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 atmosphere 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]. 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 1961-2010 to increase the number of annual insolation samples for a geographic location and reduce the uncertainty in the distribution in annual insolation.

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 SolarAnywhere 3.2 model from Clean Power Research were included from 1998-2017 for 1165 station locations [5]. 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 Koeppen-Geiger climate zone were included in the regional aggregate. Each location’s annual insolation time series data was scaled to the median of the regional distribution to emphasize the distribution shape rather than differences in the median values within the region.

Results

Fig. 2a and Fig. 2b show maps of the continental United States for annual insolation: median and P99 probability of exceedance. Overall the P99 probability of exceedance values were found to be 4 to 8% of the P50 value across the continental United States. Spatial aggregation of data increases the sample size (Fig. 2c) and provides improved estimates of interannual variability. Based on the simulations in Fig. 2d, an accurate estimate of variability in annual insolation value for a project location requires at least 5 times more samples than the ~20 years available for most locations.

One concern is that the aggregated annual insolation data may have significant spatio-temporal correlation between nearby sites or nearby years. Temporal correlations were found to be quite low (0.05 ± 0.21), and were even lower by computing the correlation coefficient for 31 sites between the data and its 1 year lag. Spatial correlations between nearby sites were found to be significant (0.67 ± 0.25), suggesting that the degrees of freedom may be somewhat less than the sample size for computing confidence intervals on distribution statistics. In addition, regional aggregation smooths out local microclimate impacts and the regional results presented here should not be used as site-specific values.

Annual insolation is the dominant driver of annual energy yield, but other secondary parameters such as temperature, wind speed, aerosol optical index, and diffuse irradiance fraction could have a significant effect on 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.

Fig. 1: (a) Koeppen-Geiger climate map of the United States with overlaid WBAN, USAF-METSTAT and CPR-SolarAnywhere analysis locations. (b) For each Koeppen-Geiger climate zone, all of the available site-years were compiled into a histogram. One sigma variability values range from 1.3 to 4.0% by climate zone.

Fig. 2: (a) Regional median annual insolation, aggregated by region and by Koeppen-Geiger climate zone. (b) Regional P99 annual insolation data, generated using a normal distribution fit to the aggregated insolation data. (c) Number of site-years aggregated for each location (d) Statistical study based on random sampling of the normal distribution demonstrating that >100 site-years of data are needed to estimate the one-sigma variability to within 5%.

References