Improved model of solar resource variability based on regional aggregation and climate zones

Gregory M. Kimball[1], Chetan Chaudhari[1], Patrick Keelin[2], John Dise[2], Mark Grammatico[2], and Ben Bourne[1]
[1]SunPower Corporation, Richmond, CA 94804, USA
[2]Clean Power Research, Napa, CA 94559, USA

***Abstract*—Solar resource measurements play a critical role in the assessment of long-term energy yield and the valuation of PV systems. Recent valuation methods have focused on the probability of exceedance statistic (PXX) for annual insolation as part of computing project risk. For example, P90 is the annual insolation value exceeded 90% of the time. However, the small sample size of annual insolation values for a given location increases uncertainty in the distribution. To assess the distribution of annual insolation values, we aggregate ground- and satellite-based data for the continental United States from 1961-2017, aggregate the data regionally and by climate zone, and report variability statistics for each location. Overall, the P99 annual insolation values were found to range from -4 to -8% of the P50 value across the continental United States.**

I. Introduction

Reducing uncertainty in interannual variability of solar insolation is a key challenge for predicting the long-term energy yield of PV systems [1]. Most satellite-based insolation models report data from periods of less than 20 years [2] and only a few ground-based sensor systems have been operating for longer periods up to 40 years [3]. Even based on 8 years of data, detailed maps of the coefficient of variance for annual insolation have been published for the continental United States [2]. Other reports have generated "synthetic" years comprised of 3-month periods from a 19 year dataset to improve the sampling of interannual variability [4]. In this study, we aggregate ground-based [3] and satellite-based [5] data in a 200-km radius and from the years 1961-2017 to increase the sample size of annual insolation data and reduce the uncertainty in interannual variability.

II. Data and Methods

Annual insolation data across the United States were compiled from several sources. First, data of NSRDB volume 1 were retrieved from the period 1961-1990 for 239 primary and secondary WBAN stations [3]. Second, data from the NSRDB 1991-2010 Update were included for 1020 USAF site locations using the METSTAT model [5]. Third, data from the SolarAnywhere 3.2 model from Clean Power Research were included from 1998-2017 for 44 WBAN station locations [6]. Hourly averaged global horizontal irradiance (GHI) data were used to generate annual insolation values. Between the three data sources, long-term median insolation values showed correlation coefficients of >0.98.

To aggregate regional insolation data, all locations in a 200-km radius were compiled and those within the same Köppen-Geiger climate zone were included in the regional aggregate. Each location's annual insolation time series data was scaled to the location's median, which was determined by fitting a normal distribution to the data. The scaled annual insolation data was then aggregated to emphasize the distribution shape rather than differences in the median values within the region.

III. Results

Figure 1 shows the results of a statistical study of estimating median and standard deviation values for a location based on interannual data. Estimating the median value of annual insolation requires a smaller sample size than estimating the variability in annual insolation. A fixed number of points were sampled from a normal distribution with location parameter

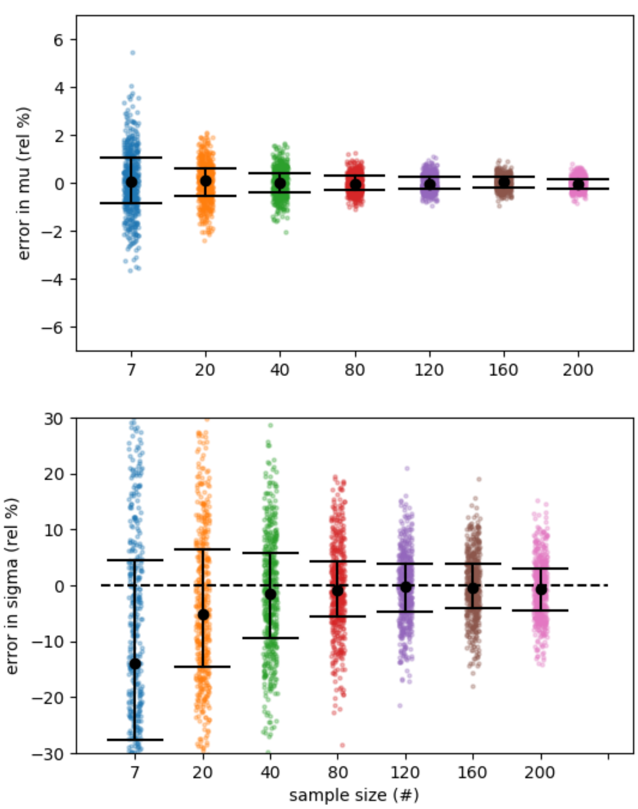

Fig. 1: Statistical study based on random sampling of the normal distribution demonstrating that 10 points are sufficient for estimating the median with 1%, but that >100 points are needed to estimate the standard deviation to within 5%.

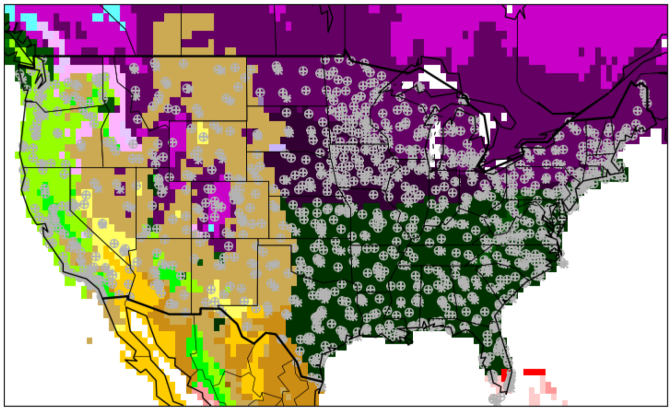

Fig. 2: (a) Köppen-Geiger climate map of the United States with overlaid WBAN, USAF-METSTAT and CPR-SolarAnywhere analysis locations.

1600 and scale factor 60, where the values were chosen based typical solar resource variability data in the United States. The sampling procedure was completed 500 times, resulting in estimates of median resource with an interquartile range of $\pm 1\%$ for 7 points and estimates of standard deviation with an interquartile range of $\pm 10\%$ for 40 points. Greater than 100 points of data were needed to reduce the interquartile range of standard deviation values to $\pm 5\%$.

Figure 2 shows the climate regions of the continental United states along with the NSRDB's WBAN and USAF locations. Although the continuous climate regions cover broad areas in the eastern United States, south Florida, the Rocky Mountains and the west coast show more spatial nonuniformity. Station coverage is denser in the eastern United States than in the mountain west. CPR-SolarAnywhere analysis locations were chosen to correspond with the list of WBAN and USAF sites.

Figure 3 shows the full data set aggregated by climate zone. The dominant climate zones in the United States have sample size >5000, but some of the more rare climate zones only have a few hundred observations. Standard deviation values were found to be 2.5-3.0% for the dominant Cfa, Dfa, and Dfb climate types in the eastern United States. Lower standard deviation values of 1.3-1.9% were found for the desert southwest, and the Csb zone common in California exhibited standard deviation of 3.3%.

Figure 4 shows the median and P99 annual insolation values computed for the continental United States. Figure 4a echoes familiar maps of typical solar resource, and Figure 4b shows the P99 probability of exceedance statistic based on normal distribution fits to regionally-aggregated data. Median annual insolation values peak around 2100 kWh/m2/yr in the desert southwest, and show a minimum in the Pacific northwest around 1200 kWh/m2/yr. Overall the P99 probability of exceedance values were found to be -4 to -8% of the P50 value across the continental United States.

IV. Discussion

In comparison with previous reports, the data and methods in this paper show more geographically uniform and somewhat higher interannual variability. Gueymard et al [2] prepared variability maps of the continental United States based on 7 years of satellite data processed using the SUNY model. Figure 5 compares similar maps and color scales between the two reports. We find somewhat higher variability of 2.0-2.5% across the mountain west, and a fairly uniform variability of 3.0-3.5% across the eastern United States. Western Oregon is found to have a large region of high variability up to 5% and previous reports of high variability in south Texas and central New York from previous reports were not reproduced.

Inspection of the data in Figure 3 occasionally suggests some more complex structure than a symmetric normal distribution. To explore this concept, we generated a range of skew-normal distributions to assess how many data points would be required to converge on the correct sigma and shape factor. We found that >500 points would be required to distinguish between normal and skew-normal distributions, a sample size greater than we would have without increasing the regional area to a radius of >400 km. Furthermore, if annual insolation values can be modeled as the sum of many random weather events, we would expect the central limit theorem to drive annual insolation values towards a normal distribution. We also explored aggregating the regional data without scaling the data from each site. This approach yields larger variability results because it conflates variability in median insolation with variability in interannual insolation across a region.

Based on the simulations in Figure 1, an accurate estimate of the P99 annual insolation value for a project location requires at least 5 times more samples than the 20 years available for most locations. Spatial aggregation of data increases the sample size and allows distribution shape to be assessed, but also raises several issues.

One concern is that the annual insolation data may have significant spatio-temporal correlation between nearby sites or nearby years. High spatio-temporal correlation reduces the degrees of freedom in the annual insolation samples and overstates the certainty in distribution statistics. Some authors have reported that annual insolation appears to undergo decade-long trends [7] [8], while others have asserted that annual insolation is independent of previous years [9]. Temporal correlations were evaluated by computing the correlation coefficient for 31 sites between the data and its 1-year lag. The 1-year temporal correlation coefficients were found to be 0.05 ± 0.21, which is comparable to the results found after shuffling the year index (-0.003 ± 0.23). Spatial correlations were evaluated by computing the correlation coefficient for 566 combinations of nearby sites for the same year. The spatial correlation coefficients were found to be 0.67 ± 0.25, which are significantly stronger than the results found after shuffling the year index (-0.006 ± 0.29). The partial correlation between nearby sites suggests that the degrees of freedom may be somewhat less than the sample size for computing confidence

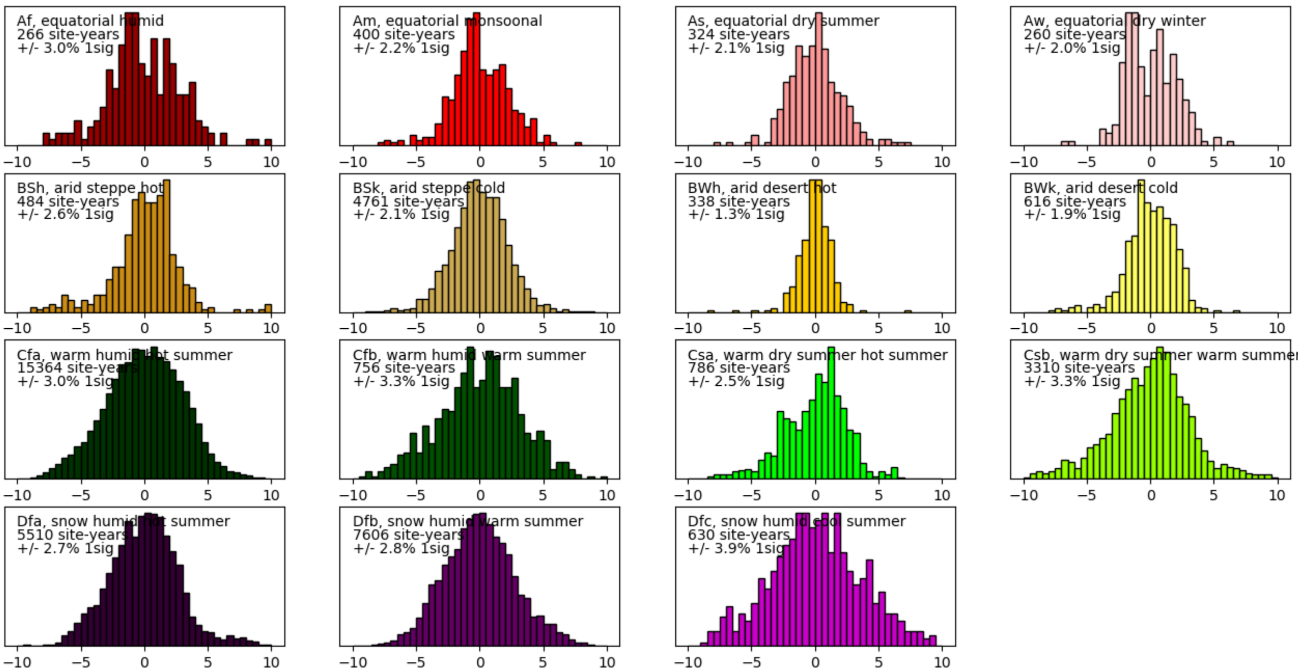

Fig. 3: For each Köppen-Geiger climate zone, the available site-years were compiled into histograms. X-axis labels represent relative percent deviation from the median. One sigma variability values range from 1.3 to 4.0% by climate zone.

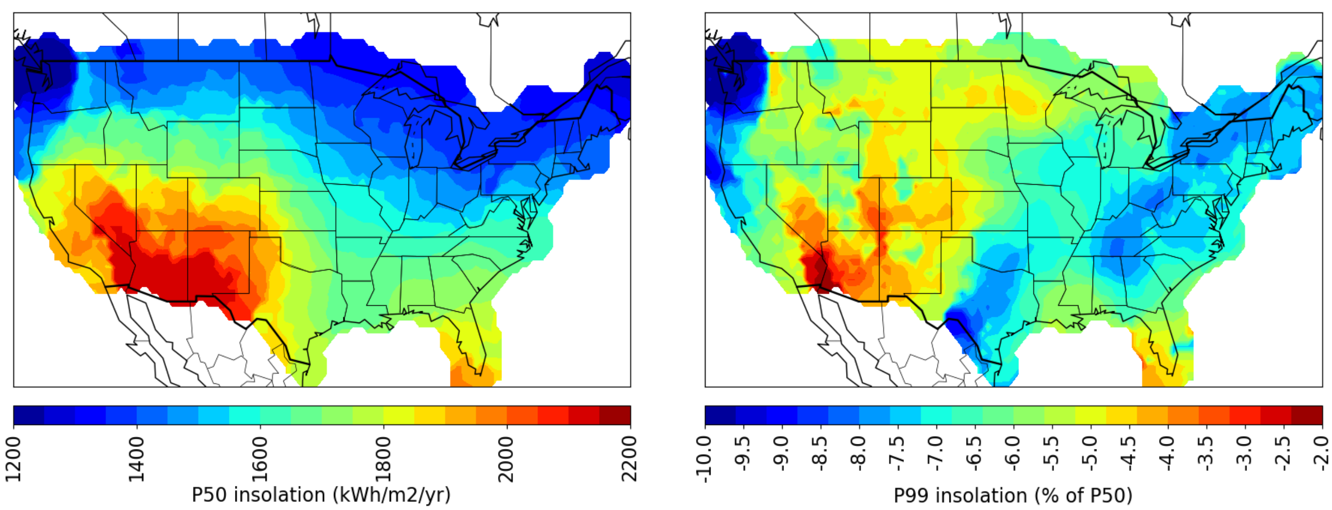

Fig. 4: (a) Regional median annual insolation, aggregated by region and by Köppen-Geiger climate zone. (b) Regional P99 annual insolation data, generated using a normal distribution fit to the aggregated insolation data.

intervals on distribution statistics.

Another important consideration with spatial aggregation is that nearby sites may not share the same distribution of annual insolation. Many satellite-based and numerical weather prediction models provide gridded estimates with 10x10 km resolution, suggesting that the median annual insolation is likely to change for nearby sites, but changes in median insolation do not necessarily require changes in the interannual distribution. The significant spatial correlation of 0.67 ± 0.25 between sites suggests that the distribution is likely to be similar across a region. Selecting nearby sites for aggregation only if they share the same Köppen-Geiger climate zone [10] was performed to mitigate the risk of aggregating across sites with significantly different distributions of interannual insolation.

Variability in annual insolation is an important component

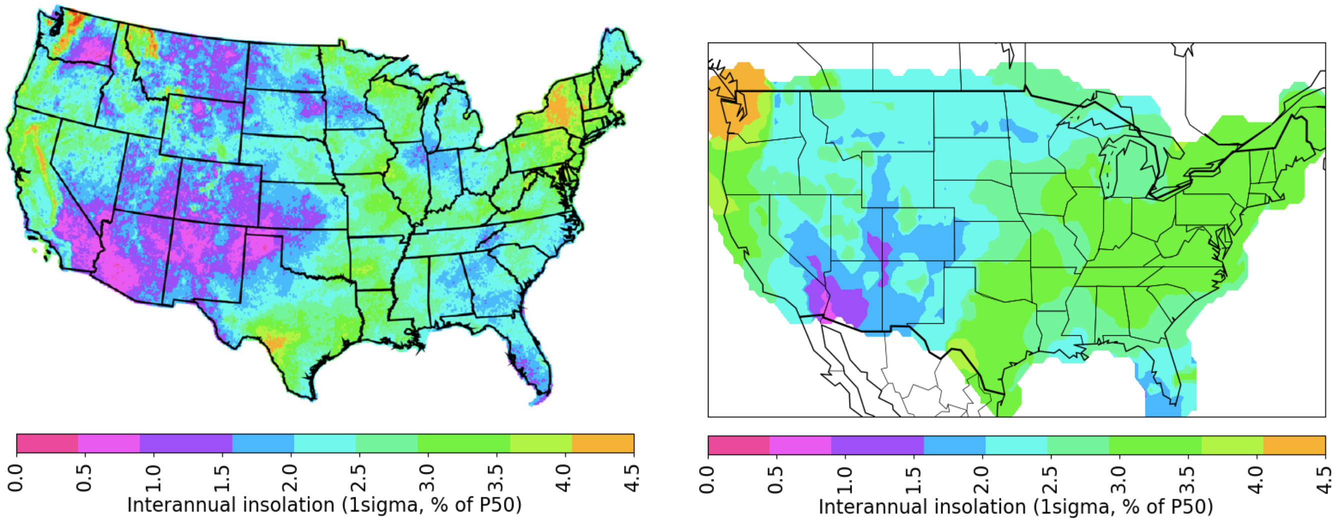

Fig. 5: Variability in interannual insolation from (a) Gueymard et al [2] and (b) values from this work. Color bins have been matched between the two studies.

of variability in annual energy yield. Additional modeling steps would be needed to convert the distribution of annual insolation values into an expected range of PV energy yield values. Hourly data on irradiance, temperature, wind speed and other environmental variables are used as inputs to a PV energy production model. Secondary loss factors such as clipping, soiling loss, and snow losses also play an important role in variability in PV energy yield.

V. Conclusion

Solar resource measurements play a critical role in the assessment of long-term energy yield and the valuation of PV systems. Based on a series of satellite- and ground-based annual insolation data sources spanning from 1961-2017, we compile data by region and by Köppen-Geiger climate zone to model interannual variability in insolation. From the aggregated data, the P99 annual insolation values were found to range from -4 to -8% of the P50 value across the continental United States.

References

[1] B. Muller, W. Heydenreich, N. Reich, C. Reise, and B. Farnung, "Investment risks of utility-scale PV: Opportunities and limitations of risk mitigation strategies to reduce uncertainties of energy yield predictions," pp. 1–5, IEEE, June 2015.

[2] C. A. Gueymard and S. M. Wilcox, "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*, vol. 85, pp. 1068–1084, May 2011.

[3] "NSRDB-Vol. 1. User's Manual-National Solar Radiation Data Base," 1992.

[4] J. Dise, "Advances in Long-Term Solar Energy Prediction and Project Risk Assessment Methodology Through Non-Normally Distributed Probabilities of Exceedance," *8th PVPMC*, vol. 2017.

[5] S. Wilcox, "National Solar Radiation Database 1991–2010 Update: User's Manual," 2012.

[6] R. Perez, J. Schlemmer, K. Hemker, S. Kivalov, A. Kankiewicz, and C. Gueymard, "Satellite-to-irradiance modeling - a new version of the SUNY model," pp. 1–7, IEEE, June 2015.

[7] M. Wild, D. Folini, F. Henschel, N. Fischer, and B. Müller, "Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems," *Solar Energy*, vol. 116, pp. 12–24, June 2015.

[8] B. Müller, M. Wild, A. Driesse, and K. Behrens, "Rethinking solar resource assessments in the context of global dimming and brightening," *Solar Energy*, vol. 99, pp. 272–282, Jan. 2014.

[9] T. Tomson, "Dynamic behavior of solar radiation," in *Modeling Solar Radiation at the Earth Surface*, Springer, 2008.

[10] M. Peel, "Updated world map of the Köppen-Geiger climate classification," *Hydrology and Earth System Sciences Discussions*, vol. 2007, pp. pp.439–473.