

## REPORTING OF IRRADIANCE MODEL RELATIVE ERRORS

Thomas E. Hoff  
Clean Power Research  
Napa, CA 94558  
[tomhoff@cleanpower.com](mailto:tomhoff@cleanpower.com)

Richard Perez  
ASRC, 251 Fuller Rd  
Albany, NY, 12203  
[rperez@albany.edu](mailto:rperez@albany.edu)

Jan Kleissl  
UCSD  
La Jolla, CA 92093  
[jkleissl@ucsd.edu](mailto:jkleissl@ucsd.edu)

David Renne  
NREL (Emeritus)  
Golden, CO 80401  
[drenne@mac.com](mailto:drenne@mac.com)

Joshua S. Stein  
Sandia National Labs  
Albuquerque, NM 87123  
[jsstein@sandia.gov](mailto:jsstein@sandia.gov)

### ABSTRACT

Metrics used in assessing irradiance model accuracy such as Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) are precisely defined. Their relative (%) counterpart, however, can be subject to interpretation and may cover a wide range of values for a given set of data depending on reporting practice. This article evaluates different approaches for the reporting of relative dispersion accuracy of a model and formulates recommendations for the most appropriate approach.

### 1. INTRODUCTION

Statistical methods for calculating and quantifying error have long been established across a wide range of sciences and industries. Whether quantifying the accuracy of an electrical meter, the tolerance of a precision part, or the expected range of forecasted temperatures, the methods for determining error are generally accepted. It is somewhat surprising, then, that these same methods have proved confusing and sometimes misleading when applied to commonly used diurnal quantities in the solar energy field. Error calculations related to solar irradiance and PV power production, for example, are complicated by observations taken during nighttime and other low solar conditions. These conditions are often of little interest to the solar researcher, but they do cover the majority of time over a multi-day test period. Since these observations are subject to

very low absolute error, their inclusion and weighting have a large impact on overall relative error.

As part of recent European and International Energy Agency (IEA) tasks [1], [2], a group of experts have developed recommendations for reporting irradiance model accuracy [3], [4]. Root Mean Square Error (RMSE), Mean Bias Error (MBE) and Kolmogorov Smirnov Integral (KSI) are the three key recommended validation metrics. These respectively provide a measure of model's dispersion (RMSE), overall bias (MBE), and ability to reproduce observed frequency distributions (KSI).

In many contexts, however, relative error is more commonly desired than absolute error. While the IEA tasks developed recommendations for absolute errors, they have not developed recommendations for reporting error in percentage terms, aside from using the informally (but not universally) accepted approach of dividing RMSE by the day-time mean of the considered irradiance. This is unfortunate because users in the utility industry desire to understand error in relative terms rather than absolute terms.

A simplified reporting approach for the %KSI metric was proposed in a recent article [5]. The present note focuses on the relative dispersion error metrics (RMSE and MAE) with the objective of setting a standard for reporting these metrics in the industry and research community to facilitate comparison between forecast models. Forecast model error also depends on meteorological conditions, forecast horizon, and averaging interval. This is not an attempt to create a

metric that makes forecasts comparable across these dependencies. Rather the focus is on which metric should be chosen to compare two forecasts at the same site, same forecast horizon, and same averaging interval. This discussion only focuses on methods concerned with expressing the relative error between two time series with a single statistic.

## 2. METHODS

### 2.1 Absolute Dispersion Errors

#### 2.1.1 Root Mean Square Error (RMSE)

The RMSE is defined to be the square root of the sum of the squares of the difference between modeled and reference irradiances using some time interval (e.g., hourly) over some time period (e.g., one year) divided by the number of observations.

$$RMSE = \sqrt{\frac{\sum_{t=1}^N (I_t^{test} - I_t^{ref})^2}{N}} \quad (1)$$

where  $I_t^{test}$  is the test irradiance at time  $t$ ,  $I_t^{ref}$  is the reference irradiance at time  $t$ , and  $N$  is the number of observations.

One ambiguity with the RMSE calculation (as well as all other error calculations that involve any sort of averaging) is that a decision is required as to whether or not to include all values. The prevalent practice in the solar resource community has been to only include daytime values, sometimes filtered by solar zenith angle less than  $80^\circ$  to avoid shading and/or sensor cosine response issues under low sun angles.

#### 2.1.2 Mean Absolute Error (MAE)

The MAE is defined to be the sum of the absolute differences between modeled and reference irradiances using some time interval over some time period divided by the number of observations.

$$MAE = \frac{\sum_{t=1}^N |I_t^{test} - I_t^{ref}|}{N} \quad (2)$$

### 2.2 Relative (Percent) Errors

Quantifying relative error requires that absolute error (i.e., RMSE or MAE) be divided by a normalizing number. To emphasize, the normalization is not carried out for each  $I_t^{test}$  and  $I_t^{ref}$  pair, but rather using a single number representative of typical irradiances during the entire time series. Three possible candidates to use in the denominator to calculate *Percent Error* are:

- Average irradiance (Avg.).
- Weighted average irradiance (Weighted Avg.).
- Capacity (Maximum Nominal irradiance)

#### 2.2.1 Average Irradiance

Average irradiance equals the sum of the irradiance values divided by the number of observations.

$$Average = \frac{\sum_{t=1}^N I_t^{ref}}{N} \quad (3)$$

#### 2.2.2 Weighted Average Irradiance

Weighted Average irradiance may be used to assign more importance to high-level irradiance observations. It is defined to be the sum of the irradiance values weighted by a factor.

$$Weighted\ Average = \sum_{t=1}^N W_t I_t^{ref} \quad (4)$$

One meaningful way to weight the irradiance is by its magnitude. That is, let

$$W_t = \frac{I_t^{ref}}{\sum_{t=1}^N I_t^{ref}} \quad (5)$$

Substituting Equation (5) into Equation (4) results in a Weighted Average of

$$Weighted\ Average = \frac{\sum_{t=1}^N (I_t^{ref})^2}{\sum_{t=1}^N I_t^{ref}} \quad (6)$$

Unlike for the simple average, the day-time weighted average equals the 24-hour weighted averages since the weight of night-time points is zero.

#### 2.2.3 Capacity

A third option is to use the peak irradiance or Capacity (C). For global horizontal irradiance, for example, the Capacity would be  $1,000 \text{ Wm}^{-2}$ .

The wind industry has adopted this approach of normalizing to installed generating capacity for the reporting of output prediction errors [6].

#### 2.2.4 Percent Error Calculation Methods

With two measures of dispersion (RMSE and MAE) and three normalizing means, there are six possible methods to calculate *Percent Error*. These methods are summarized in TABLE 1. TABLE 2 presents the mathematical definitions used to calculate *Percent Error* by combining Equations (1) through (6) (see appendix for the detailed derivations).

**TABLE 1. POSSIBLE PERCENT ERROR CALCULATION METHODS**

	<b>RMSE</b>	<b>MAE</b>
<b>Average</b>	RMSE/Avg.	MAE/Avg.
<b>Weighted Average</b>	RMSE/Weighted Avg.	MAE/Weighted Avg.
<b>Capacity</b>	RMSE/Capacity	MAE/Capacity

**TABLE 2. MATHEMATICAL DEFINITIONS OF PERCENT ERROR METHODS.**

<b>Percent Error Method</b>	<b>Definition</b>
RMSE/Avg.	$\sqrt{N} \left( \frac{1}{\sum_{t=1}^N I_t^{ref}} \right) \sqrt{\sum_{t=1}^N (I_t^{test} - I_t^{ref})^2}$
RMSE/Weighted Avg.	$\left[ \frac{1}{\sqrt{N}} \right] \left[ \frac{\sum_{t=1}^N I_t^{ref}}{\sum_{t=1}^N (I_t^{ref})^2} \right] \sqrt{\sum_{t=1}^N (I_t^{test} - I_t^{ref})^2}$
RMSE/Capacity	$\left( \frac{1}{\sqrt{N}} \right) \left( \frac{1}{C} \right) \sqrt{\sum_{t=1}^N (I_t^{test} - I_t^{ref})^2}$
MAE/Avg.	$\left( \frac{1}{\sum_{t=1}^N I_t^{ref}} \right) \sum_{t=1}^N  I_t^{test} - I_t^{ref} $
MAE/Weighted Avg.	$\left[ \frac{1}{N} \right] \left[ \frac{\sum_{t=1}^N I_t^{ref}}{\sum_{t=1}^N (I_t^{test})^2} \right] \sum_{t=1}^N  I_t^{test} - I_t^{ref} $
MAE/Capacity	$\left( \frac{1}{N} \right) \left( \frac{1}{C} \right) \sum_{t=1}^N  I_t^{test} - I_t^{ref} $

2.2.5 24 hours vs. Daytime

The effect of including 24 hours in the analysis vs. only including daytime values can be analyzed using the equations presented in TABLE 2. *Total Error* (i.e.,  $\sqrt{\sum_{t=1}^N (I_t^{test} - I_t^{ref})^2}$  or  $\sum_{t=1}^N |I_t^{test} - I_t^{ref}|$ ), remains unchanged by including night-time values. However, *Absolute Error* (RMSE or MAE) is affected by the distinction since the results are obtained by dividing *Total Error* by the number of considered points. *Percent Error* is further affected by the daytime vs. 24-hour distinction since the normalizing means are different.

TABLE 3 summarizes the impact of the distinction on the selected error reporting metrics. It shows that *Percent Error* calculated using RMSE/Avg. method increases from 24 hour to daytime, the MAE/Avg. is unchanged, and *Percent Error* calculated using the other four methods decreases. In all of the changed scenarios, the change is a function of the fraction of daytime hours. For example, if there are 4,380 daytime hours in a 12-month test period, the fraction Daytime Hours is 0.5. If night time hours are considered, *Percent Error* calculated using RMSE/Avg. will increase by 41 percent ( $\sqrt{\frac{1}{0.5}}$ ), *Percent Error* calculated using RMSE/Weighted Avg. will decrease by 29 percent ( $\sqrt{0.5}$ ), and *Percent Error* calculated using MAE/Weighted Avg. or MAE/Capacity will decrease by 50 percent. The only method independent of nighttime hours is the MAE/Avg. method.

**TABLE 3. RATIO OF PERCENT ERROR USING ALL HOURS TO PERCENT ERROR USING DAYTIME HOURS.**

<b>Percent Error Method</b>	<i>Ratio of Daytime to 24h Percent Error</i>			
	$\sqrt{\frac{N^{All\ Hour}}{N^{Daytime\ Hour}}}$	100 % (No change)	$\sqrt{\frac{N^{Daytime\ Hour}}{N^{All\ Hour}}}$	$\frac{N^{Daytime\ Hour}}{N^{All\ Hour}}$
RMSE/Avg.	✓			
RMSE/Weighted Avg.			✓	
RMSE/Capacity			✓	
MAE/Avg.		✓		
MAE/Weighted Avg.				✓
MAE/Capacity				✓

3. APPLICATION EXAMPLE

An effective way to compare and contrast the six possible methods is to quantify results using an actual irradiance data set. Hourly satellite-derived global horizontal insolation (GHI) data was obtained for Hanford, CA, from January 1, 2010 to December 31, 2010. The reference data are from a high-quality ISIS ground site [7]. The modeled data are

from a satellite-based irradiance service [8]. Fig 1 plots one year’s worth of hourly modeled data vs. measured data. A perfect match would occur if all blue dots were on the red line. As can be seen from the figure, the selected modeled data are a good visual match to the reference data.

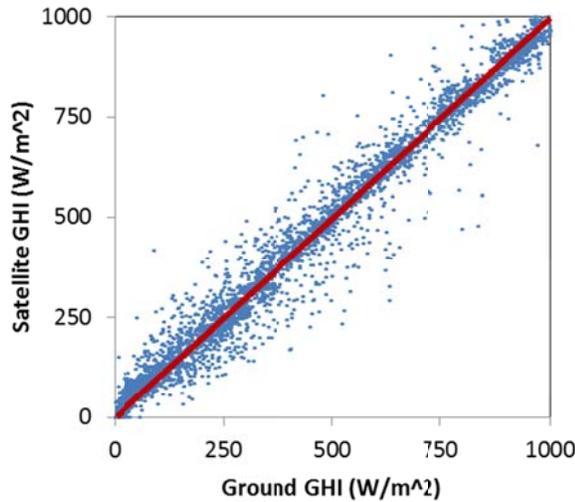


Fig. 1: Irradiance data for Hanford, CA, 2010.

Fig. 2 presents *Percent Error* for the six methods using the two scenarios of All Hours (24 hours per day) and Daytime Hours only. The “All Hours” scenarios are represented by the black bars. The “Daytime Hours” are represented by the white bars.

Several observations can be made based on the figure:

- *Percent Error* ranges by a factor of more than 10 depending upon which method and scenario is selected
  - RMSE/Avg. method using nighttime values results in a 17.0 *Percent Error*.
  - MAE/Capacity method using nighttime values results in 1.5 *Percent Error*.
- The exclusion/inclusion of nighttime values changes results for five of the six definitions; *Percent Error* is lower for one case and higher for four cases.
- Only the MAE/Avg. *Percent Error* definition is independent of the inclusion of nighttime data.

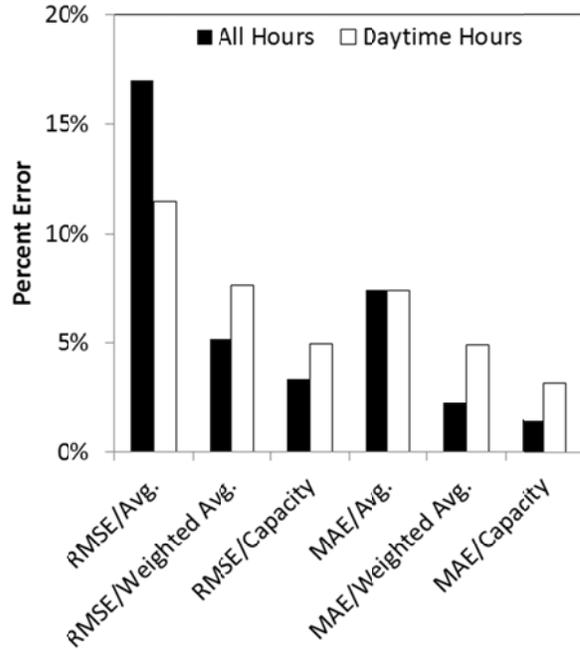


Fig. 2: Comparison of error results for six methods using “All Hours” and “Daytime Hours” for Hanford, CA, 2010

### 3. 1 Threshold Dependence

The *Irradiance Threshold* is the value below which data are excluded. Use of a threshold is relevant because while the current practice is to exclude night-time values, the industry lacks a precise definition of what is night-time. Is night-time when irradiance is 0 W/m<sup>2</sup>, 0.1 W/m<sup>2</sup>, 1 W/m<sup>2</sup>?

The 24-hour and daytime scenarios are specific threshold points, occurring respectively when irradiance is larger than, or equal to a zero *Irradiance Threshold* for the former and above the zero *Irradiance Threshold* for the latter.

Fig. 3 presents the percent of solar energy that occurs below a given *Irradiance Threshold*. It is interesting to note that much of the collectable energy resides above significant threshold levels. For example, the dashed line shows that GHI observations less than an *Irradiance Threshold* of 250 W/m<sup>2</sup> correspond to only 8 percent of the annual GHI at Hanford, CA in 2010.

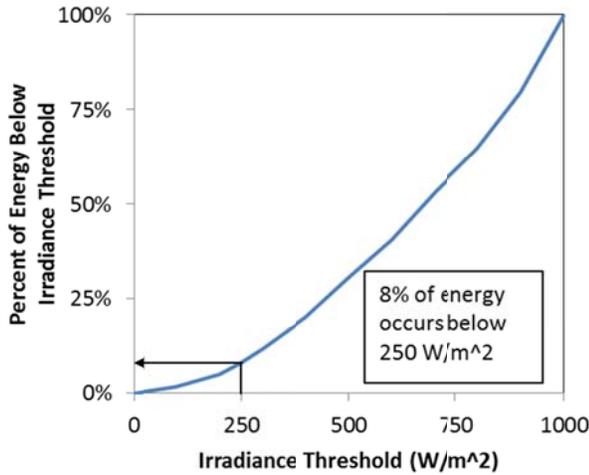


Fig. 3: Energy distribution of irradiance data for Hanford, CA, 2010.

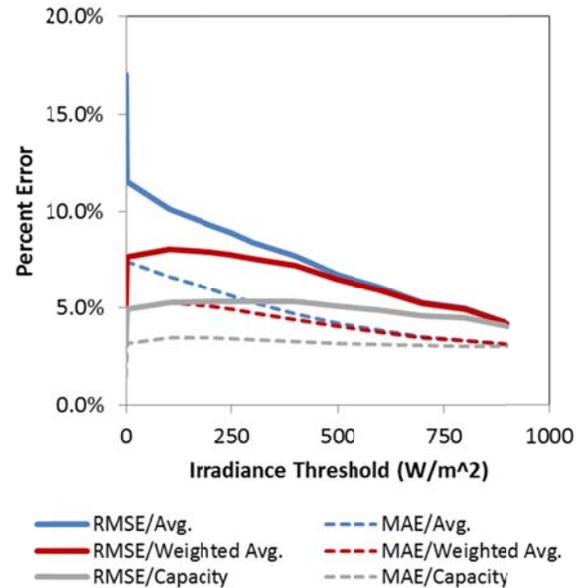


Fig. 4: Comparison of error results for Hanford, CA, 2010.

Fig. 4 presents *Percent Error* as a function of *Irradiance Threshold* for all six methods. Several observations can be made based on the figure.

- All *Percent Error* definitions based on *RMSE* converge to the same result as the *Irradiance Threshold* increases.
- All *Percent Error* definitions based on *MAE* converge to the same result as the *Irradiance Threshold* increases.
- *RMSE/Weighted Avg.* results are similar to *MAE/Avg.* when “Daytime Hours” are included.

### 3.2 RMSE vs. MAE

Aside from the *Percent Error* reporting issue, it is worthwhile to explore the question whether the *RMSE* or the *MAE* is the most appropriate method to report dispersion error.

The main difference between the two is that the *RMSE* is driven by the square of the differences unlike the *MAE*. As a result, outliers are considerably more influential on the reported accuracy when using the *RMSE* metric. In the above example the addition of four far outliers to the data set (representing 0.1 percent of the data samples) increases the *RMSE* by a factor of 1.12, but only increases the *MAE* by a factor of 1.04.

## 4. DISCUSSION

Table 4 summarizes the comparative observations made above using a subjective grading for the attributes of each relative dispersion error reporting method. The attributes we considered include:

- Whether the method is commonly accepted
- Whether it is simple to understand
- Whether it depends on the 24-hr vs. daytime only distinction
- Whether it depends on the data selection threshold
- Whether it is affected by outliers

A grade of 0 to 2 is assigned to each method to represent its strength (2) or its weakness (0) with respect to a given attribute.

The *MAE/Avg.* provides the best practical measure of relative dispersion error based on the selected evaluation criteria and the subjective evaluations. The *MAE/Avg.* is attractive in that it is independent of the number of observations and is simple to understand. The *RMSE/Capacity* method is also desirable because it is commonly accepted (the wind power industry has already adopted this method) and is simple to understand.

**TABLE 4. SUBJECTIVE EVALUATION OF RELATIVE ERROR REPORTING METHOD.**

	Com mon ly ac cepted	Sim ple to und erst and	Depe nds on nigh t time Value s	Depe nds on select ed thres hold	Affe cted by outli ers	T o t al
RMSE/Avg.	2	2	0	0	0	<b>4</b>
RMSE/Weighted Avg.	0	1	1	1	0	<b>3</b>
RMSE/Capacity	2	2	1	1	0	<b>6</b>
MAE/Avg.	1	2	2	1	1	<b>7</b>
MAE/Weighted Avg.	0	1	0	1	1	<b>3</b>
MAE/ Capacity	0	2	0	2	1	<b>5</b>

The value of agreeing on a simple-to-calculate method has the benefit that multiple predictions and forecasts can be quickly evaluated and compared. Given that irradiance and PV power predictions and forecasts will be applied to a variety of applications (resource assessment, electrical grid operations and planning, etc.), it is not expected that the single statistic proposed here will necessarily be a complete measure of forecast quality. The authors, however, feel that it is a good start towards promoting a standard metric in the industry.

**5. REFERENCES**

1. MESOR, 2002-2010 – Management and Exploitation of Solar Resource Knowledge. European Commission’s Community Research Sixth Framework Programme.
2. IEA Solar Heating and Cooling Programme, Task 36, Subtask A, (2011): Standard Qualification for Solar Resource Products.
3. Beyer G. G., J. Polo Martinez, M. Suri, J. L. Torres, E. Lorenz, S. Muller, C. Hoyer-Klick and P. Ineichen, (2009): Benchmarking of Radiation Products, Mesor Report D.1.1.3.
4. Espinar B., L. Ramirez, A. Drews, H. G. Beyer, L. Zarzalejo, J. Polo Martinez and L. Martin, (2009): Analysis of Different Comparison Parameters Applied to Solar Radiation from Satellite and German Radiometric Stations.
5. Perez, R., M. Beauharnois, K. Hemker Jr., S. Kivalov, E. Lorenz, S. Pelland, J. Schlemmer, and G. Van Knowe, (2011): Evaluation of Numerical Weather Prediction Solar Irradiance Forecasts in the US. Proc. Solar 2011, American Solar Energy Society’s Annual Conference.
6. Tastu, J., P. Oinson, E. Kotwa, H. Madsen and H. Nielsen, (2010): Spatio-Temporal Analysis and Modeling of Short-Term Wind Power Forecast Errors. Wind Energy, 14, 43-60.
7. Integrated Surface Irradiance Study (ISIS) Network, (1995-2011).
8. SolarAnywhere, available at [www.SolarAnywhere.com](http://www.SolarAnywhere.com).