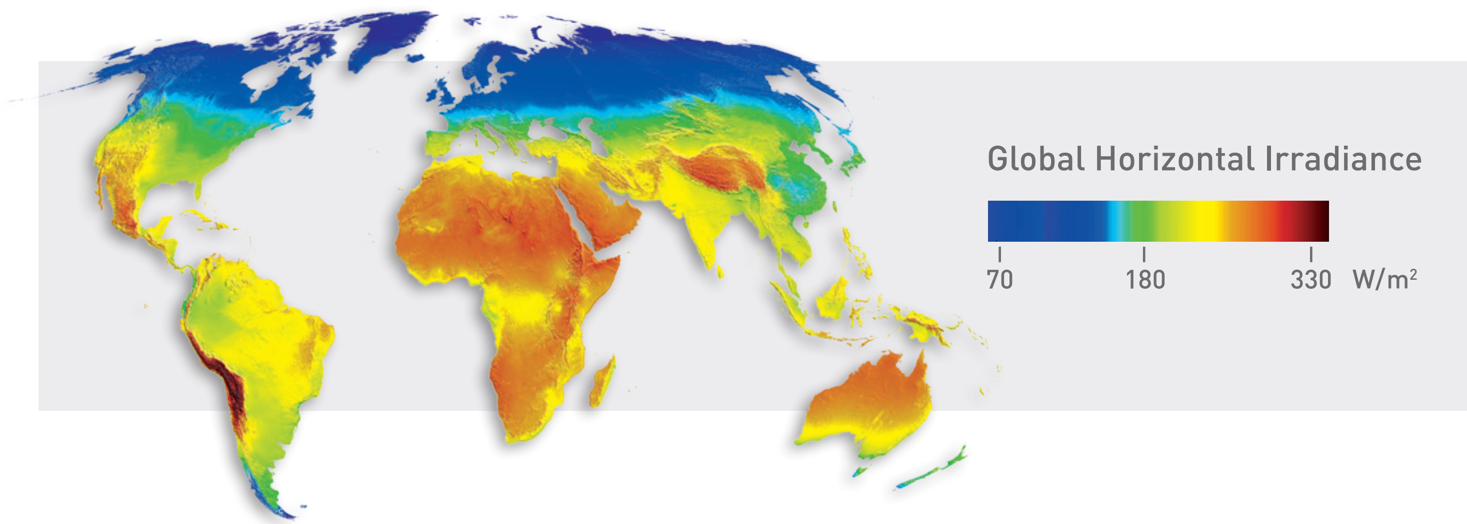
# 3TIER Global Solar Dataset: Methodology and Validation

October 2013

www.3tier.com




## INTRODUCTION

Solar energy production is directly correlated to the amount of radiation received at a project location. Like all weather-driven renewable resources, solar radiation varies rapidly over time and space and understanding this variability is crucial in determining the financial viability of a solar energy project.

The three components of irradiance most critical for determining solar installation production values are Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), and Diffuse Irradiance (DIF). Fixed panel photovoltaic (PV) installations are dependent on GHI, or the total amount of radiation received by a horizontal surface. Concentrated solar power (CSP) projects and PV tracking systems rely predominantly on DNI, which is the total amount of radiation received by a surface that is always kept perpendicular to the sun's direct rays. Most financing options for solar projects require information on the expected yearly irradiance values as projects typically have to service debt one to four times a year. However, annual averages do not provide enough information to determine accurate annual exceedance probabilities for irradiance and power production.

Depending on the characteristics of a site, studies have shown that on average, annual irradiance means can differ from the long-term mean by 5% for GHI and by as much as 20% for DNI. Thus a long-term record of solar irradiance estimates is needed to calculate a realistic variance of production values.

The existing network of surface observation stations is too sparse to quantify solar resources at most potential sites. A vast majority of stations only provide a limited short-term record of the resource (months to a few years), are rarely located near proposed sites, and are often plagued with measurement errors. Calculating site-specific solar irradiance values using geostationary satellite data is an accepted alternative. Within the global atmospheric science community, satellite derived values have proven to be more accurate than nearby surface observations for locations that are more than 25 km away from a ground station. 3TIER created a global satellite derived solar dataset to help clients determine solar variability at any site worldwide, from the prospecting stage through assessment and bankability.

In this paper, we will provide an outline of standard practices that should be followed to ensure accurate solar assessment. We will also describe the methodology 3TIER used to create its global solar dataset and provide an extensive validation, showing the accuracy of 3TIER's data by region in the Appendix.

## SOLAR DEVELOPMENT ROADMAP

Developing a solar project requires a large upfront investment. A standard development roadmap conserves time and money and ensures that only the most promising projects are constructed. Each stage of development asks different questions about the solar resource and each stage requires varying degrees of information and investment.

### Prospecting and Planning

The first step in building any solar resource project is identifying the regions most suitable for development. The price of energy, access to transmission, and environmental siting issues should all be taken into consideration, but the most essential variable is the availability of the solar resource – the "fuel" of the project. At this early stage, average annual and monthly solar irradiance values can be used to assess the overall feasibility of a particular site and to select the appropriate solar technology to be installed. Presently, basic solar annual and monthly averages can be found via online solar prospecting tools and in the form of GIS layers. These tools allow developers to quickly target the best locations for further investigation and identify red flags early in the process.

### Design and Due Diligence

Once a promising site is identified, a more in-depth analysis is required to better quantify the long-term availability of the solar resource, to design technical aspects of the project, and to secure the upfront capital for construction. A common source of solar data used for this purpose is Typical Meteorological Year (or TMY) data. A TMY dataset provides a 1-year, hourly record of typical solar irradiance and meteorological values for a specific location in a simple file format. Although not designed to show extremes, TMY datasets are based on a long time period and show seasonal variability and typical climatic conditions at a site. They are often used as an input to estimate average annual energy production. While TMY data provide a good estimate of the average solar irradiance at a site, they are not a good indicator of conditions over the next year, or even the next 5 years. The U.S. National Renewable Energy Laboratory User Manual for TMY3 data explicitly states, "TMY should not be used to predict weather for a particular period of time, nor are they an appropriate basis for evaluating real-time energy production or efficiencies for building design applications or a solar conversion system." Hourly time series covering a period of several years provide a much more complete record for calculating accurate estimates of solar resource variability.

Year-to-year variability has a significant impact on annual energy production. Many financial and rating institutions, as well as internal certification organizations, require 1-year P90 values to assess the economic feasibility of a project. A 1-year P90 value indicates the production value that the annual solar resource will exceed 90% of the time. A 1-year P90 value (as opposed to a 10-year P90 value) is typically mandatory because most solar projects have a lending structure that requires them to service debt one to four times a year, not one to four times every 10 years. If power production decreases significantly in a given year due to solar variability, debt on the project may not be able to be paid and the project could default on its loan. This is precisely what financiers are trying to avoid. The only way to determine 1-year P90 values acceptable to funding institutions is with long-term continuous data at the proposed site.

If collected properly, surface observations can provide very accurate measurements of solar radiation at high temporal resolution, but few developers want to wait the 5 to 10 years required to develop a 1-year P90 value. Satellite derived irradiance values can accurately provide a long-term, hourly time series of data without the expense and wait. However, satellite data cannot always capture the micro-scale features that affect a site. Therefore, a combination of short-term ground measurements and long-term satellite derived irradiance values is ideal for assessing variability and project risk. One method of combining short-term ground measurements with long-term satellite data is a technique known as model output statistics (or MOS). 3TIER's MOS algorithm can significantly reduce error and bias by statistically correcting our satellite derived irradiance values to the environmental context of a particular site based on available surface observations. More information on MOS can be found at:

http://www.3tier.com/en/support/glossary/#mos

**Operations and Optimization**

With more solar energy coming onto the grid everyday, effectively managing its integration is becoming increasingly important. Once a project is operational, forecasting plays a vital role in estimating hour and day ahead solar production and variability. This information is critical for estimating production, scheduling energy, managing a mixed energy portfolio, avoiding imbalance charges, and detecting reduced production days.

Some rudimentary NWP (Numerical Weather Prediction) modeling systems have been introduced for this purpose. However, 3TIER has found that basic NWP models poorly estimate cloud cover, the single variable that most directly impacts solar energy production.

To provide greater accuracy, 3TIER has developed a sophisticated method of combining our advanced NWP models with our long-term satellite derived dataset. Other factors that impact production, such as wind, temperature, and aerosols, can also be taken into account to better predict real-time solar production.

Recent solar irradiance observations can also be used to model the energy that a project should have produced. Comparing modeled production with actual production helps identify underperforming projects and explain to what extent solar variation is impacting production. This periodic, ongoing reconciliation helps pinpoint maintenance and equipment issues particularly for those with a geographically dispersed portfolio of projects.

## 3TIER SOLAR METHODOLOGY

3TIER developed and maintains a global, long-term, high-resolution solar dataset, which was created using satellite observations from around the world. As discussed earlier in this document, satellite derived data have proven to be the most accurate method of estimating surface solar irradiance beyond 25 km of a ground station. 3TIER's main source of satellite observations is weather satellites in a geo-stationary orbit. These satellites have the same orbital period as the Earth's rotation and, as a result, their instruments can make multiple observations of the same area with identical viewing geometry each hour.

3TIER uses visible satellite imagery to calculate the level of cloudiness at the Earth's surface. The resulting time series of cloudiness (or cloud index) is then combined with other information to model the amount of solar radiation at the Earth's surface. The outcome is a 15+ year dataset that provides hourly estimates of surface irradiance (GHI, DNI, and DIF) for all of the Earth's land mass at a spatial resolution of approximately 3 km.

The general methodology is similar to other satellite derived solar datasets, but a majority of the algorithms were developed in-house. 3TIER's dataset also includes several key improvements such as higher spatial and temporal resolution, empirical fitting, and a monthly time series of turbidity estimates. We have also developed a cloud-index algorithm that produces consistent results when used with the large number of different satellites that must be combined to construct a global dataset. In Figure 1, a basic flow chart presents our general methodology.

Satellite based time series of reflected sunlight are used to determine a cloud index time series for every land surface worldwide. A satellite based daily snow cover dataset is used to aid in distinguishing snow from clouds. In addition, the global horizontal clear-sky radiation (GHC), or the amount of radiation in the absence of clouds, is modeled based on the surface elevation of each location, the local time, and the measure of turbidity in the atmosphere. This latter quantity accounts for the transparency of the atmosphere and is affected by aerosols and water vapor. Unfortunately, direct observations of turbidity are made at only a few locations. 3TIER opted to use a satellite based, monthly time series of aerosol optical depth and water vapor derived from four datasets available from the Moderate Resolution Imaging Spectroradiometer (MODIS). This dataset was combined with another turbidity dataset that includes both surface and satellite observations to provide turbidity measurements that span the period of our satellite dataset and are complete for all land surfaces.

The cloud index and GHC are then combined to model GHI. This component of the process is calibrated for each satellite based on a set of high-quality surface observations. GHI estimates are then combined with other inputs to calculate DNI and DIF.

Despite the resolution of the dataset, some factors need to be taken into consideration by the user. 3TIER's global solar dataset does not directly account for local shades and shadows and, as a result, local conditions must be considered when interpreting the irradiance values. Also, in some areas with highly reflective terrain, such as salt flats and areas with permanent snow, the satellite algorithms have difficulty distinguishing clouds from the terrain. The cloudiness estimates in these areas are higher than they should be and the amount of GHI and DNI is underestimated and the DIF is overestimated. Known areas affected by this problem include highly reflective areas such as Lake Gairdner National Park in South Australia.

## VALIDATION OF 3TIER GLOBAL IRRADIANCE DATASET

An extensive validation of 3TIER's satellite data was performed using observations from 120 surface stations across the globe. In the study, 3TIER used stations from the World Climate Research Program and the Baseline Surface Radiation Network, National programs from the Indian Metrological Department, the Australian Bureau of Meteorology, the National Solar Radiation Database, and several others observational datasets. For quality control any negative or anomalously high irradiance values were removed from the observations prior to the analysis. The World Climate Research Program estimates ground solar radiation sites have inaccuracies of 6-12%. Specialized high quality research sites, such as those from the Baseline Surface Radiation Network, are possibly more accurate by a factor of two. These constraints make direct comparisons between solar radiation datasets difficult, but it is still possible to estimate the relative accuracy if the same reference observations are used. The statistics presented in the following sections were computed using only daytime irradiance values, which provide a better indication of the accuracy and value of the dataset.

### Global Validation Statistics

Table 1 provides a list of statistical metrics evaluating 3TIER's global dataset. There were 120 surface observations that had GHI measurements but only 96 stations with quality DNI measurements. The statistics are also shown in a global map including all the stations for GHI in Figures 2 and for DNI in Figure 3. The computed statistics include those most commonly used in the solar industry, such as mean bias error (MBE), mean absolute error (MAE), and hourly root mean square error (RMSE). Mean bias error (MBE) provides information about the average difference in the mean over the entire dataset when compared against observations. Mean absolute error (MAE) measures the average magnitude of the errors. Root mean square error (RMSE) also measures the average magnitude of the error, but uses quadratic weighting, which results in large errors carrying more weight. A smaller RMSE value means that the dataset more closely tracks observations on an hour-by-hour basis. Together MBE, MAE, and hourly RMSE can be used to assess the accuracy of a solar dataset compared to observations.

Globally, 3TIER GHI values show a MBE and MAE of 0.9% and 3.8% respectively and an hourly RMSE of approximately 24.1% when compared with observations. These errors are consistent across each of the six regions. This means 3TIER irradiance values can be used across the globe with the same degree of certainty for each region despite the vastly different meteorological conditions. The DNI errors are slightly higher showing a MBE and MAE of 0.7% and 8.6% respectively and an hourly RMSE of 43.1%. The higher RMSE for DNI is to be expected considering the significant hour-to-hour variability of DNI compared to GHI. It should be noted that all the mean errors are within the standard error of observations as determined by the World Climate Research Program.

### Interannual Variability and P90 values

For design and due diligence, calculating yearly P90 values is essential for project financing. Using average or TMY irradiance data is not sufficient information to understand the variability that affects solar projects. An example of this is seen in the GHI measurements for Desert Rock, Nevada (36.63N, 116.02W) and illustrated in Figure 4. Over a 10 year period, ground observations showed annual GHI variations were generally less than 5% from the mean, with one anomalous year of low GHI in 2002 where GHI was about 6% less than the long-term mean. 3TIER's satellite derived solar dataset captures this interannual variability with an annual correlation coefficient of 0.97 and an overall mean bias error of 2.5%.

A 3TIER MOS-corrected time series was derived using 1 year of ground observations from Desert Rock from 2008 and the 3TIER satellite record at that location. The MOS-corrected time series retained the high correlations with the observed data (annual correlation coefficient of 0.97) and also reduced bias from 2.5% to less than 0.5%. This MOS-corrected time series provides a long-term perspective using short-term observations and can be used as a way to calculate the 1-year P90 values.

## CONCLUSION

The development of solar projects has expanded significantly and appears to have a promising future. However, even the best locations are not immune to normal year-to-year variations in solar irradiance, which have a corresponding impact on power production and the ability of the project to service its debt. While on-site observations capture the localized nuances of solar irradiance at a particular location, they do not provide the long-term perspective required for project funding. Satellite derived solar datasets, on the other hand, accurately capture year-to-year fluctuations, but do not always capture micro-scale features. A solar resource assessment, combining both on-site observations and long-term, satellite derived data greatly reduces uncertainty, and provides the "bankable" production estimates required to secure financing. Widespread adoption of this assessment technique will ensure that only profitable solar projects are constructed and secure the future success of the solar energy industry.

## REFERENCES


[1] Gueymard, C., Wilcox, S., *Spatial and Temporal Variability in the Solar Resource.* Buffalo, NY: American Solar Energy Society, May 2009. Print.

[2] Perez, R., Ineichen, P., Moore, K., Kmiecik, M., Chain, C., George, R., Vignola, F. "A New Operational Satellite-to-Irradiance Model." *Solar Energy* 73.5 (2002): 307-317. Print.

[3] Zelenka, A., Perez, R., Seals, R., Renné, D. "Effective Accuracy of Satellite Derived Irradiance." *Theoretical and Applied Climatology.* 62.3-4 (1999): 199-207. Print.

[4] Marion, W. and Wilcox. S. *Users Manual for TMY3 Data sets.* Golden: National Renewable Energy Laboratory, 2008.

[5] Venkataraman, S., D'Olier-Lees, T. "Key Credit Factors." *Standard and Poor's Solar Credit Weekly* 29.42 (2009), 29-30, 49-50. E-article.

[6] "SSE Release 6.0 Methodology" *nasa.gov*. National Aeronautics and Space Administration. 2010. Web. 28 Sept. 2010.

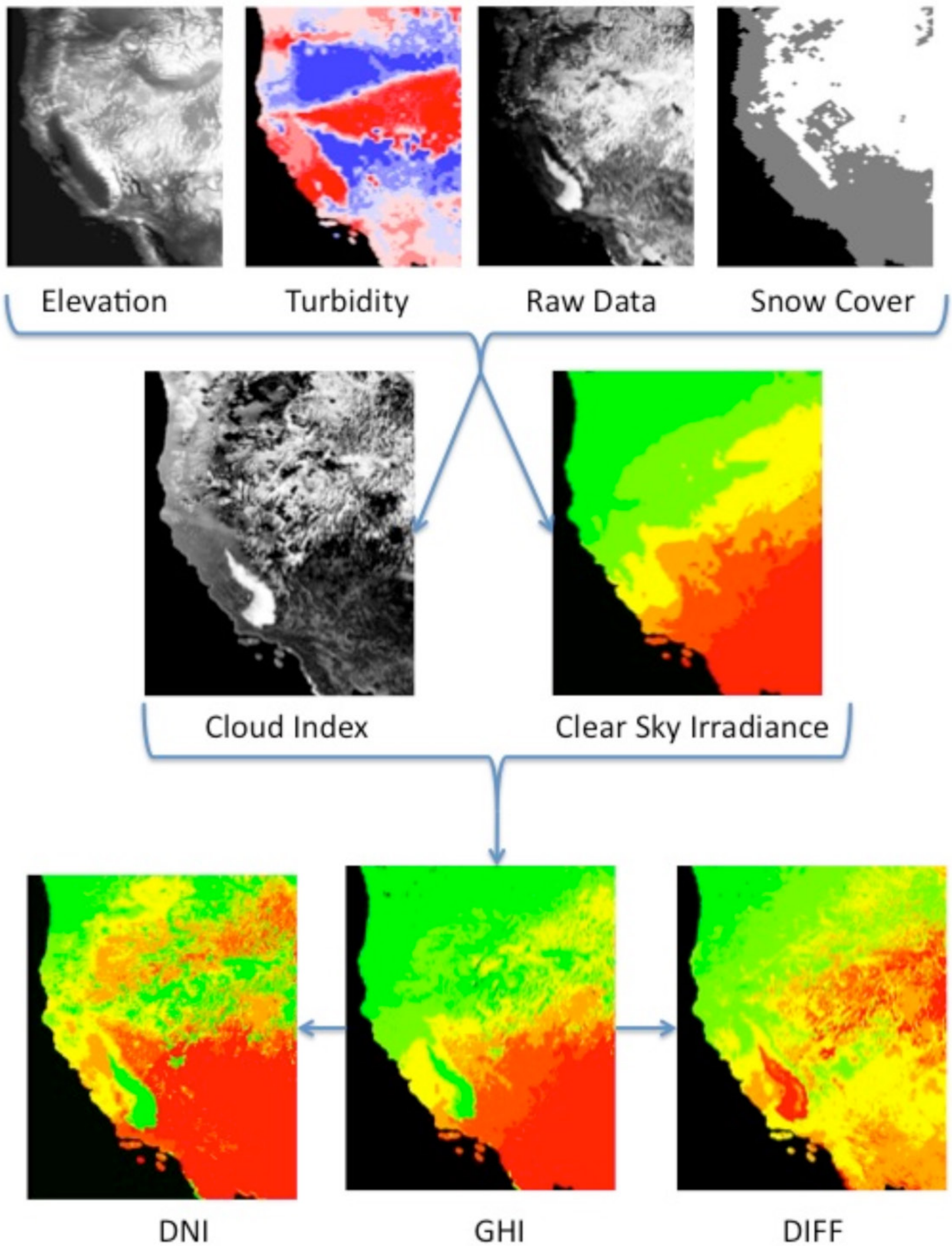
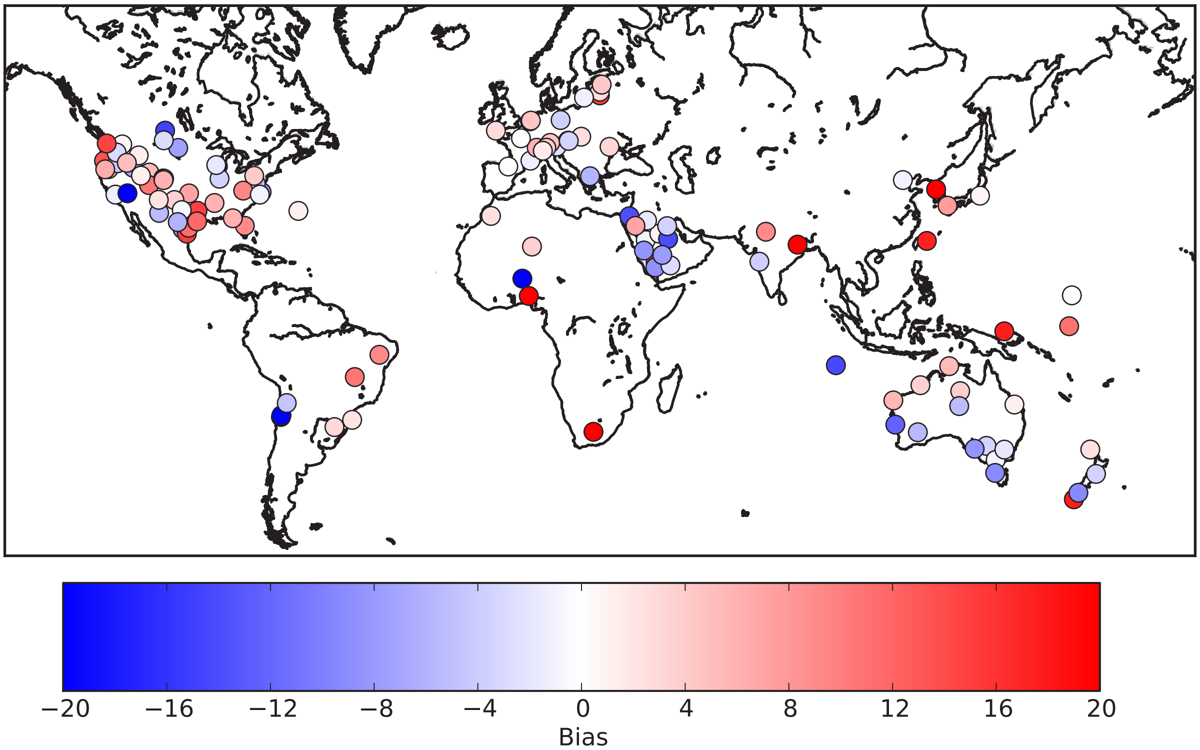



*Figure 1.* A flow chart showing 3TIER's method for calculating solar irradiance at the Earth's surface.

*Figure 2.* Global maps showing mean bias and RMSE for GHI.



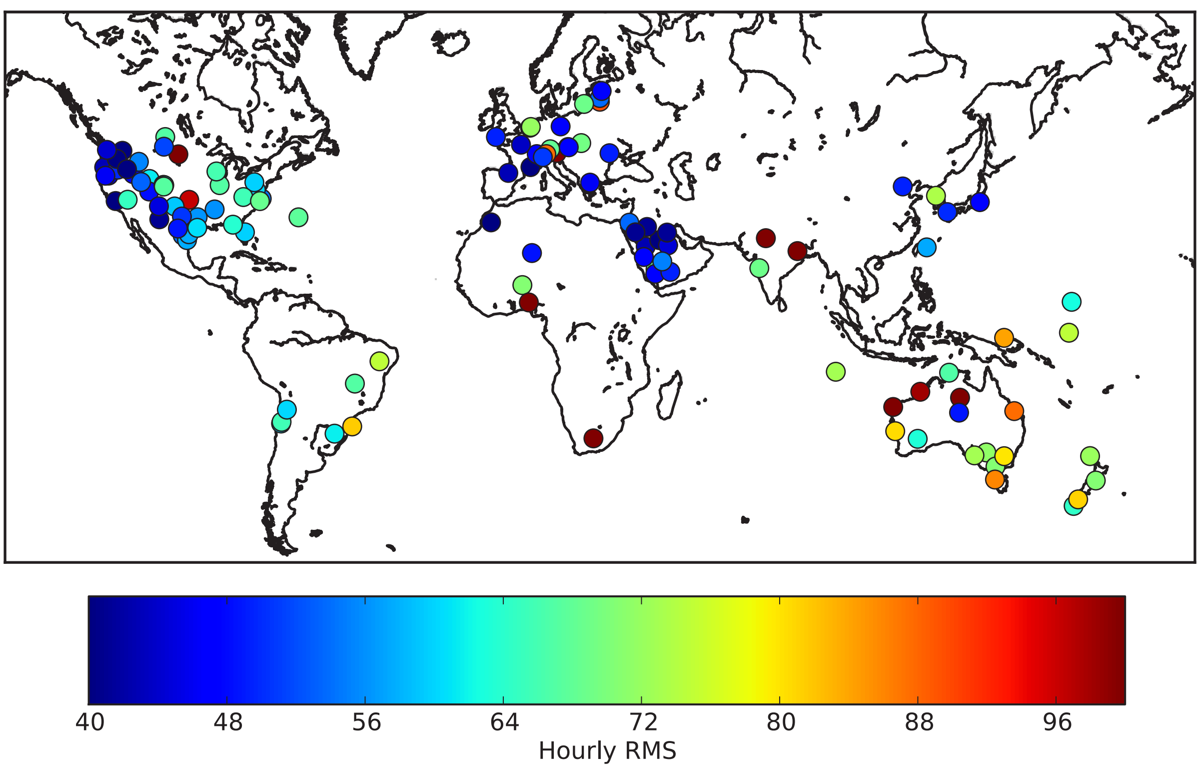
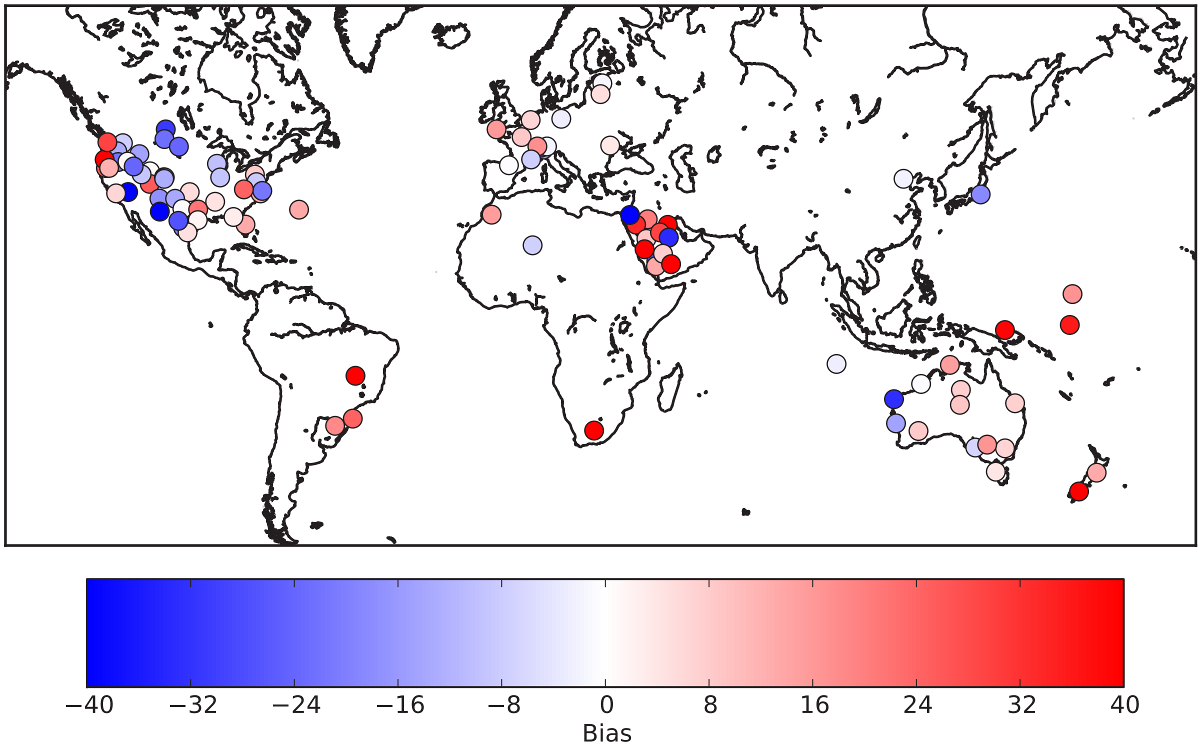



*Figure 3.* Global maps showing mean bias and RMSE for DNI.



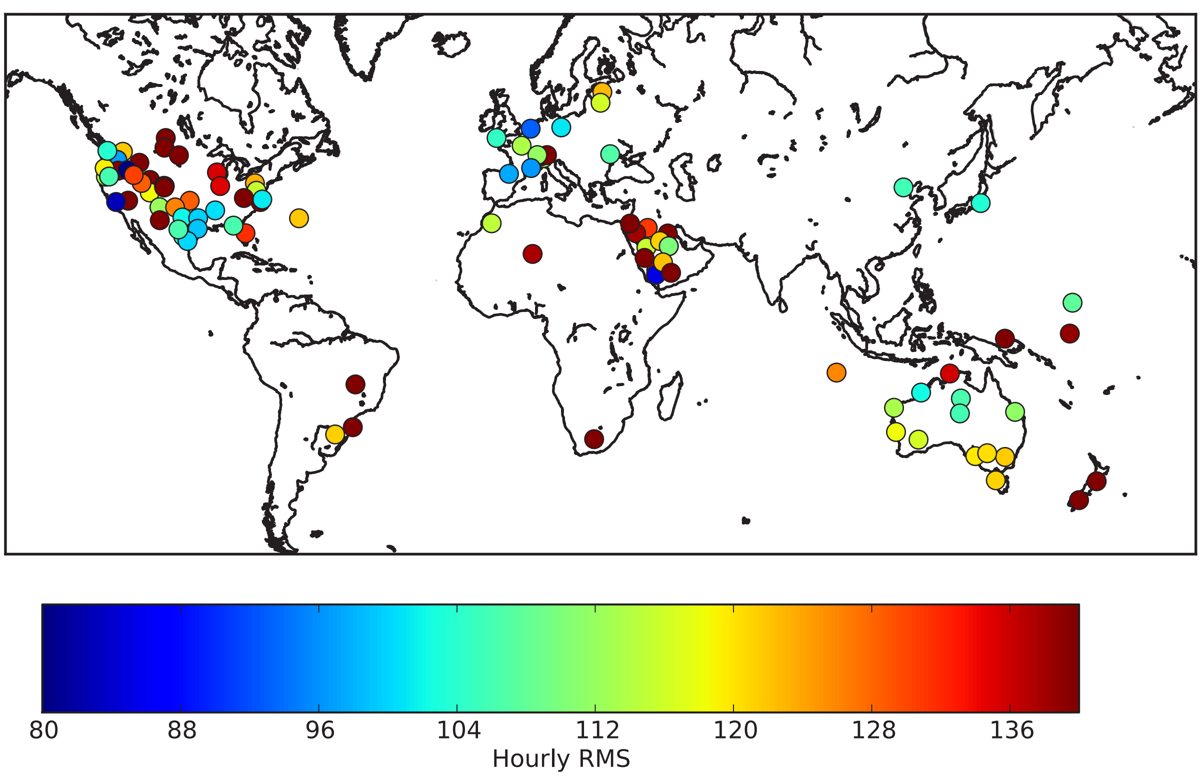
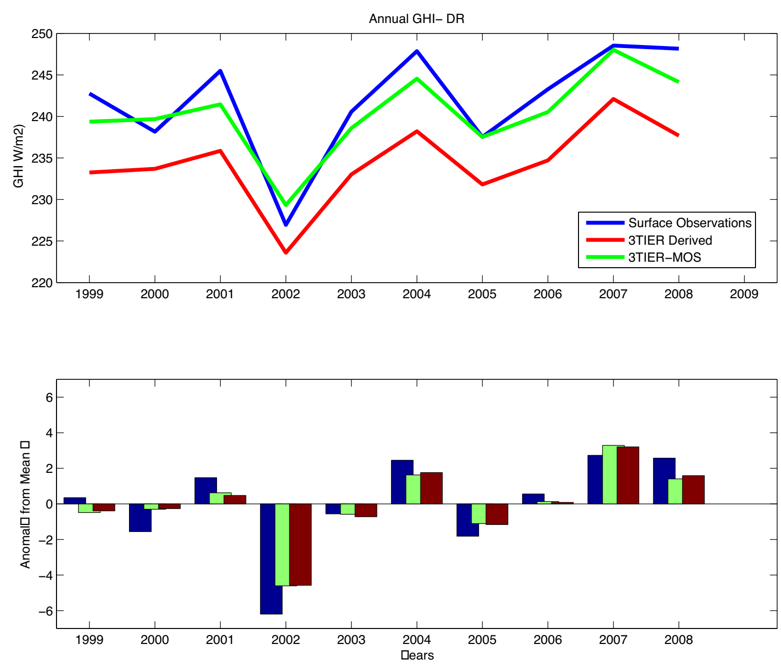



*Table 1.* Validation statistics globally and by the regions shown in Figure 3. MBE is Mean Bias Error, MAE is the mean absolute error, STDEV is the standard deviation of the errors. Median is the median of the error, Hourly RMSE is the Hourly Root Mean Square Error, and N is the number of stations used in the analysis.

| | Global | AME | ASA | S.AM | N.AM | EUR | IND |
|---|---|---|---|---|---|---|---|
| **GHI** | | | | | | | |
| MBE | 0.9% | -0.1% | 1.1% | -1.1% | 0.9% | 1.7% | 3.7% |
| MAE | 3.8% | 4.2% | 3.9% | 4.4% | 3.8% | 2.7% | 4.5% |
| STDEV | 5.0% | 5.8% | 5.4% | 5.2% | 4.9% | 3.6% | 4.2% |
| Median | 0.6% | -0.9% | 0.6% | 1.3% | 1.0% | 0.7% | 3.1% |
| Hourly RMSE | 24.1% | 15.4% | 26.6% | 21% | 22.6% | 31.3% | 29.6% |
| N | 120 | 18 | 26 | 7 | 44 | 22 | 3 |
| **DNI** | | | | | | | |
| MBE | 0.7% | 4.0% | 3.1% | 13.4% | -2.5% | 1.0% | NA |
| MAE | 8.6% | 11.9% | 6.7% | 13.4% | 8.7% | 5.9% | NA |
| STDEV | 11.6% | 13.5% | 8.1% | 3.7% | 12.4% | 7.4% | NA |
| Median | 2.5% | 5.8% | 3.0% | 15.4% | -2.1% | -0.3% | NA |
| Hourly RMSE | 43.1% | 36.8% | 40.7% | 50.6% | 43.1% | 52.8% | NA |
| N | 96 | 16 | 20 | 3 | 44 | 13 | NA |

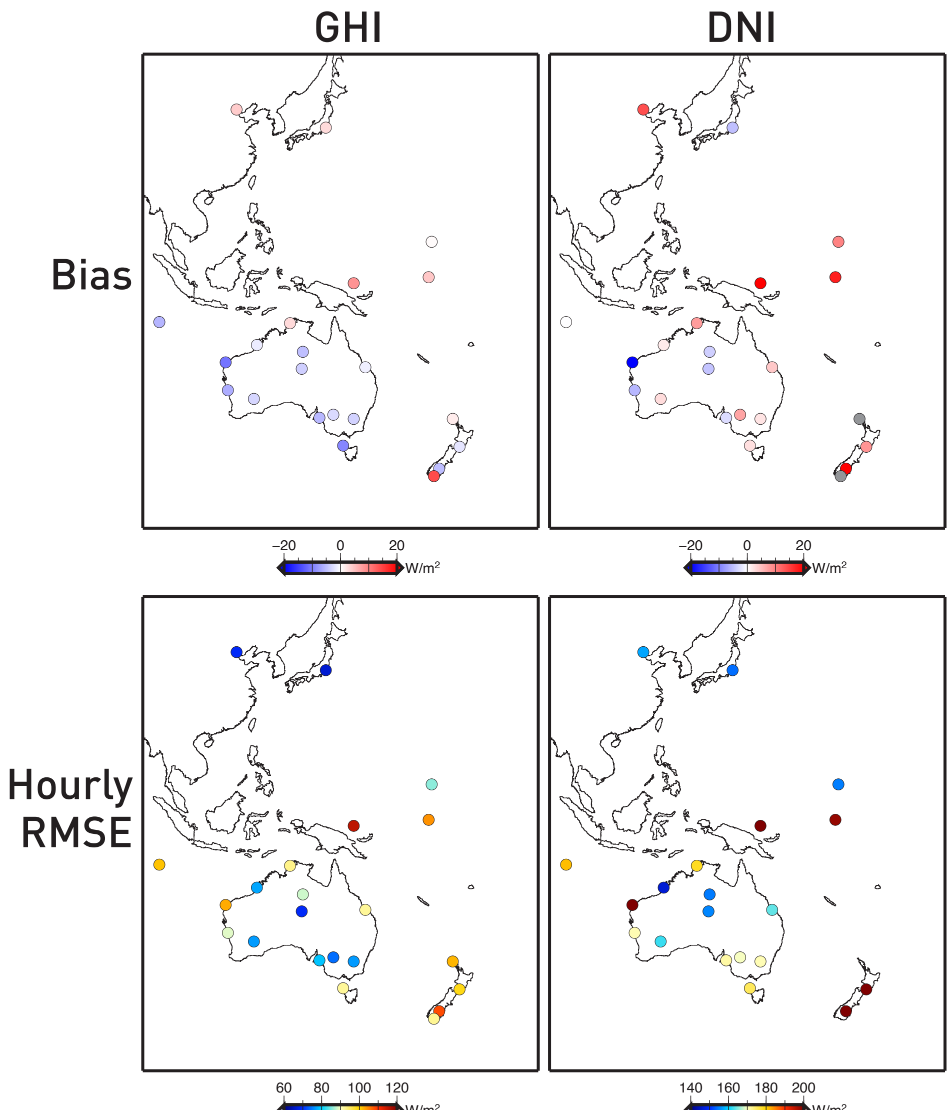
*Figure 4.* Annual observed, 3TIER derived, and 3TIER MOS GHI values for a 10-year period (1999-2008) at Desert Rock, Nevada

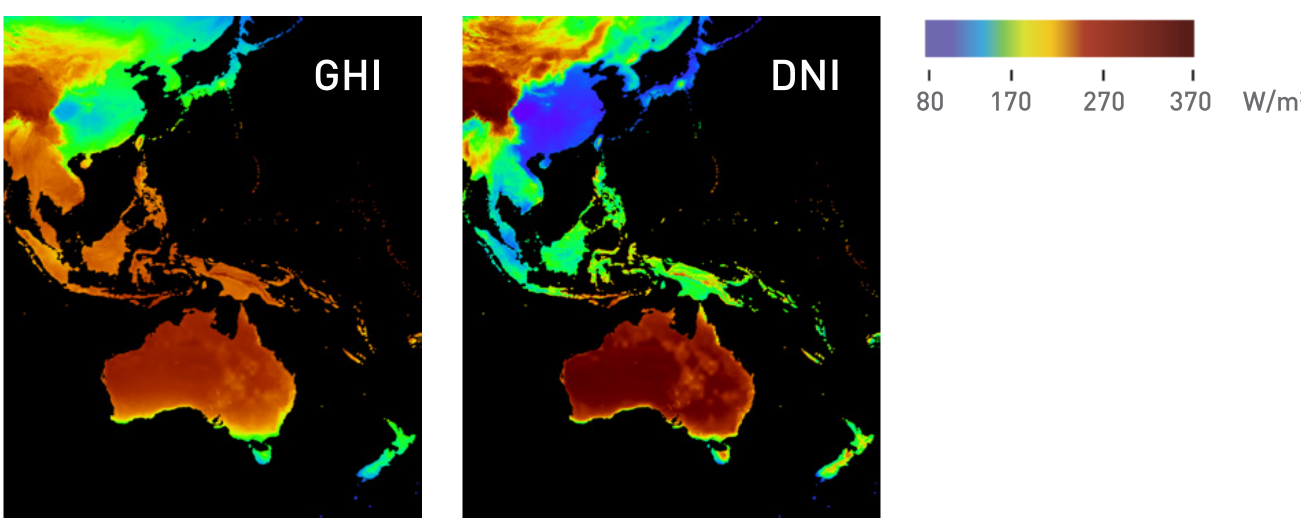


# Appendix: Regional Validations

### East Asia/Oceania (ASA)

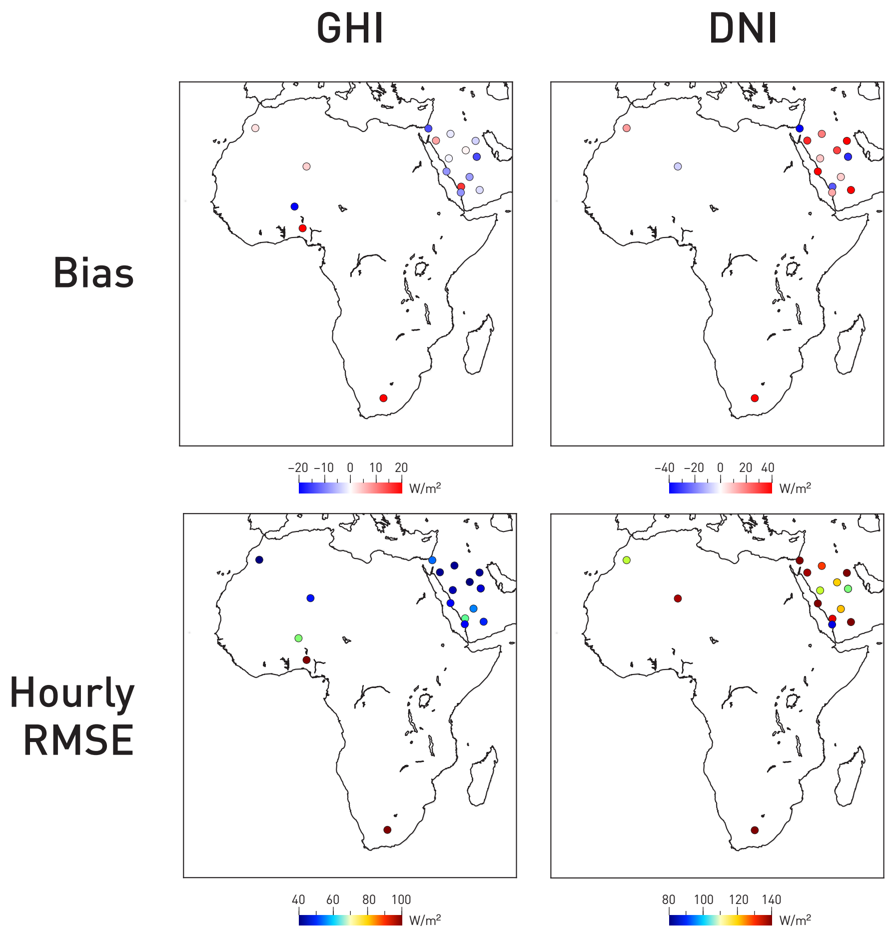
Statistics for the ASA region were derived using 13 sites in Australia, 4 in New Zealand, 4 located on islands in the Pacific, 2 in Japan, 1 in Korea, 1 in Taiwan, and 1 in China. Mean bias and hourly RMSE values for each of these sites are displayed below and the numeric averages are shown in Table 1. On average, mean bias errors for GHI and DNI are less than 4%. There is a weak trend of under prediction of GHI over the interior of Australia and over prediction to the north. Hourly RMSE values for the ASA region are consistent with global validation statistics, averaging to 27% for GHI and 41% for DNI.

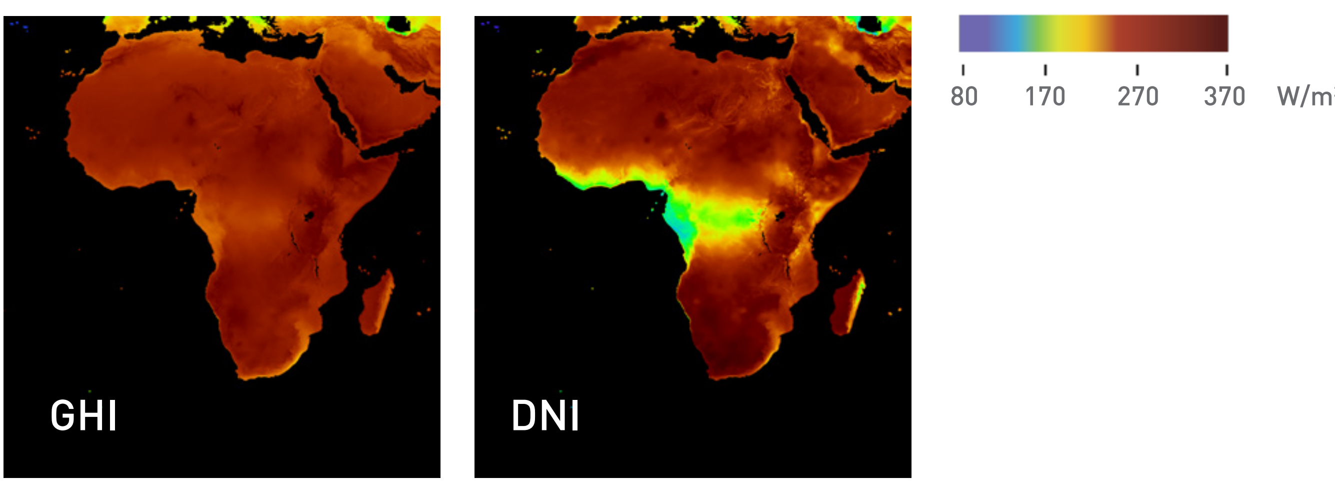


## Africa/Middle East (AME)

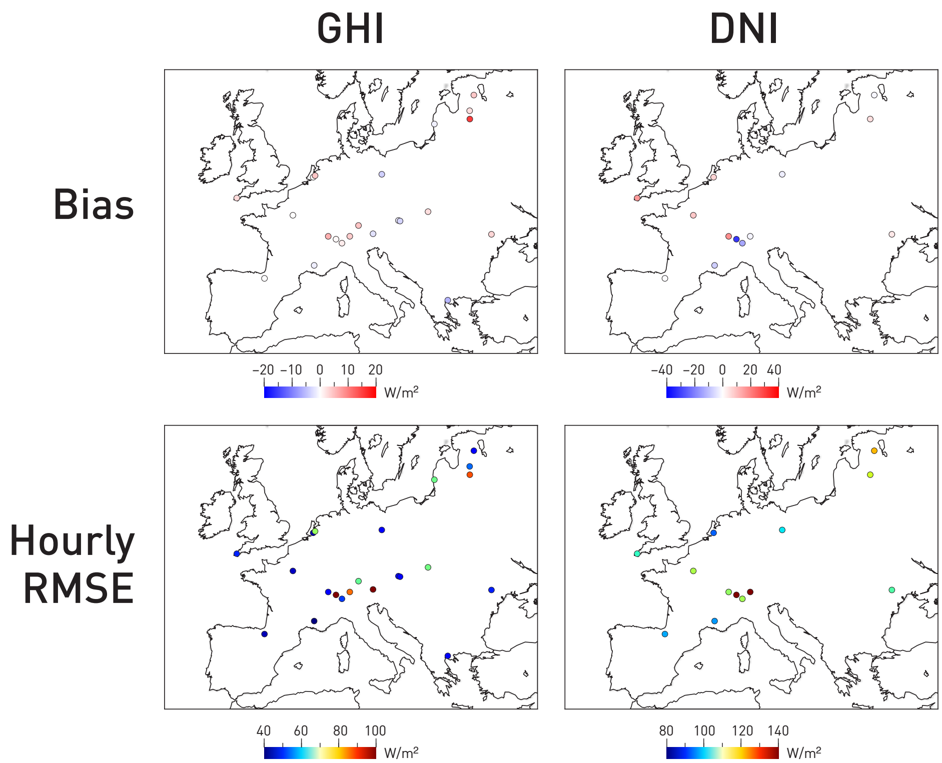
For Africa and the Middle East region (AME), 18 observational sites were available for validation, 3 in Africa and 15 over the Middle East. For the 18 sites mean bias and hourly RMSE were consistent with validation statistics globally with mean bias errors of less than 1% for GHI and 4% for DNI as shown in Table 1. The hourly RMSE values for GHI and DNI are 15% and 37% respectively.

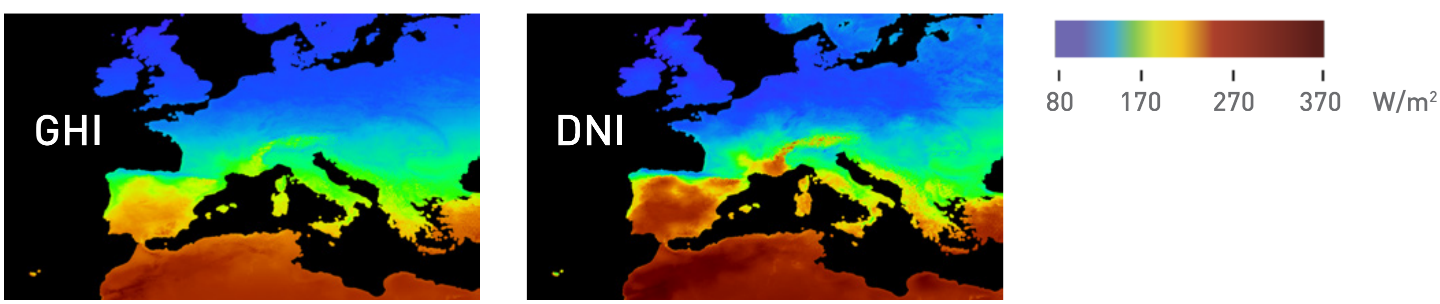


**Europe (EUR)**

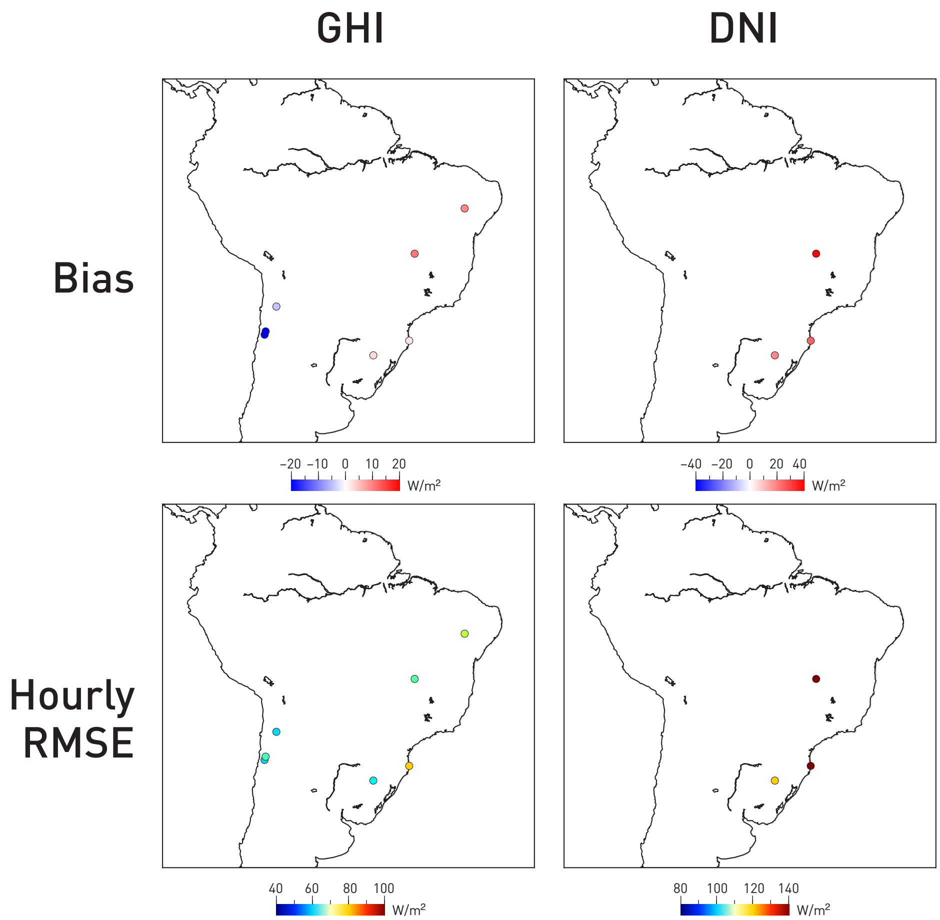
Statistics for Europe (EUR) were derived using 22 sites across the area.  Mean bias and hourly RMSE values for each of these sites are displayed in the figure below and numeric averages are shown in Table 1. On average, mean bias errors for GHI and DNI are less than 2%. Hourly RMSE for the Europe region are consistent with global validation statistics, averaging to 23% for GHI and 53% for DNI.

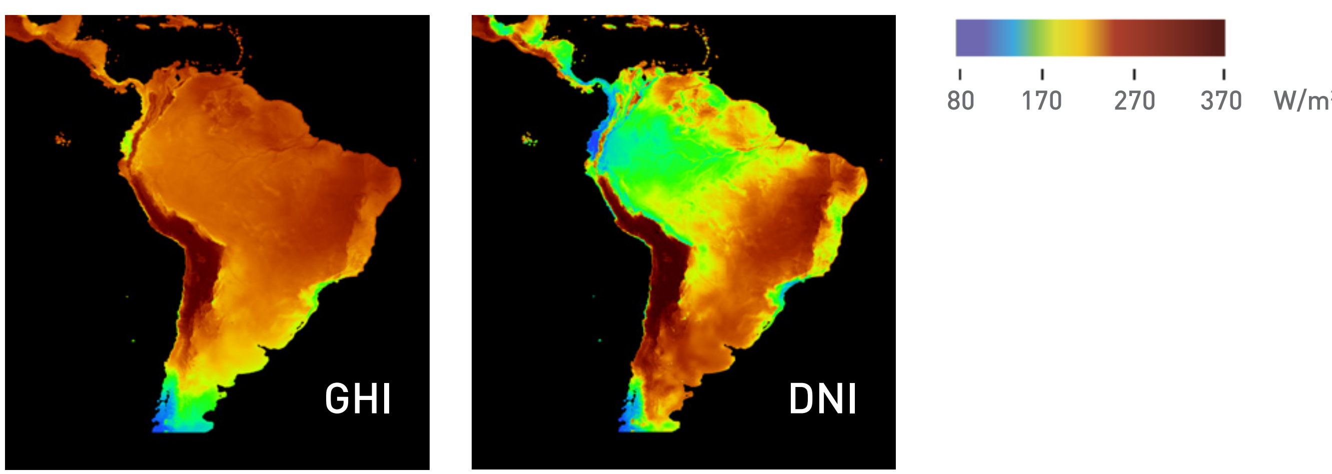


**South America (S.AM)**

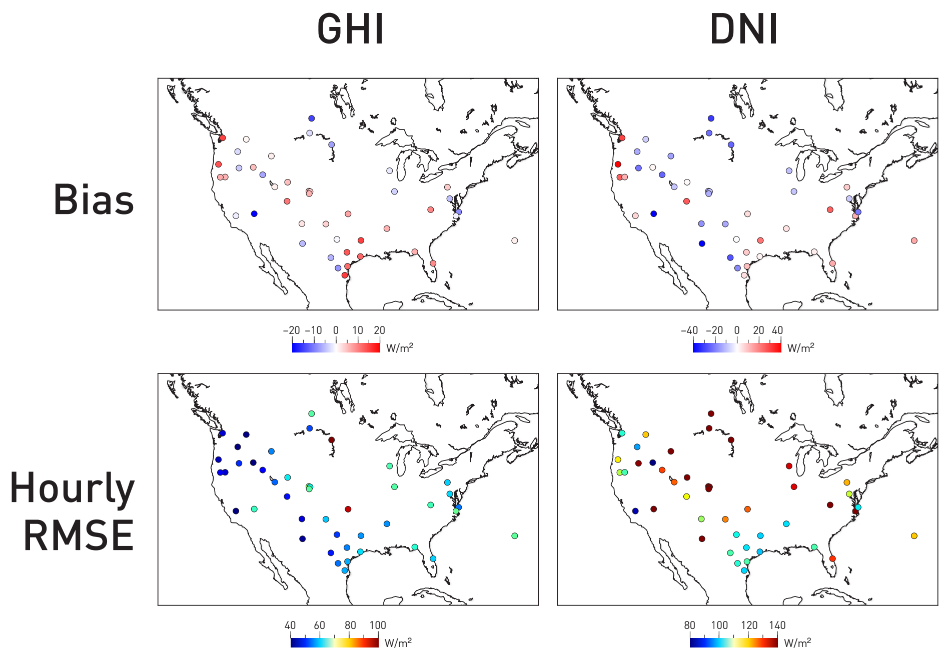
Statistics for South America (S.AM) were derived using 6 sites in Brazil and Chile. Mean bias and hourly RMSE values for each of these sites are displayed in the figure below and numeric averages are shown in Table 1. On average, mean bias errors for GHI were 3% with Hourly GHI RMSE of around 21%. The DNI errors were larger with a MBE of 13% for 3 sites. The observed data is of questionable quality, which may be contributing to the larger errors seen in this region.

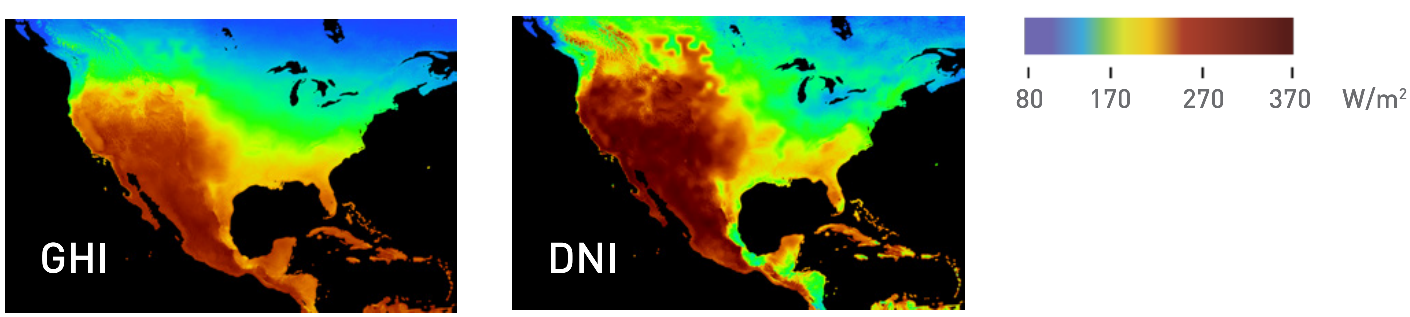


**North America (N.AM)**

Statistics for North America (N.AM) were derived using 44 sites across the area. Mean bias and hourly RMSE values for each of these sites are displayed in the figure below and numeric averages are shown in Table 1. On average, mean bias errors for GHI and DNI are less than 3%. Hourly RMSE values for the N.AM region are consistent with global validation statistics, averaging to 23% for GHI and 43% for DNI.



**India (IND)**

For validation over India, the Indian Meteorological Department (IMD) provided observed climatological values of GHI and DIF for 3 stations throughout the subcontinent. Presently, accurate DNI measurements from the IMD were not available. For GHI mean bias errors were less than 4%. These errors are consistent with the validation numbers across the globe for the 3TIER dataset.