section 10.2, no standardized method that would be equivalent to GUM, but specifically addressing modeled estimates, exists yet. Ideally, the assessment of modeled data uncertainty would attempt to replicate the developments made for measurement uncertainty, as detailed in Section 10.1. This means that the individual uncertainty of each of the six sources of error identified above would have to be quantified and ultimately combined in quadrature. This process is still in its infancy, but some considerations are developed in what follows for each source of error.

1. *Intrinsic Model's Uncertainty*

   Each radiation model—most notably of an empirical nature—constitutes an approximate (and thus imperfect) mathematical representation of complex physical processes that occur throughout the atmosphere. To accelerate calculations, most models used in practice rely on several simplifying assumptions, parameterizations, look-up tables, interpolation/extrapolation, ad hoc empirical results, etc., which decrease performance. Because this might occur under various, typically unknown, circumstances, and there is essentially no way to verify the model's output without measurements that carry uncertainty, it is difficult to attribute a specific uncertainty to the model itself. The most sophisticated models with a discrete-ordinate solver algorithm (e.g., libRadtran [Emde et al. 2016; Mayer and Kylling 2005]), are considered to provide an "exact" solution to the equation of radiative transfer, and thus should have no intrinsic uncertainty, although this cannot be demonstrated. Moreover, such models cannot normally be used in repetitive calculations because of their considerable computing requirements. Even when they are used in research projects, their irradiance predictions do not necessarily appear better than those of much simpler models, possibly because of the inherent uncertainty in the reference ground truth or error propagation from imperfect inputs (Abreu et al. 2023).

2. *Model's Input Data Uncertainty*

   Any radiation model relies on various inputs to provide an irradiance prediction. Some of these inputs are either deterministic (e.g., SZA) or can be measured very accurately, and can thus have an extremely low uncertainty if properly handled. Many important variables, however, can only be retrieved (e.g., through remote sensing) or estimated, in which case their uncertainty can be high and dependent on location and time. This is, for example, the case for all aerosol- and cloud-related variables. If the radiation processes strongly depend on such uncertain variables, the irradiance predictions will be impacted by any error in them. Further, if errors in the atmospheric input data are large over a specific area, a paradoxical situation can occur whereby the predictions of a sophisticated model that uses such inputs can be worse than those of a simple locally developed empirical model that does not depend on them (Sun et al. 2022).

3. *Uncertainty Caused by Error Propagation from Input to Output*

   Error propagations depend on the sensitivity of the model to each input, and how these sensitivities interact with each other to result in compensation or amplification of errors.

Ideally, the specific standard deviation, σ_p, for error propagation affecting a model that predicts the irradiance quantity E from n inputs X_i (i = 1, n) would be obtained from:

$$\sigma_p^2 = \left(\Delta X_1 \frac{\delta E}{\delta X_1}\right)^2 + \left(\Delta X_2 \frac{\delta E}{\delta X_2}\right)^2 + \cdots + \left(\Delta X_n \frac{\delta E}{\delta X_n}\right)^2 \quad (10\text{-}17)$$

where the ΔX_i error in input X_i is multiplied by the sensitivity of the modeled output to X_i. This formula, however, might be too pessimistic because it does not take any possible systematic compensation of errors into account. In any case, there is currently no specific data to determine these sensitivities for each model used in practice. Some preliminary studies involving a few clear-sky radiation models have been made, however, either for the particular case of the sensitivity of DNI predictions to aerosol inputs (Gueymard 2003; 2012a; Gueymard and Ruiz-Arias 2015) or for more inputs but with a rigorous radiative transfer model, namely, libRadtran (García et al. 2014). Recently, Wang et al. (2024) analyzed the impact of various key atmospheric variables (including cloud optical depth) on the estimates of hourly all-sky GHI obtained with hybrid combinations of libRadtran and machine-learning algorithms.

4. Uncertainty in Ground-Truth Measurements

As mentioned earlier, the validation of any modeled dataset is normally made by comparison with high-quality ground-based radiometric measurements. Because there is uncertainty involved in those, even a perfect model using perfect inputs would not be attributed an uncertainty lower than that of these reference ground-truth measurements. In the field, the uncertainty of irradiance measurements can be estimated using ASTM G213 (2017), for instance.

Note that many publications mention the term "model uncertainty," but this is a confusing misnomer. Based on the above discussion, a more correct terminology is "modeled prediction uncertainty" because all four sources of error a), b), c), and d) are then explicitly included. It is also important to emphasize that the prediction uncertainty is not necessarily independent from the measurement uncertainty, which complicates the picture, as demonstrated by the following example. Suppose that the predictions from two models, M1 and M2, are compared against GHI measurements obtained with a high-quality pyranometer of assumed 5% uncertainty. Unbeknownst to the analyst, however, that specific instrument is incorrectly calibrated, resulting in a systematic bias of +3% in the measurements. Unbeknownst to the analyst as well, M1 and M2 behave the same in terms of introducing randomness in their outputs, but M1 happens to be perfectly centered (no bias), whereas M2 is biased +3% for the specific inputs used at that specific location. The comparison with ground measurements would lead the analyst to the incorrect conclusion that M2 is better than M1 and that the latter's uncertainty is larger than the former's.

5. Uncertainty Caused by the Interannual Variability

Modeled irradiance predictions are typically validated against ground-based measurements that span a period of only a few months or years. To extrapolate those results to a longer

period, such as the 30-year period of climate normals, additional uncertainty must be added, depending on irradiance component and the specific interannual variability conditions of the area (see Chapter 6). Even though this uncertainty is not negligible in general, most validation results do not take it into account and are thus somewhat optimistic, unfortunately. Note, however, that this uncertainty obviously does not apply to short-term forecasts.

6. Uncertainty Caused by Long-Term Trends

Modeled predictions can only be validated against observational data from the past. Sometimes, the reference measurements that are needed for such validation are years or decades old. Various regions of the world are affected by long-term trends (e.g., dimming and brightening) that typically affect DNI more than GHI. Because what actually matters in solar resource assessments is the accuracy of modeled predictions in future decades, the mismatch between the validation period in the past and the future period of interest must be attributed an additional uncertainty. This specific uncertainty is still not known precisely and has not been considered yet in published validation results. As in the case just above, this uncertainty does not apply to short-term forecasts.

7. Uncertainty Caused by the Sun's Output Variability

As discussed in Chapter 2, the sun's output is not constant but has both short-term (daily) fluctuations and a long-term (≈11-year) cycle. At any moment, the resulting uncertainty is about ±0.2%. It affects the uncertainty of all radiometer calibrations made outdoors because, by chance, an instrument might be calibrated at a moment when the sun's output is exceptionally high or low. For modeled data, however, this source of uncertainty is relatively small and can be neglected.

It is essential to use measurements of solar radiation made at ground stations from regions in various climates (or even microclimates) with the goal of performing a detailed evaluation of the modeled dataset; however, measurements of solar radiation made at ground stations are temporally and spatially sparse, and they are expensive to operate and maintain. Further, to perform an accurate evaluation of the model's predictions, it is critical that these ground-based irradiance measurements be of high quality and rely on low-uncertainty radiometers that follow the best practices for collection, operation, maintenance, and quality assurance.

Studies such as those by Cebecauer et al. (2011b); Gueymard (2014); Habte, Sengupta, and Lopez (2017); Suri and Cebecauer (2014); Thevenard and Pelland (2013) discussed quantification methods aimed at a comprehensive representation of prediction uncertainty. Various error statistics (bias, random error metrics) can be used to evaluate the effective uncertainty of modeled data when also considering the uncertainty in the ground-based irradiance measurements.

Following Gueymard and Wilcox (2011); Habte, Sengupta, and Lopez (2017), the interannual variability metric can be formalized as follows:

$$\mathrm{SD} = \sqrt{\left(\frac{1}{n}\sum_{i=1}^{n}\ (a_i - \hat{a})^2\right)} \tag{10-18}$$

(10-19)

$$U_{inter-annual\ variability}(\%) = \mathrm{COV}(\%) = \frac{\mathrm{SD}}{\hat{a}} \bullet 100$$

where SD is the standard deviation, and a_i is the average irradiance of the individual year, i, of the considered n years. The mean irradiance during the selected long-term period is represented by â.

In parallel, the accuracy of satellite-derived modeled data can be determined using various other statistical indicators, such as the Kolmogorov-Smirnov test (Massey Jr. 1951). The Kolmogorov-Smirnov test is a rigorous nonparametric method that is used for benchmarking satellite-retrieved GHI and DNI against ground-based observations (Espinar et al. 2009; Gueymard 2014). Directly derived from it is the Kolmogorov-Smirnov test integral, which calculates the area differences between two cumulative distribution frequencies to determine the deviation, for example, between satellite-derived data and ground measurements (Espinar et al. 2009; Beyer et al. 2009). Another indicator is OVER (estimate of the area between the CDFs over a critical value distance), which assimilates the original Kolmogorov-Smirnov test; it attempts to find values that are above a specific critical value. Unlike MBD or RMSD, OVER discriminates between values that are either statistically similar or dissimilar (depending on whether they are above a specific critical value). This test has the advantage of being nonparametric and is therefore not distribution-dependent. It compares the two distributions of irradiance to evaluate their resemblance. In the future, more elaborate methods, such as those used in the meteorological community to quantify the performance of weather forecasts (Murphy 1993), can be expected to appear and be adopted more often in large-scale solar resource assessment studies.

10.5 Modeled Data Uncertainty Estimation Challenges

Satellite-derived irradiance datasets have various embedded sources of uncertainty (Cebecauer et al. 2011a; 2011b). Most importantly, irradiance values obtained from satellite-based models use spaceborne observations of clouds. The satellite pixel represents a certain area, typically 1–100 km^2. Depending on that size, some subpixel variability and cloud-induced parallax effects could contribute to higher random errors in both GHI and DNI, as suggested by Cebecauer et al. (2011a); Habte, Sengupta, and Lopez (2017); and Zelenka et al. (1999). In intermittent cloud situations, the resolution of satellite images has limited ability to adequately describe properties of small and scattered clouds. This problem can be exacerbated when a physical retrieval method is used to first characterize the cloud optical properties for a given pixel, which can result in actual partly cloudy periods being classified as cloudless, thus yielding significant positive bias in DNI, for instance (Salazar et al. 2020).

In tropical rainforest climates, it is often challenging to find cloudless situations for characterizing the surface albedo, which is often used as a reference based on which of the pixel's overall cloudiness characteristics can be eventually quantified. Conversely, for geostationary satellite observations at high latitudes, the low satellite viewing angles introduce errors in the detection of cloud position and properties (the satellite sensor most often sees clouds from the side rather than from the top). For intermittent cloud situations, a major part of the observed random errors (evaluated by RMSD) is driven by inadequacies in the cloud-related portions of the radiative transfer algorithms.

Adequate specification of aerosols is another area of concern (Cebecauer et al. 2011b). Aerosols tend to affect DNI three to four times more than GHI, depending on the relative proportions of absorption and scattering for the specific aerosol mixture of the moment and location (Gueymard 2012b). For example, mineral dust is mostly scattering, whereas black carbon is partly absorbing. At any instant, the aerosol optical depth (AOD) varies spectrally, so the common use of a single broadband AOD could result in additional uncertainties (see Chapter 5, Section 5.5 for more information on AOD). When monthly (or "climatological") AOD averages are used, they could introduce significant errors in long-term DNI estimates (Ruiz-Arias et al. 2016). This is more likely to happen over areas of biomass burning, severe urban air pollution, or dust storms, where an aerosol climatology tends to smooth out episodic high-AOD events; therefore, it is advantageous to use AOD data with daily or subdaily resolution in advanced modeling approaches (Cebecauer et al. 2011b; Gueymard et al. 2018).

In regions with variable or complex landscape patterns (e.g., high spatial variability caused by land/water mosaics, complex urbanization, or mountains), the surface reflectance properties change rapidly, both over the space and time domains and even over distances that are shorter than the satellite's spatial resolution (Gueymard et al. 2021) (see Chapter 5, Section 5.11 for more information on this topic). Compared to neighboring rural or natural landscapes, large urban or industrial areas have much higher and temporarily changing concentrations of aerosols and water vapor. Over mountains, rapid changes in elevation also induce rapid changes in the concentration of key atmospheric constituents and in cloud properties. In addition, 3D effects and terrain shading are local complexities that must be considered and approximated by solar radiation models.

Another difficulty inherent to satellite-derived datasets is the poor discrimination between clouds and snow-covered surfaces when using only the visible imagery. This is because both situations have a high reflectance in the visible spectrum; thus, a clear-sky scene over snowy ground might look like an overcast sky, resulting in a strong overestimation or underestimation of both GHI and DNI, depending on the situation (Perez et al. 2002; Vignola and Perez 2004). One such adverse situation is known as the "Eugene syndrome" (Gueymard and Wilcox 2011). The use of multiple channels in the visible and infrared can solve this issue.

Finally, specular reflections of significant intensity, especially from sandy deserts or snowy/icy surfaces during certain times of the day, could result in incorrectly interpreting the satellite image as temporarily cloudy and thus in an underestimation of both GHI and DNI. Theoretically, this issue can be resolved by estimating the probability of specular reflection for such areas and factoring that into the calculation of surface radiation.

10.5.1 Indicative Uncertainty of Modern Satellite-Based Models

As an example, experience based on 189 validation sites shows that state-of the-art semiempirical satellite models can estimate the annual GHI with bias of about ±4% when normalized to daytime irradiation (Suri and Cebecauer 2014). This bias value depends on topography and climate. It can be higher (up to at least ±8%) in: (1) complex tropical regions; (2) areas with high atmospheric pollution, high latitudes, high mountains, or complex terrain; and (3) regions with low sun angles and occurrences of snow. Typical bias for DNI estimates at most sites is approximately twice that of GHI.

Regarding random errors, the main sources of increased uncertainty are clouds and, to a lesser extent, changes in snow cover and increased dynamics of aerosols. Over arid and semiarid areas or during sunny seasons, the RMSD of hourly GHI values normally range from 7–20%. In more cloudy regions with more intricate weather patterns, higher dynamics of atmospheric constituents, complex landscapes, or middle latitudes, the hourly RMSD increases to 15–30%. Over high mountains, high latitudes, or during seasons with low sun angles and frequent occurrences of snow, the relative RMSD for GHI can be 25–35% or more. Similar patterns of RMSD exist for the hourly DNI but with approximately twice the errors mentioned for GHI. In arid and semiarid zones, which are of the highest interest for concentrating solar energy technologies, RMSDs for the hourly DNI ranging from 18–30% are typical. In cloudier regions, with significant dynamics exhibited by aerosols, RMSD can reach 25–45%. Finally, at high latitudes and over mountains, RMSD could exceed 45%.

With continuous progress in satellite sensors and radiation models, it can be expected that the accuracy in satellite-derived databases will continue to improve, as suggested by recent validation results (Babar et al. 2018; 2019; Bright 2019; Kamath and Srinivasan 2020; Shi et al. 2018; Urraca et al. 2018). In Urraca et al. (2017), satellite data are even used to test some aspects of ground measurements using the positive-quality aspects of satellite-based irradiance data. A general validation of this "reverse QC" approach for ground measurements still needs to be undertaken, however.

10.6 Evaluation and Uncertainty of Irradiance and PV Power Forecasts

The evaluation of solar irradiance forecasts provides users with the necessary information about forecast quality and helps them choose from different forecasting products or assess the risk when using a particular forecast as a basis for decisions. This section first addresses the evaluation of deterministic irradiance or solar power forecasts that provides an overall indication of the uncertainty of a specific forecast model. Probabilistic solar forecasts assigning uncertainty estimates to each individual forecast value are described in Chapter 9, Section 9.7. Methods for probabilistic forecast evaluation, including the assessment of reliability, resolution, and sharpness are given below in Section 10.6.3.

As described in Section 10.2, the quality of forecasts, both deterministic and probabilistic, is evaluated by assessing their similarity to reference data. Most often, irradiance measurements are used as reference data. They are commonly referred to as ground-truth data, though they are also affected by a certain degree of uncertainty (see Section 10.3) Alternatively, satellite-retrieved irradiance values or the output of a detailed physical model might serve as reference. The uncertainty of the reference data should always be kept in mind when interpreting the results of forecast evaluations.

An extensive overview of forecast-verification methods is given by Jolliffe and Stephenson (2011). The choice of appropriate metrics and concepts for the evaluation of solar irradiance and power forecasts is the subject of ongoing discussions within the solar forecasting community; see Hoff and Perez (2012); Kleissl et al. (2013). Recently, Yang et al. (2020) proposed applying the well-established Murphy-Winkler framework for distribution-oriented

forecast verification as a standard practice to analyze and compare deterministic solar forecasts. In parallel, Lauret et al. (2019) addressed the evaluation of probabilistic solar forecasts.

In this chapter, the most standard evaluation methods for solar forecasting are outlined. These include: (1) statistical error metrics (Section 10.2.3); (2) basic visual assessment (Section 10.2.4); (3) comparison to reference models using the skill score (Section 10.6.1); (4) analyzing forecasts as a function of different influencing parameters (e.g., location, solar elevation, cloud conditions; Section 10.6.2); and (5) introduction to probabilistic forecast evaluation (Section 10.6.3).

10.6.1 Skill Score

The skill score (also referred to as forecast skill) is used to quantify the forecast performance relative to a reference model. RMSE is normally used for this comparison; other scores, such as MAE or MSE, are also often used. The skill score is defined as the difference between the score of the reference model and the forecast model divided by the difference between the score of the reference model and a perfect model; note that a perfect model yields zero RMSE. For RMSE, the skill score, ss_{RMSE}, is calculated as:

$$ss_{RMSE} = \frac{RMSE_{ref} - RMSE}{RMSE_{ref}}, \tag{10-20}$$

where $RMSE_{ref}$ refers to the reference model, and RMSE refers to the investigated forecasting algorithm (Coimbra and Pedro 2013). The skill score's value of 1 then indicates a perfect forecast, and a skill score of 0 means that the investigated algorithm has the same RMSE as the reference forecast. A negative value indicates performance that is worse than the reference. Skill scores might be applied for comparisons to a simple reference model and also for intercomparisons of different forecasting approaches (i.e., improvement scores).

In solar radiation forecasting, persistence is the simplest and most widely used reference model to evaluate forecast skill. Several definitions of persistence of solar irradiance are given in Chapter 9, Section 9.2.1, including simple persistence, scaled persistence (which accounts for solar geometry changes), and more-advanced concepts, such as smart persistence. Simple persistence, which does not account for solar geometry, is not recommended as a baseline for forecast lead times other than 24 hours (or multiples of it). Scaled or smart persistence are a much better choice in general. Alternatively, if long-term irradiance measurements are available, combinations of climatology and persistence can be used, as recommended by Yang et al. (2020) as an advanced reference model for forecast evaluation.

10.6.2 Analysis of Solar Forecasts Using Statistical Metrics From Different Perspectives

Solar forecasts can be analyzed with statistical metrics calculated over various scales. For example, the statistical metrics described in Section 10.2.3 can be calculated over global, temporal, or spatial scales to assess the performance of solar forecasts. The global-scale metrics computed with all available modeled-observed data across all locations and times are used to evaluate the overall performance of solar forecasts.

It is also useful to group forecast evaluations with respect to forecast lead time (temporal-scale metrics), that is, to compute error metrics with all available data at a given lead time to evaluate

the error evolution as a function of the forecast lead time. Evaluation of forecast performance in dependence of forecast lead time is shown for several examples and a variety of different forecasting algorithms in Chapter 9 (e.g., Figures 9-8, 9-13, 9-19, 9-20). This type of analysis is particularly helpful in identifying the most suitable forecast models for different lead times.

Other important considerations include analyzing forecast quality as a function of space or cloud conditions (Section 10.6.2.1), solar position and time of day (Section 10.6.2.2), or cloud variability and spatiotemporal averaging (Section 10.6.2.3).

10.6.2.1 Spatial Evaluation of Forecasts and Taylor Diagrams for Different Cloud Conditions

For the evaluation of solar forecasts as a function of space, the statistical metrics are calculated with all available pairs of predictions and observations at each location. Figure 10-7 shows an example of 2D maps for R^2, RMSE, MAE, and MBE for the DNI forecasts simulated by the combination of the Fast All-sky Radiation Model for Solar applications with DNI (FARMS-DNI) (Xie et al. 2022; Yang et al. 2022) and WRF-Solar (Jimenez et al. 2016). The NSRDB (Sengupta et al. 2018) was used to analyze the spatial distribution of the statistical metrics. More generally, high-quality satellite-derived solar radiation datasets essentially offer the opportunity to conduct in-depth analyses of the accuracy of gridded solar forecasts over a wide range of regions.

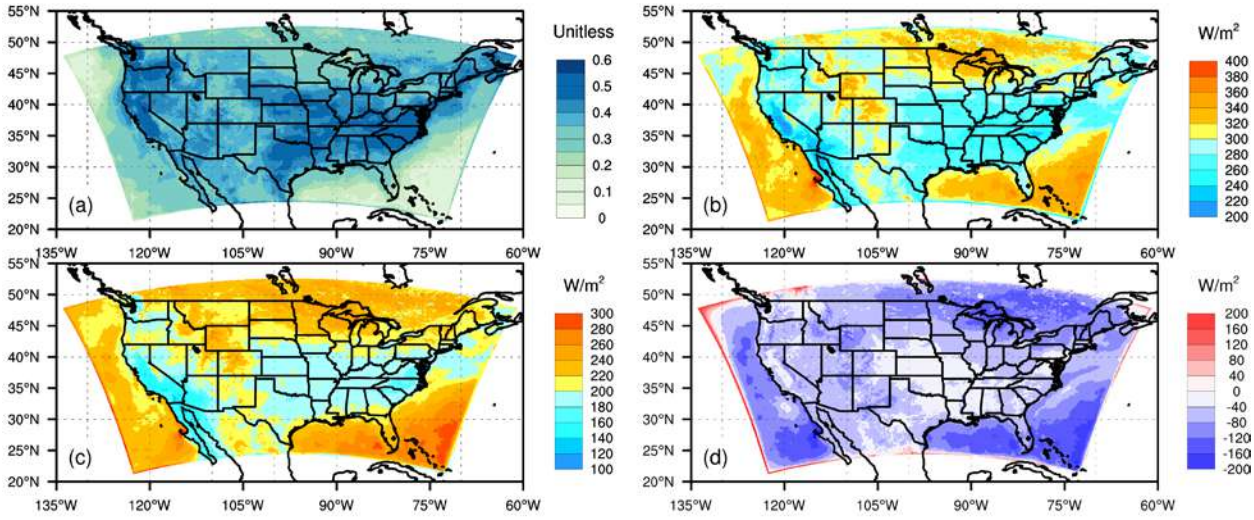

Figure 10-7. 2D maps for (a) R^2, (b) RMSE, (c) MAE, and (d) MBE of DNI forecasts simulated by the combination of FARMS-DNI and WRF-Solar

Statistical metrics of DNI forecasts are calculated against the NSRDB data for each grid point. The evaluation is performed using 365 sets of day-ahead forecasts spanning 2018.

Image by NREL

A statistical summary of model performance for the prediction of solar irradiance can be obtained with the Taylor diagram (Taylor 2001). This diagram quantifies the performance of forecasts (or other modeled data using three statistical metrics computed from modeled-observed pairs: the Pearson correlation coefficient, the standard deviation, and RMSE. Figure 10-8 shows an example of a Taylor diagram (using a normalized standard deviation) representing the performance of each member of ensemble GHI forecasts composed of 20 members simulated

with the WRF-Solar ensemble prediction system (EPS). The member obtaining the best performance is the one that lies closest to the reference point compared to other models. For example, in Figure 10-8, WRF-Solar accurately represents clear-sky scenes (red dots), thus the ensemble members lie much closer to the reference point than for cloudy-sky or all-sky conditions.

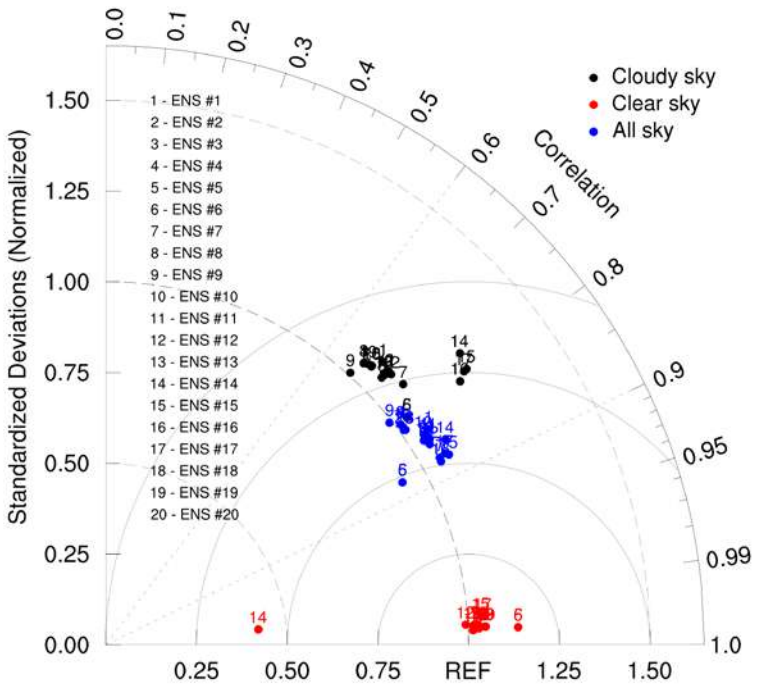

Figure 10-8. Taylor diagram for 20 ensemble members simulated with WRF-Solar under three sky conditions

Data from WRF-Solar EPS development (Sengupta et al. 2022).

10.6.2.2 Analysis of Forecast Error with Respect to Solar Elevation

Solar irradiance has a deterministic component, which results from the daily and seasonal course of the sun, and a nondeterministic component because of, for example, clouds. Both the deterministic and nondeterministic signals influence the forecast error signal. To investigate the solar irradiance forecast errors, valuable additional information is obtained by evaluating not only GHI (or DNI) but also the nondeterministic part of solar radiation, which is primarily caused by errors in the representation of clouds. To this aim, the analyzed variable is often selected to be the forecast error based on the clear-sky index rather than based on GHI.

The forecast performance of the clear-sky index can be illustrated by examples from an observational dataset of hourly pyranometer measurements from 18 weather stations of the German Weather Service (DWD) from March 2013 to February 2014 (Lorenz et al. 2016) and forecasts from two NWP models:

- High-resolution deterministic global Integrated Forecasting System (IFS) model, operated at the ECMWF with a spatial resolution of 0.125° and 3-hourly outputs; here, forecast horizons up to 24 hours are used, issued every day at 00:00 UTC.

- High-resolution Limited Area Model, operated for the area of Scandinavia (HIRLAM-SKA), operated at the Danish Meteorological Institute, with a spatial resolution of 3 km, hourly outputs, and forecast horizons from 4–9 hours ahead, issued daily at 00:00, 06:00, 12:00, and 18:00 UTC.

Figure 10-9 shows the RMSE and MBE of the clear-sky index, K_c, as a function of the cosine of the solar zenith angle (Figure 10-9, left) and the time of day (Figure 10-9, right) for the two different NWP model forecasts (IFS and SKA). The two models show similar behavior: RMSE increases with low SZA or, equivalently, during morning and evening hours, as is also the case with the magnitude of bias. This error pattern is very often caused by deficient modeling of the atmospheric transport of radiation for low solar altitudes. This limitation is a well-known flaw of the two-stream schemes used in most NWP models. Other model limitations also exist, such as 3D effects and atmospheric refraction issues whose impact is enhanced at low solar altitudes.

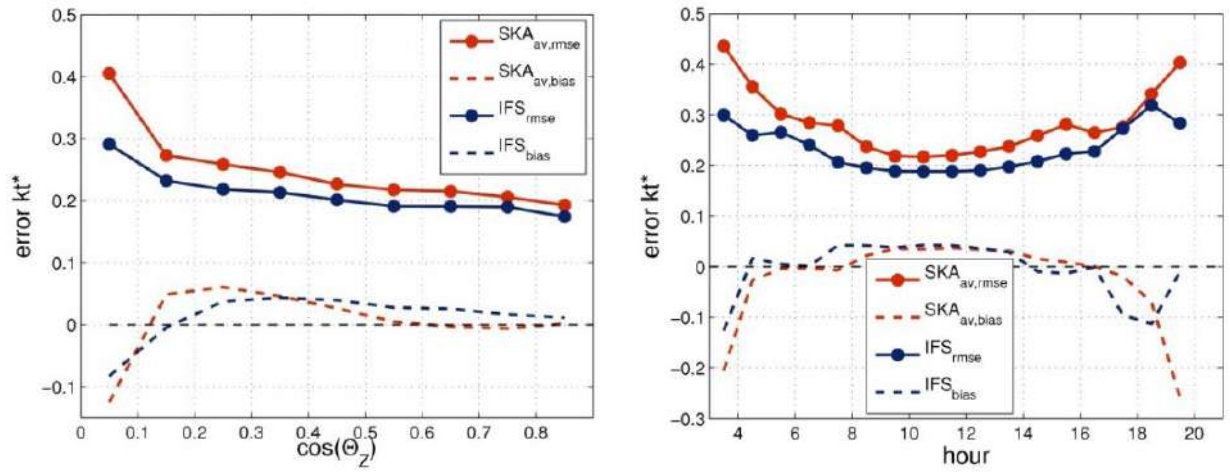

Figure 10-9. Clear-sky index (here noted as kt*) forecast error as a function of (left) cosine of SZA (noted here θ_z) and (right) hour of the day for the forecasts issued by the IFS and SKA NWP models (blue and red lines, respectively)

Solid lines show RMSE, and dashed lines show MBE (bias). The evaluated period is from March 1, 2013 to February 28, 2014.

Image by Elke Lorenz

10.6.2.3 Analysis of Forecast Error with Respect to Cloud Variability and Spatiotemporal Averaging

Forecasts generally show good agreement with measurements during clear-sky periods or even completely overcast days because both basically have a quasi-constant clear-sky index. In contrast, cloud variability strongly impacts solar forecasting accuracy. Thus, considerable deviations from the measurements are typically observed during days with variable cloudiness. An evaluation of the SKA forecast errors as a function of the measurement-derived K_c variability, here represented by the standard deviation of K_c over a 5-hour period, is shown in Figure 10-10. The evaluation also shows this dependence for multiple spatial and temporal averaging configurations of the SKA forecasts.

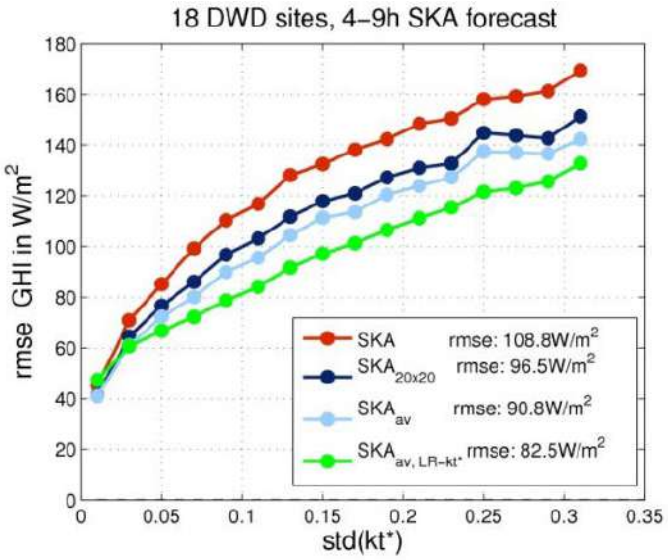

Figure 10-10. RMSE of various versions of the SKA forecasts as a function of the standard deviation of measurement-based clear-sky index, std(kt*)

Red line: SKA, original output. Dark blue: Nearest grid point, SKA_{20x20} averaged throughout 20 by 20 grid points. Light blue: SKA_{av} 5-hour sliding mean of the clear-sky index of the forecasts of the average throughout 20 by 20 grid points. Green: $SKA_{av,\ LR.kt^*}$: linear regression of the clear-sky index of the forecasts applied to SKA_{av}. The evaluated period is from April 3, 2013–February 28, 2014. Training set: Last 30 days, all 18 DWD sites.

Image by Elke Lorenz

Overall, Figure 10-10 shows that:

1. The forecast error increases with enhanced cloud variability.
2. Spatial and temporal forecast averages result in reduced RMSE values, going from negligible reductions under very stable conditions to large reductions under highly variable conditions.

Regarding the first point, the solar radiation forecast error shows a clear dependency with respect to cloud variability and, more generally, with respect to cloud conditions (Figure 10-8). Combining the error trend in the dependence of cloud conditions and the solar elevation has been proposed as an efficient method to reduce the systematic error in NWP model forecasts using a postprocessing model output statistic (MOS). In particular, Lorenz et al. (2009) used a polynomial function with cos(SZA) and K_c as independent variables to parameterize the forecast bias error from historical forecasts relative to observations and ultimately to subtract the parameterized error from operational forecasts. This approach has also been adapted and evaluated for other NWP models and different climates (Mathiesen and Kleissl 2011; Müller and Remund 2010; Pelland et al. 2013).

Regarding the second point, high-resolution irradiance forecasts frequently show phase shifts when compared to measurements, in particular for variable cloud conditions, as described also in Lorenz et al. (2016):

> "Phase shifts are caused by displacement errors in cloud prediction. Even small errors in cloud position can result in large errors for high-resolution forecasts resolving also small-scale cloud features—which is often referred to as 'double

penalty' effect. Spatial as well as temporal averaging reduces large fluctuations and forecast errors in variable situations and consequently also the RMSE."

Looking at this from another perspective, spatial and temporal averaging is in effect a way to create a new forecast by averaging forecasts of neighboring points in space and time. Under variable cloud conditions, the correlations among these forecasts are small, leading to random error cancellations during the averaging process. In contrast, under stable conditions, the correlation among neighboring pixels is high and the cancellation of random errors is reduced.

When using averaging for RMSE reduction, the optimal area size and time interval depend on the correlation structure among neighboring forecasts, both in time and space. Multiple studies have been conducted on this topic. For instance, a detailed evaluation of irradiance forecasts from the Canadian GEM model resulted in a reduction of forecast errors in the range from 10% to 15% when the model outputs were averaged throughout several hundred kilometers (Pelland et al. 2013). A similar improvement was achieved with WRF forecasts provided by Meteotest using averages over an area of 50 km by 50 km (Müller and Remund 2010). In parallel, Mathiesen and Kleissl (2011) reported an averaging area of 100 km by 100 km as suitable for irradiance forecasts using either the GFS or North American Mesoscale forecast system models. The benefit of horizon-dependent smoothing filters for Cloud Motion Vectors (CMV) forecasts was also shown by Lorenz et al. (2004), Aicardi et al. (2022), and Kühnert et al. (2013)

It is emphasized here that spatial and temporal averaging effects also have a strong impact on RSME when comparing solar irradiance forecasts of NWP models with different output resolutions. This should be considered in model intercomparisons, where different models can be compared on a similar spatial and temporal scale in addition to their original output resolution.

Temporal and spatial averaging can be also considered for nowcasts based on an all-sky imager (ASI). It has been found that in a nowcasting system with four ASIs during days with many transient clouds, the DNI RMSE for forecasts that are 10 minutes ahead is reduced in half, from 13.0% to 6.5%, by using averages of 4 km^2 and 15 minutes with respect to pixel-wise forecasts (Kuhn et al. 2018)

Despite the positive impact of spatiotemporal averaging on reducing the RMSE of a forecast, there are also negative effects. A first negative impact exists on the frequency distribution of forecasted data because the averaging process reduces extreme forecasted values and distorts the original frequency distribution of the forecast data. A second impact is that the capability to reproduce irradiance variability by the forecasts is obviously reduced. These different implications of averaging should be considered when evaluating and selecting a forecasting system for a given application. Whereas, for energy trading, RMSE or MAE are the most critical error metrics, ramp forecasting requires forecasts reflecting the high-resolution irradiance variability.

10.6.3 Verification of Probabilistic Solar Forecasts

10.6.3.1 Properties Required for a Skillful Probabilistic System

Several attributes characterize the quality[62] of probabilistic forecasts. Here, the focus is on reliability, resolution, and sharpness—the main properties used to assess the quality of probabilistic forecasts.

Reliability or calibration refers to the statistical consistency between forecasts and observations; in other words, a forecast system has a high reliability if the forecast probability and observed frequency agree. The reliability property is an important prerequisite because nonreliable forecasts would lead to a systematic bias in subsequent decision-making processes (Pinson et al. 2007).

Resolution measures the ability of a forecasting model to generate predictive distributions that depend on forecast conditions. Put differently, the more distinct the observed frequency distributions for various forecast situations are from the full climatological distribution, the more resolution the forecast system has. Climatological forecasts are perfectly reliable but have no resolution. Consequently, a skillful probabilistic forecasting system should issue reliable forecasts and should exhibit high resolution.

Sharpness refers to the concentration of predictive distributions and can be measured by the average width of the prediction intervals. Unlike reliability or resolution, sharpness is a function of only the forecasts and does not depend on the observations. Consequently, a forecasting system can produce sharp forecasts yet be useless if the probabilistic forecasts are unreliable.

10.6.3.2 Probabilistic Verification Tools

A number of visual diagnostic tools and error metrics are used for verifying probabilistic forecasts. Table 10-3 lists the diagnostic tools used to analyze probabilistic forecasts, for which Lauret et al. (2019) provided pros and cons, as well as detailed information about their implementation. Note that some tools were initially designed for a specific type of forecast (i.e., an ensemble or quantile forecast).

[62] Quality refers to the correspondence between forecasts and observations.

Table 10-3. Visual Diagnostic Tools for Probabilistic Forecasts

Diagnostic Tool	Remarks
Reliability Diagram	Initially designed for the reliability assessment of quantile forecasts. Can be used for ensemble forecasts if members are assigned specific probability levels; see Lauret et al. (2019).
Rank Histogram	Initially designed for the reliability assessment of ensemble forecasts. Can be extended to quantile forecasts if quantiles are evenly spaced.
Probability Integral Transform Histogram	Represents a reliability assessment of quantile forecasts
Sharpness Diagram	Plots the average width of the prediction intervals for different nominal coverage rates. Sharpness can only contribute to a qualitative evaluation of the probabilistic forecasts. Even if narrow prediction intervals are preferred, sharpness cannot be seen as a property for verifying the quality of probabilistic forecasts but is more likely the consequence of a high resolution.

Numerical scores provide summary measures for the evaluation of the quality of probabilistic forecasts. Table 10-4 enumerates the main scoring rules for evaluating the quality of probabilistic forecasts of a continuous variable. All the scores listed in the table are proper scoring rules (Gneiting and Raftery 2007), ensuring that perfect forecasts are given the best score value. Lauret et al. (2019) gives a detailed definition of each score.

Table 10-4. Forecast Metrics for Probabilistic Forecasts

Forecast Metric	Remarks
Continuous Ranked Probability Score (CRPS)	Can be normalized to define a skill score (CRPS skill score). Can be further partitioned into the two main attributes: reliability and resolution.
Ignorance Score	Local score (i.e., the score depends only on the value of the predictive distribution at the observation). Cannot be normalized.
Interval Score	Specifically designed for interval forecasts.
Quantile Score	Forecast performance of specific quantiles.

Some frequently used diagnostic tools and numerical scores to evaluate probabilistic forecasts are detailed next (see Lauret et al. [2019] and Yang et al. [2020] for descriptions of other metrics).

10.6.3.2.1 Reliability Diagram

A reliability diagram is a graphical verification display used to assess the reliability attribute of quantile forecasts. Quantile forecasts are evaluated one by one, and their observed frequencies are reported versus their forecast probabilities (Figure 10-11). Such a representation is appealing because the deviations from perfect reliability (the diagonal) can be visually assessed (Pinson et al. 2010); however, because of both the finite number of pairs of observation/forecast and also possible serial correlation in the sequence of forecast-verification pairs, observed proportions are not expected to lie exactly along the diagonal, even if the density forecasts are perfectly reliable. Pinson et al. (2010) proposed a method to add consistency bars to the reliability diagram. This addition can help users gain more confidence in their (possibly subjective) judgment regarding the reliability of the different models. Figure 10-12 shows an example of reliability diagram with consistency bars. In this example, the forecasts cannot be considered reliable because the line Figure 10-11 corresponding to the forecasts falls outside the consistency bars. More elaborate reliability diagrams are proposed by Yang (2019a; 2019b).

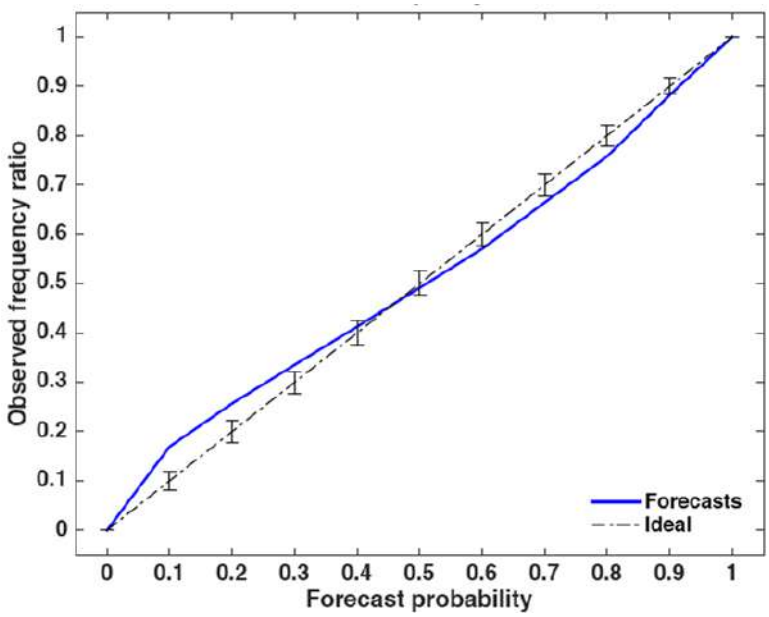

Figure 10-11. Example of a reliability diagram

Consistency bars for a 90% confidence level around the ideal line are individually computed for each nominal forecast probability. Image by University of La Réunion Laboratory of Physics and Mathematical Engineering for ENERGY, the ENVIRONMENT and BUILDINGS (PIMENT)

10.6.3.2.2 Rank Histogram

A rank histogram is a graphical display initially designed for assessing the reliability of ensemble forecasts (Wilks 2011). Rank histograms help users to visually assess the statistical consistency of the ensemble—that is, if the observation can be seen statistically like another member of the ensemble (Wilks 2011). A flat rank histogram is a necessary condition for ensemble consistency and shows an appropriate degree of dispersion of the ensemble. Underdispersed or overdispersed ensembles lead to U-shaped or hump-shaped rank histograms, respectively (Figure 10-12).

In addition, some unconditional biases can be revealed by asymmetrical (triangle-shaped) rank histograms. It must be stressed that one should be cautious when analyzing rank histograms. As

shown by Hamill (2001), a perfectly flat rank histogram does not mean that the corresponding forecast is reliable. Further, when the number of observations is limited, consistency bars can also be calculated with the procedure proposed by Bröcker and Smith (2007).

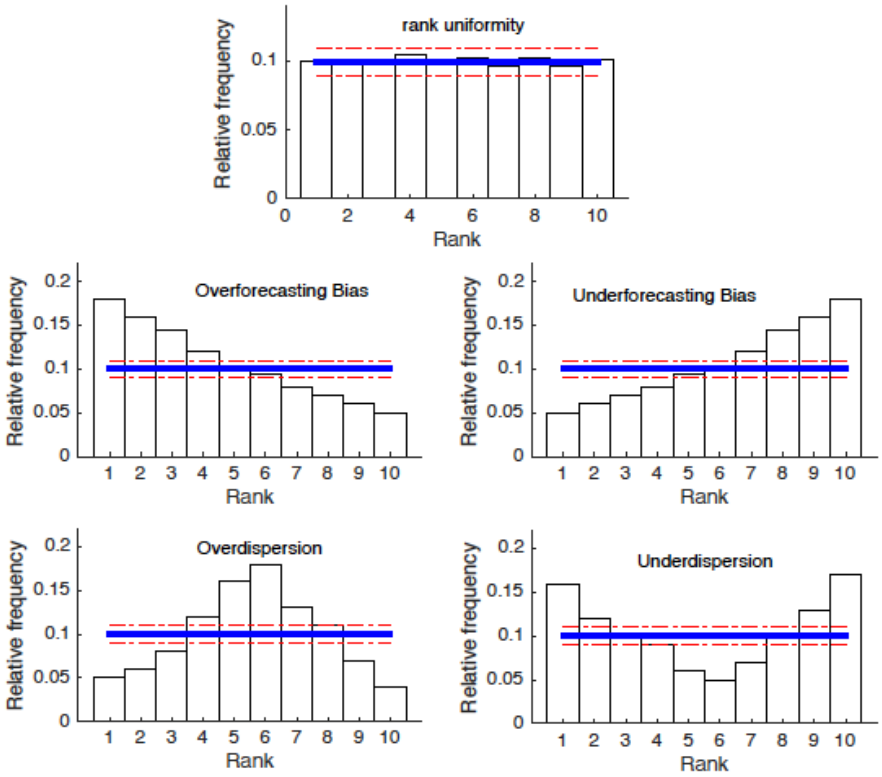

Figure 10-12. Illustrative examples of rank histograms for an ensemble of *M* = 9 members

The horizontal solid blue line denotes the statistical consistency of the ensemble. The dashed-dotted lines represent the consistency bars.

Figure modified from Wilks (2011)

10.6.3.2.3 Overall Skill Assessment With the Continuous Ranked Probability Score

The most common skill score for evaluating the quality of predictive densities of continuous variables is the CRPS, whose formulation is:

$$\mathrm{CRPS} = \frac{1}{N}\sum_{i=1}^{N}\int_{-\infty}^{+\infty}\left[\hat{F}^{i}_{fcst}(y) - F^{i}_{yobs}(y)\right]^{2} dy, \tag{10-21}$$

where $\hat{F}^{i}_{fcst}(y)$ is the predictive cumulative distribution function (CDF) of the variable of interest, x (e.g., GHI), and $F^{i}_{yobs}(y)$ is a CDF of the observation (i.e., a step function that jumps from 0 to 1 at the point where the forecast variable, y, equals the observation, y_{obs}). The squared difference between the two CDFs is averaged over the N forecast/observation pairs. Note that CRPS is negatively oriented (smaller values are better) and has the same dimension as the forecasted variable.

Figure 10-13(a) shows three hypothetical predictive probability density functions (PDFs), and Figure 10-13(b) plots the corresponding predictive CDFs. The black thick line in Figure 10-13(b) represents the CDF of the observation, $F^{i}_{yobs}(y)$. Because CRPS represents the integrated squared difference between the two CDFs, the pair of observation/forecasts labeled "1" will be assigned the best score. Conversely, forecasts indicated by labels 2 and 3 will lead to a higher CRPS. Indeed, although it has the same degree of sharpness as Forecast 1, Forecast 2 is not centered on the observation (i.e., this is a biased forecast). Conversely, Forecast 3 is centered on the observation, but is less sharp than Forecasts 1 or 2. In summary, CRPS rewards the concentration of probability around the step function located at the observed value (Hersbach 2000).

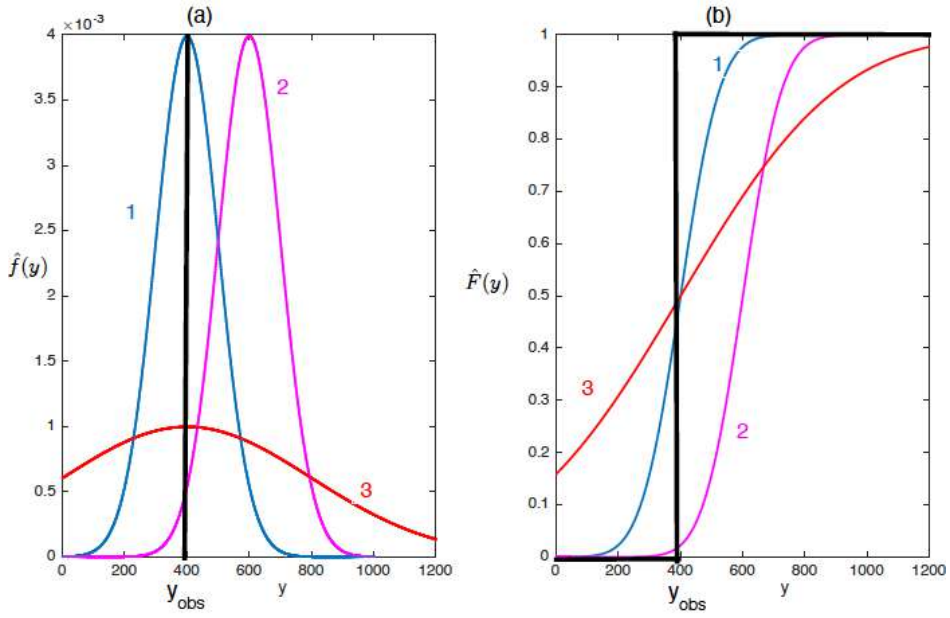

Figure 10-13. Schematic of the CRPS skill score

Three forecast PDFs are shown in relation to the observed variable in (a). The corresponding CDFs are shown in (b), together with the step function CDF of the observation (black heavy line). Forecast 1's PDF would produce a small (i.e., good) CRPS. This would not be the case for Forecast 2 or Forecast 3.

Illustration modified from Wilks (2011)

CRPS can be further partitioned into the two main attributes of probabilistic forecasts described above: reliability and resolution. The decomposition of the CRPS leads to:

$$\text{CRPS} = \text{RELIABILITY} - \text{RESOLUTION} + \text{UNCERTAINTY}. \tag{10-22}$$

The uncertainty term cannot be modified by the forecast system and depends only on the observation's variability (Wilks 2011). Because CRPS is negatively oriented, the goal of a forecast system is to minimize the reliability term and maximize the resolution term as much as possible. Hersbach (2000) and Lauret et al. (2019) detail the procedures for calculating the different terms (reliability and resolution, respectively) for ensemble and quantile forecasts.

It must be stressed that the decomposition of CRPS provides quantitative overall measures of reliability and resolution, providing additional and valuable insight into the performance of a forecasting system.

Similarly, to obtain skill scores used for evaluating deterministic forecasts (Coimbra and Pedro 2013), a CRPS skill score (CRPSS) can be derived to quantify the improvement brought by a new method over a reference easy-to-implement model, such as:

$$\mathrm{CRPSS} = 1 - \frac{\mathrm{CRPS}_{new\ method}}{\mathrm{CRPS}_{Reference}}. \tag{10-23}$$

Negative values of CRPSS indicate that the new proposed method fails to outperform the reference baseline model, and, conversely, positive values of CRPSS mean that the new method outperforms the reference model. Further, the higher the CRPSS, the better the improvement. Note that the uncertainty part of the decomposition of CRPS (which corresponds to the score of the climatology) can be used as a reference baseline model. CRPSS and mean-normalized CRPS are also discussed by Yang (2020).

10.6.3.2.4 Interval Score

The interval score (IS) specifically assesses the quality of interval forecasts. As shown by Eq. 10-24, the interval score rewards narrow prediction intervals but penalizes (with a penalty term that increases with increasing nominal coverage rate) the forecasts for which the observation, x_{obs}, is outside the interval. For a $(1-\alpha)\times 100\%$ nominal coverage rate, the interval score is obtained as:

$$\mathrm{IS}_{\alpha} = \frac{1}{N}\sum_{i=1}^{N}\left(U^{i}-L^{i}\right) + \frac{2}{\alpha}\left(L^{i}-x_{obs}^{i}\right)I_{x_{obs}^{i}<L^{i}} + \frac{2}{\alpha}\left(x_{obs}^{i}-U^{i}\right)I_{x_{obs}^{i}>U^{i}}, \tag{10-24}$$

where I_u is the indicator function (I_u = 1 if U is true and 0 otherwise), and U^i and L^i represent the upper $\left(\tau = 1-\frac{\alpha}{2}\right)$ and lower $\left(\tau = \frac{\alpha}{2}\right)$ quantiles, respectively.

A plot of interval scores for different nominal coverage rates might offer a consistent evaluation of the quality of interval forecasts. Consequently, such a plot could advantageously replace the sharpness diagram.

10.7 Available Diagnostic Tools

To evaluate various datasets, many statistical metrics (e.g., correlation, RMSE, MAE, or MBE) can be calculated by common scripts, given their simple formulas. Moreover, some programming languages provide various user-friendly library functions to calculate the statistical metrics (Table 10-5).

Table 10-5. Functions for Statistical Metrics in R, Python, and MATLAB

Programming Language	Remarks
R	Name of library function: Metrics Documentation: https://cran.r-project.org/web/packages/Metrics/Metrics.pdf Documentation for the Taylor diagram: https://search.r-project.org/CRAN/refmans/plotrix/html/taylor.diagram.html.
Python	Name of library functions: NumPy and scikit-learn Documentation: NumPy: https://numpy.org/doc/stable/ scikit-learn: https://scikit-learn.org/0.21/documentation.html Documentation for Taylor diagrams: https://metplotpy.readthedocs.io/en/develop/Users_Guide/taylor_diagram.html.
MATLAB	Link: https://www.mathworks.com/help/matlab/referencelist.html?type=function&category=descriptive-statistics&s_tid=CRUX_topnav Documentation for the Taylor diagram: https://www.mathworks.com/matlabcentral/fileexchange/20559-taylor-diagram.

References

Abreu, E. F. M., C.A. Gueymard, P. Canhoto, and M.J. Costa. 2023. "Performance assessment of clear-sky solar irradiance predictions using state-of-the-art radiation models and input atmospheric data from reanalysis or ground measurements." *Solar Energy* 252, 309–321. https://doi.org/10.1016/j.solener.2023.01.051.

Aicardi, D., P. Musé, and R. Alonso-Suárez. 2022. "A comparison of satellite cloud motion vectors techniques to forecast intraday hourly solar global horizontal irradiation." *Solar Energy* 233, 46–60. https://doi.org/10.1016/j.solener.2021.12.066.

Al-Rasheedi, M., C.A. Gueymard, A. Ismail, and T. Hussain. 2018. "Comparison of two sensor technologies for solar irradiance measurement in a desert environment." *Solar Energy* 161, 194–206. https://doi.org/10.1016/j.solener.2017.12.058.

ASTM E824. 2018. *ASTM E824-10R18E1 Test Method for Transfer of Calibration From Reference to Field Radiometers*. ASTM International. https://doi.org/10.1520/E0824-10R18E01.

ASTM G167. 2023. *ASTM G167-15 Test Method for Calibration of a Pyranometer Using a Pyrheliometer*. ASTM International. https://doi.org/10.1520/G0167-15.

ASTM G213. 2017. *ASTM G213-17(2023) Standard Guide for Evaluating Uncertainty in Calibration and Field Measurements of Broadband Irradiance with Pyranometers and Pyrheliometers*. ASTM International. https://doi.org/10.1520/G0213-17R23.

Babar, B., R. Graversen, and T. Boström. 2018. "Evaluating CM-SAF solar radiation CLARA-A1 and CLARA-A2 datasets in Scandinavia." *Solar Energy* 170, 76–85. https://doi.org/10.1016/j.solener.2018.05.009.

Babar, B., R. Graversen, and T. Boström. 2019. "Solar radiation estimation at high latitudes: Assessment of the CMSAF databases, ASR and ERA5." *Solar Energy* 182, 397–411. https://doi.org/10.1016/j.solener.2019.02.058.

Badosa, J., J. Wood, P. Blanc, C.N. Long, L. Vuilleumier, D. Demengel, and M. Haeffelin. 2014. "Solar irradiances measured using SPN1 radiometers: Uncertainties and clues for development." *Atmospheric Measurement Techniques* 7(12), 4267–4283. https://doi.org/10.5194/amt-7-4267-2014.

Balenzategui, J. L., J. de Lucas, J. Cuenca, A. González-Leiton, M. Molero, F. Fabero, J.P. Silva, E. Mejuto, R. Muñoz, A. Arce, and E. Prieto. 2022a. "Characterization of absolute cavity radiometers for traceability to SI of solar irradiance." *Measurement Science and Technology* 33(11), 115009. https://doi.org/10.1088/1361-6501/ac849d.

Balenzategui, J. L., M. Molero, J.P. Silva, F. Fabero, J. Cuenca, E. Mejuto, and J. De Lucas. 2022b. "Uncertainty in the Calibration Transfer of Solar Irradiance Scale: From Absolute Cavity Radiometers to Standard Pyrheliometers." *Solar* 2(2), 158–185. https://doi.org/10.3390/solar2020010.

Beyer, H. G., J. Polo Martinez, M. Suri, J.L. Torres, E. Lorenz, S. Müller, C. Hoyer-Klick, and P. Ineichen. 2009. *Report on Benchmarking of Radiation Products* (D1.1.3, Contract No. 038665). https://www.academia.edu/29559582/Report_on_Benchmarking_of_Radiation_Products.

Bright, J. M. 2019. "Solcast: Validation of a satellite-derived solar irradiance dataset." *Solar Energy* 189, 435–449. https://doi.org/10.1016/j.solener.2019.07.086.

Bröcker, J., and L.A. Smith. 2007. "Increasing the Reliability of Reliability Diagrams." *Weather and Forecasting* 22(3), 651–661. https://doi.org/10.1175/WAF993.1.

Cebecauer, T., R. Perez, and M. Šúri. 2011a. *Comparing Performance of Solargis and SUNY Satellite Models Using Monthly and Daily Aerosol Data*. Solar World Congress, Kassel, Germany. https://doi.org/10.18086/swc.2011.24.05.

Cebecauer, T., M. Šúri, and C.A. Gueymard. 2011b. "Uncertainty sources in satellite-derived direct normal irradiance: How can prediction accuracy be improved globally." *Proceedings of the SolarPACES 2011 Conference*, 20–23.

Coimbra, C. F. M., and H.T.C. Pedro. 2013. "Stochastic-Learning Methods." In J. Kleissl (Ed.) *Solar Energy Forecasting and Resource Assessment*. Elsevier. https://doi.org/10.1016/B978-0-12-397177-7.00015-2.

Darnell, W. L., W.F. Staylor, S.K. Gupta, and F.M. Denn. 1988. "Estimation of Surface Insolation Using Sun-Synchronous Satellite Data." *Journal of Climate* 1(8), 820–835. https://doi.org/10.1175/1520-0442(1988)001<0820:EOSIUS>2.0.CO;2.

Darnell, W. L., W.F. Staylor, S.K. Gupta, N.A. Ritchey, and A.C. Wilber. 1992. "Seasonal variation of surface radiation budget derived from International Satellite Cloud Climatology Project C1 data." *Journal of Geophysical Research: Atmospheres* 97(D14), 15741–15760. https://doi.org/10.1029/92JD00675.

Emde, C., R. Buras-Schnell, A. Kylling, B. Mayer, J. Gasteiger, U. Hamann, J. Kylling, B. Richter, C. Pause, T. Dowling, and L. Bugliaro. 2016. "The libRadtran software package for radiative transfer calculations (version 2.0.1)." *Geoscientific Model Development* 9(5), 1647–1672. https://doi.org/10.5194/gmd-9-1647-2016.

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

Forstinger, A., S. Wilbert, A.R. Jensen, B. Kraas, C. Fernandez-Peruchena, C.A. Gueymard, D. Ronzio, D. Yang, E. Collino, J. Polo, et al. 2023. *Worldwide Benchmark of Modelled Solar Irradiance Data*. Report IEA-PVPS T16-05: 2023. IEA PVPS. https://iea-pvps.org/key-topics/worldwide-benchmark-of-modelled-solar-irradiance-data/.

García, R. D., O.E. García, E. Cuevas, V.E. Cachorro, P.M. Romero-Campos, R. Ramos, and A.M. de Frutos. 2014. "Solar radiation measurements compared to simulations at the BSRN Izaña station. Mineral dust radiative forcing and efficiency study." *Journal of Geophysical Research: Atmospheres* 119(1), 179–194. https://doi.org/10.1002/2013JD020301.

Gneiting, T. 2011. "Making and Evaluating Point Forecasts." *Journal of the American Statistical Association* 106(494), 746–762. https://doi.org/10.1198/jasa.2011.r10138.

Gneiting, T., and A.E. Raftery. 2007. "Strictly Proper Scoring Rules, Prediction, and Estimation." *Journal of the American Statistical Association* 102(477), 359–378. https://doi.org/10.1198/016214506000001437.

Gueymard, C. A. 2003. "Direct solar transmittance and irradiance predictions with broadband models. Part II: validation with high-quality measurements." *Solar Energy* 74(5), 381–395. https://doi.org/10.1016/S0038-092X(03)00196-8.

Gueymard, C. A. 2012a. "Temporal variability in direct and global irradiance at various time scales as affected by aerosols." *Solar Energy* 86(12), 3544–3553. https://doi.org/10.1016/j.solener.2012.01.013.

Gueymard, C. A. 2012b. "Visibility, aerosol conditions, and irradiance attenuation close to the ground—Comments on "Solar radiation attenuation in solar tower plants" by J. Ballestrin and A. Marzo." *Solar Energy* 86(5), 1667–1668. https://doi.org/10.1016/j.solener.2011.12.027.

Gueymard, C. A. 2014. "A review of validation methodologies and statistical performance indicators for modeled solar radiation data: Towards a better bankability of solar projects." *Renewable and Sustainable Energy Reviews* 39, 1024–1034.

Gueymard, C. A., A. Habte, M. Sengupta. 2018. "Reducing Uncertainties in Large-Scale Solar Resource Data: The Impact of Aerosols." *IEEE Journal of Photovoltaics* 8(6), 1732–1737. https://doi.org/10.1109/jphotov.2018.2869554.

Gueymard, C. A., V. Lara-Fanego, M. Sengupta, and A. Habte. 2021. "Surface albedo spatial variability in North America: Gridded data vs. local measurements." *Solar Energy* 227, 655–673. https://doi.org/10.1016/j.solener.2021.05.012.

Gueymard, C. A., and D.R. Myers. 2008. "Solar Radiation Measurement: Progress in Radiometry for Improved Modeling." In V. Badescu (Ed.), *Modeling Solar Radiation at the Earth's Surface: Recent Advances*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77455-6_1.

Gueymard, C. A., and D.R. Myers. 2009. "Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling." *Solar Energy* 83(2), 171–185. https://doi.org/10.1016/j.solener.2008.07.015.

Gueymard, C. A., and J.A. Ruiz-Arias. 2015. "Validation of direct normal irradiance predictions under arid conditions: A review of radiative models and their turbidity-dependent performance."

Renewable and Sustainable Energy Reviews 45, 379–396. https://doi.org/10.1016/j.rser.2015.01.065.

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

Habte, A., M. Sengupta, A. Andreas, I. Reda, and J. Robinson. 2016. *The Impact of Indoor and Outdoor Radiometer Calibration on Solar Measurements*. NREL/CP-5D00-66659, National Renewable Energy Laboratory, Golden, CO. https://www.nrel.gov/docs/fy16osti/66659.pdf.

Habte, A., M. Sengupta, A. Andreas, I. Reda, and J. Robinson. 2017. "Radiometer calibration methods and resulting irradiance differences." *Progress in Photovoltaics: Research and Applications* 25(7), 614–622. https://doi.org/10.1002/pip.2812.

Habte, A., M. Sengupta, A. Andreas, S. Wilcox, and T. Stoffel. 2016. "Intercomparison of 51 radiometers for determining global horizontal irradiance and direct normal irradiance measurements." *Solar Energy* 133, 372–393. https://doi.org/10.1016/j.solener.2016.03.065.

Habte, A., M. Sengupta, and A. Lopez. 2017. *Evaluation of the national solar radiation database (NSRDB): 1998-2015*. Golden, CO: NREL. https://www.nrel.gov/docs/fy17osti/67722.pdf.

Habte, A., M. Sengupta, I. Reda, A. Andreas, and J. Konings. 2014. *Calibration and measurement uncertainty estimation of radiometric data*. NREL/CP-5D00-62214, National Renewable Energy Laboratory, Golden, CO. https://www.nrel.gov/docs/fy15osti/62214.pdf.

Hamill, T. M. 2001. "Interpretation of Rank Histograms for Verifying Ensemble Forecasts." *Monthly Weather Review* 129(3), 550–560. https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Hersbach, H. 2000. "Decomposition of the Continuous Ranked Probability Score for Ensemble Prediction Systems." *Weather and Forecasting* 15(5), 559–570. https://doi.org/10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

Hodson, T. O. 2022. "Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not." *Geoscientific Model Development* 15(14), 5481–5487. https://doi.org/10.5194/gmd-15-5481-2022.

Hoff, T. E., and R. Perez. 2012. "Modeling PV fleet output variability." *Solar Energy* 86(8), 2177–2189. https://doi.org/10.1016/j.solener.2011.11.005.

Hoff, T. E., R. Perez, J. Kleissl, D. Renne, and J. Stein. 2013. "Reporting of irradiance modeling relative prediction errors." *Progress in Photovoltaics: Research and Applications* 21(7), 1514–1519. https://doi.org/10.1002/pip.2225.

ISO 9059. 1990. *ISO 9059:1990 Solar energy—Calibration of field pyrheliometers by comparison to a reference pyrheliometer*. https://www.iso.org/standard/16628.html.

ISO 9060. 2018. *ISO 9060:2018 Solar energy—Specification and classification of instruments for measuring hemispherical solar and direct solar radiation*. https://www.iso.org/obp/ui/en/#iso:std:iso:9060:ed-2:v1:en.

ISO 9846. 1993. *ISO 9846:1993 Solar energy—Calibration of a pyranometer using a pyrheliometer*. https://www.iso.org/standard/17724.html.

ISO 9847. 2023. *ISO 9847:2023 Solar energy—Calibration of pyranometers by comparison to a reference pyranometer*. https://www.iso.org/standard/78800.html.

ISO/IEC. 2008. "Uncertainty of measurement: Part 3. Guide to the expression of uncertainty in measurement." In *ISO/IEC Guide 983: 2008*. https://www.iso.org/standard/50461.html.

ISO/TR 9901. 2021. *ISO/TR 9901:2021 Solar energy—Pyranometers—Recommended practice for use*. https://www.iso.org/standard/81937.html.

Jimenez, P. A., J.P. Hacker, J. Dudhia, S.E. Haupt, J.A. Ruiz-Arias, C.A. Gueymard, G. Thompson, T. Eidhammer, and A. Deng. 2016. "WRF-Solar: Description and Clear-Sky Assessment of an Augmented NWP Model for Solar Power Prediction." *Bulletin of the American Meteorological Society* 97(7), 1249–1264. https://doi.org/10.1175/BAMS-D-14-00279.1.

Jolliffe, I. T., and D.B. Stephenson. (Eds.). 2011. *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (2nd ed.). Hoboken: Wiley.

Kamath, H. G., and J. Srinivasan. 2020. "Validation of global irradiance derived from INSAT-3D over India." *Solar Energy* 202, 45–54. https://doi.org/10.1016/j.solener.2020.03.084.

Kleissl, J., C. Coimbra, and R. Marquez. 2013. "Overview of Solar Forecasting Methods and a Metric for Accuracy Evaluation." In J. Kleissl (Ed.) *Solar Energy Forecasting and Resource Assessment*. https://doi.org/10.1016/B978-0-12-397177-7.00008-5.

Konings, J., and A. Habte. 2016. *Uncertainty evaluation of measurements with pyranometers and pyrheliometers*. Solar World Congress, Daegu, Korea. https://doi.org/10.18086/swc.2015.07.15.

Kuhn, P., B. Nouri, S. Wilbert, C. Prahl, N. Kozonek, T. Schmidt, Z. Yasser, L. Ramirez, L. Zarzalejo, A. Meyer, et al. 2018. "Validation of an all-sky imager–based nowcasting system for industrial PV plants." *Progress in Photovoltaics: Research and Applications* 26(8), 608–621. https://doi.org/10.1002/pip.2968.

Kühnert, J., E. Lorenz, and D. Heinemann. 2013. "Satellite-Based Irradiance and Power Forecasting for the German Energy Market." In *Solar Energy Forecasting and Resource Assessment*. Elsevier. https://doi.org/10.1016/B978-0-12-397177-7.00011-5.

Lauret, P., M. David, and P. Pinson. 2019. "Verification of solar irradiance probabilistic forecasts." *Solar Energy* 194, 254–271. https://doi.org/10.1016/j.solener.2019.10.041.

Lorenz, E., A. Hammer, and D. Heinemann. 2004. *Short term forecasting of solar radiation based on satellite data*. https://api.semanticscholar.org/CorpusID:16951232.

Lorenz, E., J. Hurka, D. Heinemann, H.G. Beyer. 2009. "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2(1), 2–10. https://doi.org/10.1109/JSTARS.2009.2020300.

Lorenz, E., J. Kühnert, D. Heinemann, K.P. Nielsen, J. Remund, and S.C. Müller. 2016. "Comparison of global horizontal irradiance forecasts based on numerical weather prediction models with different spatio-temporal resolutions." *Progress in Photovoltaics: Research and Applications* 24(12), 1626–1640. https://doi.org/10.1002/pip.2799.

Lorenz, E., T. Scheidsteger, J. Hurka, D. Heinemann, and C. Kurz. 2011. "Regional PV power prediction for improved grid integration." *Progress in Photovoltaics: Research and Applications* 19(7), 757–771. https://doi.org/10.1002/pip.1033.

Markovics, D., and M.J. Mayer. 2022. "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction." *Renewable and Sustainable Energy Reviews* 161, 112364. https://doi.org/10.1016/j.rser.2022.112364.

Massey Jr, F. J. 1951. "The Kolmogorov-Smirnov test for goodness of fit." *Journal of the American Statistical Association*, 46(253), 68–78. https://doi.org/10.2307/2280095.

Mathiesen, P., and J. Kleissl. 2011. "Evaluation of numerical weather prediction for intraday solar forecasting in the continental United States." *Solar Energy* 85(5), 967–977. https://doi.org/10.1016/j.solener.2011.02.013.

Mayer, B., and A. Kylling. 2005. "Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use." *Atmospheric Chemistry and Physics* 5(7), 1855–1877. https://doi.org/10.5194/acp-5-1855-2005.

Michalsky, J., E.G. Dutton, D. Nelson, J. Wendell, S. Wilcox, A. Andreas, P. Gotseff, D. Myers, I. Reda, and T. Stoffel. 2011. "An Extensive Comparison of Commercial Pyrheliometers under a Wide Range of Routine Observing Conditions." *Journal of Atmospheric and Oceanic Technology* 28(6), 752–766. https://doi.org/10.1175/2010JTECHA1518.1.

Michalsky, J., E. Dutton, M. Rubes, D. Nelson, T. Stoffel, M. Wesley, M. Splitt, and J. DeLuisi. 1999. "Optimal Measurement of Surface Shortwave Irradiance Using Current Instrumentation." *Journal of Atmospheric and Oceanic Technology* 16(1), 55–69. https://doi.org/10.1175/1520-0426(1999)016<0055:OMOSSI>2.0.CO;2.

Müller, S. C., and J. Remund. 2010. "Advances in Radiation Forecast Based on Regional Weather Models MM5 and WRF." *25th European Photovoltaic Solar Energy Conference and Exhibition / 5th World Conference on Photovoltaic Energy Conversion*. https://doi.org/10.4229/25THEUPVSEC2010-4BV.1.75.

Murphy, A. H. 1993. "What is a good forecast? An essay on the nature of goodness in weather forecasting." *Weather and Forecasting* 8(2), 281–293. https://doi.org/10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Myers, D. R., T.L. Stoffel, I. Reda, S.M. Wilcox, and A.M. Andreas. 2002. "Recent progress in reducing the uncertainty in and improving pyranometer calibrations." *J. Sol. Energy Eng.* 124(1), 44–50. https://doi.org/10.1115/1.1434262.

Pelland, S., G. Galanis, and G. Kallos. 2013. "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model." *Progress in Photovoltaics: Research and Applications* 21(3), 284–296. https://doi.org/10.1002/pip.1180.

Perez, R., P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, and F. Vignola. 2002. "A new operational model for satellite-derived irradiances: Description and validation." *Solar Energy* 73(5), 307–317. https://doi.org/10.1016/S0038-092X(02)00122-6.

Perez, R., R. Seals, P. Ineichen, R. Stewart, D. Menicucci. 1987. "A new simplified version of the perez diffuse irradiance model for tilted surfaces." *Solar Energy* 39(3), 221–231. https://doi.org/10.1016/S0038-092X(87)80031-2.

Pinson, P., P. McSharry, and H. Madsen. 2010. "Reliability diagrams for non-parametric density forecasts of continuous variables: Accounting for serial correlation." *Quarterly Journal of the Royal Meteorological Society* 136(646), 77–90. https://doi.org/10.1002/qj.559.

Pinson, P., H. Nielsen, J.K. Møller, H. Madsen, and G.N. Kariniotakis. 2007. "Non-parametric probabilistic forecasts of wind power: Required properties and evaluation." *Wind Energy* 10(6), 497–516. https://doi.org/10.1002/we.230.

Reda, I. 2011. *Method to Calculate Uncertainty Estimate of Measuring Shortwave Solar Irradiance using Thermopile and Semiconductor Solar Radiometers*. NREL/TP-3B10-52194. National Renewable Energy Laboratory, Golden, CO. https://doi.org/10.2172/1021250.

Reda, I., M. Dooraghi, and A. Habte. 2013. *NREL Pyrheliometer Comparison: September 16 to 27, 2013 (NPC-2013)*. NREL/TP-3B10-60749, National Renewable Energy Laboratory, Golden, CO. https://doi.org/10.2172/1111204.

Reda, I., D. Myers, and T. Stoffel. 2008. "Uncertainty Estimate for the Outdoor Calibration of Solar Pyranometers: A Metrologist Perspective." *NCSLI Measure* 3(4), 58–66. https://doi.org/10.1080/19315775.2008.11721448.

Renné, D., R. Perez, A. Zelenka, C.H. Whitlock, and R.C. DiPasquale. 1999. "Use of Weather and Climate Research Satellites for Estimating Solar Resources." In *Advances in Solar Energy: An Annual Review of Research and Development* (Vol. 13). American Solar Energy Society.

Ruiz-Arias, J. A., C.A. Gueymard, F.J. Santos-Alamillos, and D. Pozo-Vázquez. 2016. "Worldwide impact of aerosol's time scale on the predicted long-term concentrating solar power potential." *Scientific Reports* 6(1), 1–10. https://doi.org/10.1038/srep30546.

Salazar, G., C.A. Gueymard, J.B. Galdino, O. de C. Vilela, and N. Fraidenraich. 2020. "Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil." *Renewable and Sustainable Energy Reviews* 117, 109478. https://doi.org/10.1016/j.rser.2019.109478.

Sengupta, M., P.A. Jimenez, J-H Kim, J. Yang, and Y. Xie. 2022. *Probabilistic Cloud Optimized Day-Ahead Forecasting System Based on WRF-Solar (Final Report)*. NREL/TP-5D00-81904, National Renewable Energy Laboratory, Golden, CO. https://doi.org/10.2172/1855782.

Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby. 2018. "The National Solar Radiation Data Base (NSRDB)." *Renewable and Sustainable Energy Reviews* 89, 51–60. https://doi.org/10.1016/j.rser.2018.03.003.

Shi, H., W. Li, X. Fan, J. Zhang, B. Hu, L. Husi, H. Shang, X. Han, Z. Song, and Y. Zhang. 2018. "First assessment of surface solar irradiance derived from Himawari-8 across China." *Solar Energy* 174, 164–170. https://doi.org/10.1016/j.solener.2018.09.015.

Stoffel, T. L., I. Reda, D.R. Myers, D. Renne, S. Wilcox, and J. Treadwell. 2000. "Current issues in terrestrial solar radiation instrumentation for energy, climate, and space applications." *Metrologia* 37(5), 399. https://doi.org/10.1088/0026-1394/37/5/11.

Sun, X., D. Yang, C.A. Gueymard, J.M. Bright, and P. Wang. 2022. "Effects of spatial scale of atmospheric reanalysis data on clear-sky surface radiation modeling in tropical climates: A case study for Singapore." *Solar Energy* 241, 525–537. https://doi.org/10.1016/j.solener.2022.06.001.

Suri, M., and T. Cebecauer. 2014. "Satellite-based solar resource data: Model validation statistics versus user's uncertainty." *ASES SOLAR 2014 Conference, San Francisco*.

Taylor, K. 2001. "Summarizing multiple aspects of model performance in a single diagram." *Journal of Geophysical Research* 106, 7183–7192. https://doi.org/10.1029/2000JD900719.

Thevenard, D., and S. Pelland. 2013. "Estimating the uncertainty in long-term photovoltaic yield predictions." *Solar Energy* 91, 432–445. https://doi.org/10.1016/j.solener.2011.05.006.

Urraca, R., A.M. Gracia-Amillo, T. Huld, F.J. Martinez-de-Pison, J. Trentmann, A.V. Lindfors, A. Riihelä, and A. Sanz-Garcia. 2017. "Quality control of global solar radiation data with satellite-based products." *Solar Energy* 158, 49–62. https://doi.org/10.1016/j.solener.2017.09.032.

Urraca, R., T. Huld, A. Gracia-Amillo, F.J. Martinez-de-Pison, F. Kaspar, and A. Sanz-Garcia. 2018. "Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data." *Solar Energy* 164, 339–354. https://doi.org/10.1016/j.solener.2018.02.059.

Vignola, F., and R. Perez. 2004. *Solar Resource GIS Data Base for the Pacific Northwest using Satellite Data*. University of Oregon.

Vuilleumier, L., M. Hauser, C. Félix, F. Vignola, P. Blanc, A. Kazantzidis, and B. Calpini. 2014. "Accuracy of ground surface broadband shortwave radiation monitoring." *Journal of Geophysical Research: Atmospheres* 119(24), 13,838-13,860. https://doi.org/10.1002/2014JD022335.

Wang, L., Y. Lu, Z. Wang, H. Li, and M. Zhang. 2024. "Hourly solar radiation estimation and uncertainty quantification using hybrid models." *Renewable and Sustainable Energy Reviews* 202: 114727. https://doi.org/10.1016/j.rser.2024.114727.

Wilbert, S., S. Kleindiek, B. Nouri, N. Geuder, A. Habte, M. Schwandt, and F. Vignola. 2016. "Uncertainty of rotating shadowband irradiometers and Si-pyranometers including the spectral irradiance error." *AIP Conference Proceedings* 1734: 150009. https://doi.org/10.1063/1.4949241.

Wilcox, S., and D.R. Myers. 2008. *Evaluation of radiometers in full-time use at the national renewable energy laboratory solar radiation research laboratory*. NREL/TP-550-44627, National Renewable Energy Laboratory, Golden, CO. https://doi.org/10.2172/946331.

Wilks, D. S. 2011. *Statistical methods in the atmospheric sciences* (Vol. 100). Academic press.

WMO, W. 2018. "Guide to instruments and methods of observation." *World Meteorological Organization WMO*. https://library.wmo.int.

Xie, Y., J. Yang, M. Sengupta, Y. Liu, and X. Zhou. 2022. "Improving the prediction of DNI with physics-based representation of all-sky circumsolar radiation." *Solar Energy* 231, 758–766. https://doi.org/10.1016/j.solener.2021.12.016.

Yang, D. 2019a. "On postprocessing day-ahead NWP forecasts using Kalman filtering." *Solar Energy* 182, 179–181. https://doi.org/10.1016/j.solener.2019.02.044.

Yang, D. 2019b. "Standard of reference in operational day-ahead deterministic solar forecasting." *Journal of Renewable and Sustainable Energy* 11(5), 053702. https://doi.org/10.1063/1.5114985.

Yang, D. 2020. "Ensemble model output statistics as a probabilistic site-adaptation tool for solar irradiance: A revisit." *Journal of Renewable and Sustainable Energy* 12(3), 036101. https://doi.org/10.1063/5.0010003.

Yang, D., S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H.G. Beyer, R. Blaga, J. Boland, J.M. Bright, C.F.M. Coimbra, et al. 2020. "Verification of deterministic solar forecasts." *Special Issue on Grid Integration* 210, 20–37. https://doi.org/10.1016/j.solener.2020.04.019.

Yang, D., Y. Gu, M.J. Mayer, C.A. Gueymard, W. Wang, J. Kleissl, M. Li, Y. Chu, and J. Bright. 2024. "Regime-dependent 1-min irradiance separation model with climatology clustering." *Renewable and Sustainable Energy Reviews* 189, 113992. https://doi.org/10.1016/j.rser.2023.113992.

Yang, J., Y. Xie, M. Sengupta, Y. Liu, and H. Long. 2022. "Parameterization of cloud transmittance for expeditious assessment and forecasting of all-sky DNI." *Journal of Renewable and Sustainable Energy*14(6), 063703. https://doi.org/10.1063/5.0127454.

Zelenka, A., R. Perez, R. Seals, and D. Renné. 1999. "Effective accuracy of satellite-derived hourly irradiances." *Theoretical and Applied Climatology* 62(3), 199–207. https://doi.org/10.1007/s007040050084.