QUALITY ANALYSIS OF GLOBAL HORIZONTAL IRRADIANCE DATA FROM 3500 U.S. GROUND-BASED WEATHER STATIONS

James Hall
JHtech
PO Box 877
Divide, CO 80814

Jeffrey Hall
JHtech
PO Box 877
Divide, CO 80814

ABSTRACT

Currently the US has over 3500 ground sites that measure solar radiation and make their hourly and daily observations publicly accessible. These sites are professionally maintained by universities and government agencies for specific purposes such as agriculture, water management and environmental monitoring. Wider use of this resource by the renewable energy community has been limited by a general lack of knowledge, disparate data formats and concerns about accuracy. This paper compares the accuracy of the solar radiation observations from these networks to other commonly used data sources, describes our work to aggregate this data into a single nationwide solar energy resource with appropriate quality control, and demonstrates the utility of the data in a solar forecasting application.

1. INTRODUCTION

For the past decade solar radiation data from NASA satellites and the datasets from the National Renewable Energy Laboratory (NREL) have been the mainstays in the design of solar power systems.

Meanwhile many universities and government agencies have built their own solar radiation networks for specific purposes such as agriculture, water management and environmental monitoring. Examples include CIMIS (California Irrigation Management Information System) and the Oklahoma Mesonet. Some networks cover only a portion of a state; others have 10-20 sites spread across the entire US. Most overlap, so it is common for one state to have sites from several networks. By itself, a single network may not provide a significant design resource. However in aggregate there are over 3500 such sites across the nation (Fig. 1). Mapping a 75km (47 mile) radius around the each site, we see nearly every location in the US is relatively close to one or more of these sites.(Fig. 2).

One concern about the data from these sites is the accuracy of the solar radiation measurements. A few sites have high quality instrumentation accurate within ±2%, but typically they use mid-quality transducers with about ±5% accuracy. Maintenance issues and data transmission errors can also increase the inaccuracies. However before dismissing these mid-quality sites as not sufficiently accurate or reliable for solar design, we should compare their accuracy to the datasets commonly used in solar engineering.

2. COMPARISON I

The accuracy of a dataset can be determined by comparing observations to a highly accurate reference. Even small differences in location can affect the amount of solar radiation on the ground, especially for short time intervals, so comparisons should be done at exactly the same location. This can prove challenging for solar radiation observations where test sites may be nearby, but rarely co-located with the reference. In addition, high accuracy reference measurements of solar radiation are not widely available. These constraints make absolute comparisons between solar radiation datasets difficult, but it is still possible to estimate the relative accuracy of various datasets if all comparisons are made to the same high-quality reference.

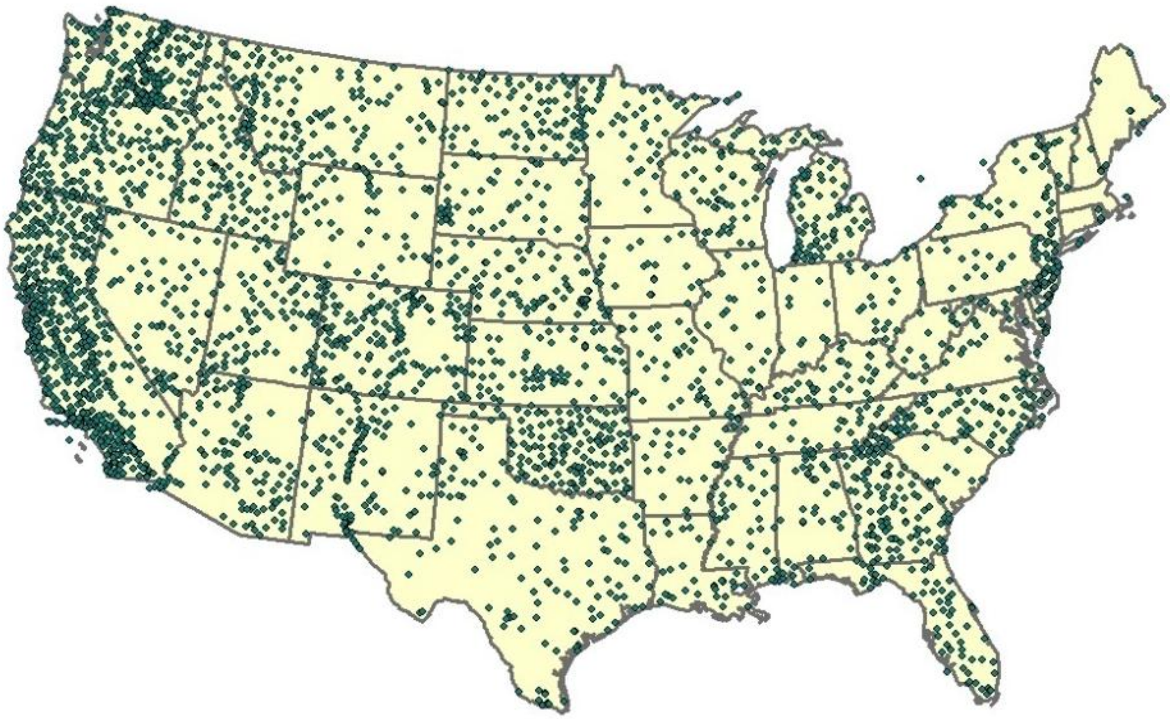

Fig. 1: Over 3500 medium-quality sites measure solar radiation across the US.

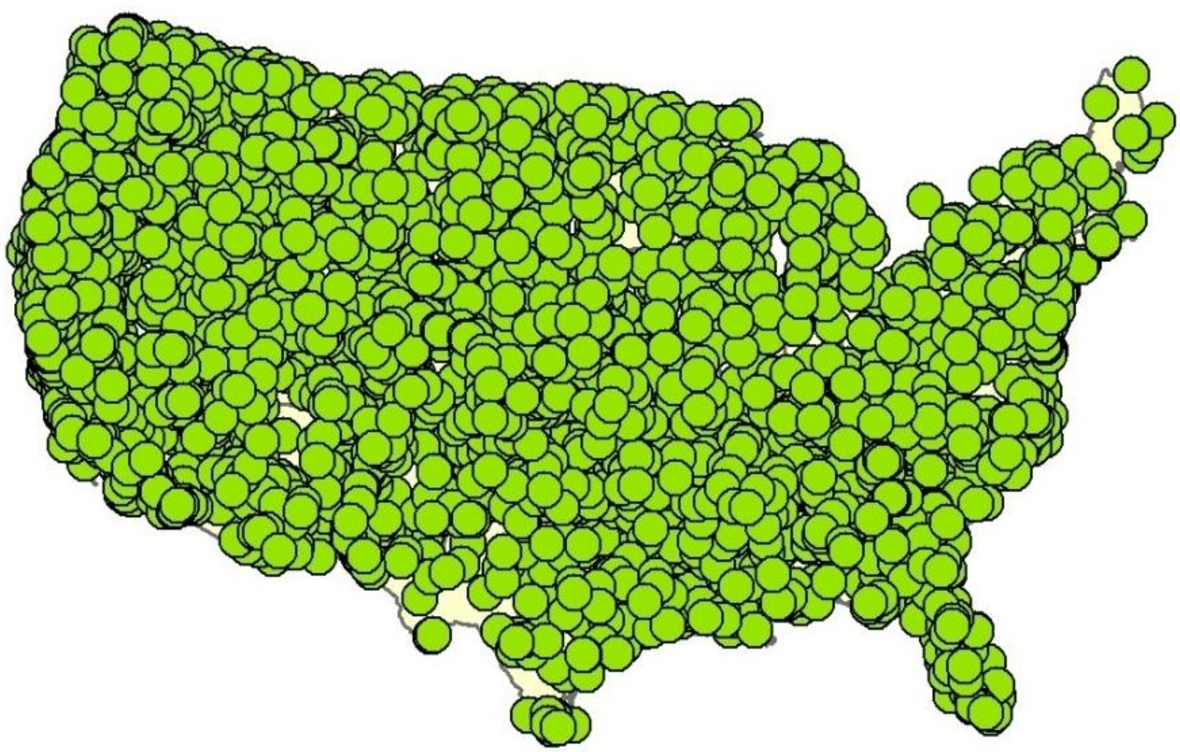

Fig. 2: Nearly every location in the US is within 75km of a medium-quality site; often data from several nearby sites are available.

To this end, we obtained data from 13 high-quality reference sites scattered across the country that were used in the 2003-2005 National Solar Radiation Database (NSRDB). The locations of these reference sites along with a link to an interactive map can be found in Table 1 of the Appendix. Sixteen medium-quality stations were found within 40 km and 200 m elevation of these reference sites, and data for the daily averages of GHI (global horizontal insolation) were compared. This resulted in 13,242 direct comparisons of daily solar radiation between the medium-quality sites and the high-quality reference stations.

Relative to the nearby high-quality sites, the average total error in the daily GHI from 16 medium-quality stations was 9.8% rMAE (relative mean absolute error). This total error includes both the bias and precision errors (Fig. 3).

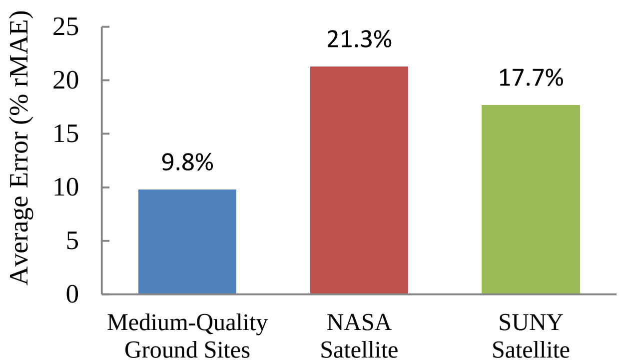

Fig 3: Average Total Error in Daily GHI

Since measurements from the reference sites also contain some error, the true accuracy of the medium-quality sites may be somewhat lower, but this technique provides a useful comparison.

How do these errors compare to satellite-based solar radiation datasets? Using the same data from the NRSDB high-quality stations as a reference, we repeated this comparison for corresponding measurements of global horizontal insolation from two sources of satellite-based data: SUNY[1] (State University of New York) and NASA Agro-Climate data[2]. The average total error for the SUNY dataset relative to the 13 reference sites was 17.7% rMAE. The average total error for the NASA dataset relative to the 13 reference was 21.3% rMAE (Fig. 3). These percentages include both bias and precision errors.

3. COMPARISON II

A second comparison used research-quality solar measurements from USCRN (US Climate Reference Network) as the reference. Seven of the earliest USCRN sites were operating in the southern half of the US during 2002-2005, a period of time that overlaps with NREL's TMY3 (Typical Meteorological Year) dataset. By design, the TMY3 data for a location is composed of monthly observations selected from different years to represent a typical annual climate series. We identified 19 different TMY3 months from seven locations that could be directly compared with nearby USCRN sites. Comparable SUNY and NASA observations were obtained for the 10km grid containing the reference site along with GHI data from nine medium-quality ground sites. Each medium-quality site was located within 36km (20 miles) and 150 m (500ft) elevation of the USCRN reference. Where more than one medium-quality site was available, the observations were averaged. The locations of the sites can be found in the Appendix, Table 2.

The total errors (rMAE) and bias errors (rME or relative mean error) in the daily observations of GHI are shown in figure 4. The NASA observations had the highest total error (27%), TMY3 and SUNY had similar errors (19%) and the medium-quality ground measurements had the lowest error (9%). Bias errors for the datasets are also shown in figure 4.

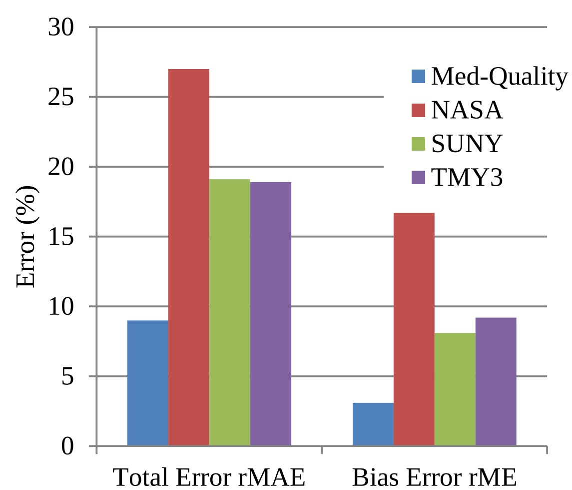

Fig 4: Total Errors and Bias Errors in 570 Daily GHI Observations Compared to USCRN Reference

Leaving out TMY3, a more comprehensive comparison can be made between the NASA, SUNY and ground-based data. Using the same seven USCRN sites as the reference, we compared all available daily observations from 2002-2005; a total of 7770 comparisons. The graph in Figure 5 shows that the daily observations from the medium-quality sites had less than half the total error (8.8%) of the NASA and SUNY observations (21.3% and 18.9%). The respective bias errors are also shown on the graph.

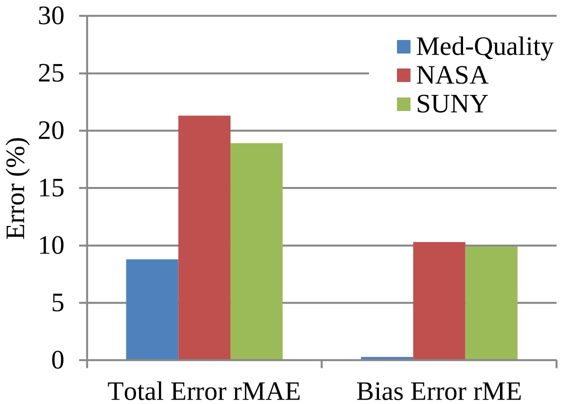

Fig 5: Total Errors and Bias Errors in 7770 Daily GHI Observations Compared to USCRN Reference

These results are similar to other published comparisons. NASA estimates that their measurements of daily solar radiation have a root mean square error of 35 W/m^2 (roughly 20% total error) [3]. Other researchers found 19% total error in the NASA daily observations[4].

Within the USCRN network there are several pairs of stations in close proximity. This provides an opportunity to determine if the apparent differences between two nearby ground measurements are due to sensor accuracy or because the sites are not exactly co-located. This comparison used data from the paired USCRN stations in Lincoln NE, Newton GA, Stillwater OK, and Asheville NC. The separation distances between the pairs were 29, 10, 2, and 10 km respectively (18, 6, 1.5, 6 miles). Detailed location information is found in the Appendix, Table 3.

The GHI sensors used by the USCRN sites are rated at <1% non-linearity and ±2% stability per year; however Figure 6 shows relatively high errors between the high-quality sites in close proximity. This suggests that the calculated errors in the ground-based observations of GHI may be more influenced by physical separation than by the difference between high-quality and medium-quality sensors.

Overall, these findings indicate that the sources of solar radiation data commonly used in solar engineering contain about twice the average total error found in the data from medium-quality ground observation networks.

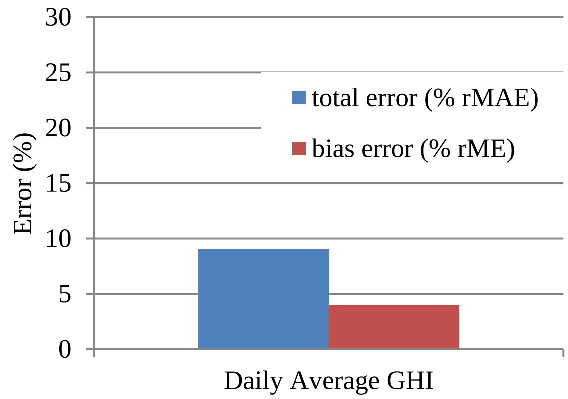

Fig 6: Error in Daily GHI Observations between Closely Located USCRN Sites

4. DATA ACCESS AND QUALITY CONTROL

Past experience has shown that accessing and aggregating the data from these networks of medium-quality solar radiation observations can be challenging. Typically there are multiple networks in a region, and much of the value comes from integrating all of the regional data into a coherent dataset. However each network has its own measurement units, data format and delivery method. Nearly every week one or more of the networks change their data format or the method of accessing the data. Since most networks are operated for specific private purposes, bulk downloads and ease of public access are generally not priorities for the network operators.

Only a few of the networks provide quality control of their data. However by gathering the data from many networks across the entire US over a long time period, our experience has shown that several statistical techniques can often identify data that might be questionable. One level of quality control compares each observation to a long-term historical database for the same region and day of the year. A second level of quality control compares each observation to those from other stations in the area at similar elevations.

5. SAMPLE APPLICATION

One demonstration of the utility of medium-quality solar radiation data is a pilot study[5] to forecast solar radiation at Fontana, CA. Previous forecasting attempts have been hampered by a perceived lack of near-real time data for this region. Currently there are over 160 medium-quality sites in the greater Los Angeles area reporting hourly solar radiation. This large number of observation sites provides a significant advantage in any forecasting attempt.

The pilot forecasting study utilized solar radiation data from mid-quality sites near Fontana, CA as well as meteorological observations from regional airports and the National Digital Forecast Database. The model was composed of two adaptive components: one predicted solar radiation based on meteorological observations and the second forecast solar radiation based on seasonal pattern recognition. Results from the model are shown in Fig. 7. The forecasting results from this pilot study were significantly better than previously published results in true out-of-sample testing[5].

6. CONCLUSIONS

Higher accuracy, nationwide coverage and the demonstrated utility of the data in solving real-world problems should encourage practitioners in the renewable energy sector to take a careful look at these networks of medium-quality solar radiation measurement sites.

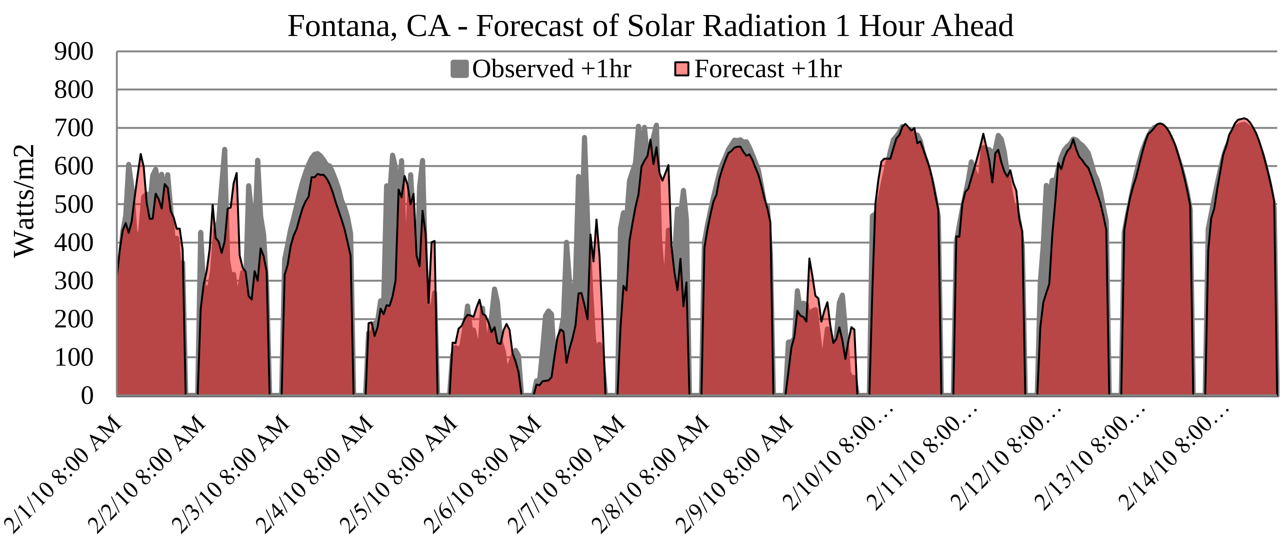

Fig 7: Sample of GHI Forecasts Utilizing Data from Medium-Quality Sites in the Los Angeles Basin

7. REFERENCES

(1) SUNY gridded data can be accessed at: ftp://ftp.ncdc.noaa.gov/pub/data/nsrdb-solar

(2) NASA Agro-Climate data can be accessed at: http://power.larc.nasa.gov/

(3) NASA Surface meteorology and Solar Energy (SSE) Release 6.0 Methodology, 2010, http://power.larc.nasa.gov/documents/SSE_Methodology.pdf

(4) Yang, et. al., Usability of NASA Satellite Imagery-Based Daily Solar Radiation for Crop Yield Simulation and Management Decisions. 2010, http://adsabs.harvard.edu/abs/2007AGUFMPA33A1031Y

(5) James Hall, Forecasting Solar Power with Adaptive Models – A Pilot Study, available online at: http://www.solardatawarehouse.com/forecasting%20solar%20power.pdf

8. APPENDIX

TABLE 1: LOCATIONS OF THE NSRDB SITES USED IN COMPARISON I

National Solar Radiation Database High-Quality Station	Medium Quality Station	Separation distance (km)	Elevation Difference (m)
Bismark Airport, ND	ND059-1 (46.77, -100.92)	11.4	90
Spokane Intl Airport, WA	WA063-1 (47.41, -117.53)	9.3	97
Wolf Point Intl, MT	MT085-1 (48.12, -105.08)	20.5	105
Hermiston, OR	OR059-3 (45.82, -119.53)	11.5	52
Burns Airport, OR	OR025-10 (43.51, -119.3)	22	70
Boise Air Terminal, ID	ID001-1 (43.6, -116.18)	2.3	128
Dillon Airport, MT	MT003-1 (45.56, -107.44)	17.1	66
Medford Intl Airport, OR	OR029-5 (42.33, -122.94)	25.1	186
University of Illinois	IL019-1 (40.05, -88.37)	12.1	6
University of Illinois	IL019-2 (40.08, -88.23)	1.1	0
Bluefield/Mercer, WV	VA197-1 (37.01, -81.18)	29.4	29
Bluefield/Mercer, WV	WV089-2 (37.53, -81)	35.8	28
Hanford Airport, CA	CA031-1 (36.49, -119.78)	26	14
Ponca City Airport, OK	OK047-1 (36.41, -97.7)	26.4	14
Ponca City Airport, OK	OK047-2 (36.41, -97.7)	28.6	34
Broomfield/Jeffco, CO	CO123-4 (40, -104.85)	36.2	148

An interactive map showing the location of these sites and the surrounding terrain can be found at:
http://maps.google.com/maps/ms?ie=UTF8&hl=en&msa=0&msid=117745141622161862329.0004642e9c9f5254c656c&ll=40.380028,-95.097656&spn=31.79519,85.78125&t=p&z=4

TABLE 2: LOCATIONS OF THE USCRN SITES USED IN COMPARISON II.

USCRN Reference Station	Medium-Quality Ground Station	TMY3
IL Champaign 9SW (40.0526, -88.3729)	IL019-1 (40.05, -88.37) IL019-2 (40.08, -88.23)	U of I Bondville Surfrad (40.06 -88.37)
OK Stillwater 2W (36.1181, -97.0914)	OK119-4 (36.12, -97.1)	Stillwater Regional AP (36.15, -97.083)
GA Newton 8W (31.33127, -84.4706)	GA Newton 11 SW (31.1923, -84.4465) note: USCRN site, other sites not available 2003-2005	Albany Dougherty AP (31.533, -84.183)
AZ Elgin 5S (31.5907, -110.509)	AZ019-1 (31.78, -110.64)	Davis Monthan AFB (32.167, -110.83)
NE Lincoln 8ENE (40.8484, -96.5651)	NE109-2 (40.85, -96.6) NE109-8 (40.78, -96.6)	Lincoln Municipal AP (40.833, -96.767)
CO Nunn 7NNE (40.8066, -104.755)	CO123-9 (40.87, -104.73)	Cheyenne Municipal AP (41.15, -104.8)
LA Monroe 26N (32.8833, -92.1165)	LA073-1 (32.51, -92.35)	Monroe Regional AP (32.517, -92.033)

TABLE 3: LOCATIONS OF THE PAIRED USCRN SITES USED IN COMPARISON II

USCRN Reference Station	USCRN Paired Station
NC Ashville 8SSW (35.4945, -82.6142)	NC Ashville 13S (35.4185, -82.5567)
OK Stillwater 2W (36.1181, -97.0914)	OK Stillwater 5WNW (36.1346, -97.1082)
GA Newton 8W (31.3127, -84.4706)	GA Newton 11 SW (31.1923, -84.4465)
NE Lincoln 8ENE (40.8484, -96.5651)	NE Lincoln 11SW (40.6954, -96.8541)