Conference Proceedings

ASES National Solar Conference 2016

San Francisco, CA, USA July 10-13, 2016

Solar Energy Assessments: When is a Typical Meteorological Year Good Enough?

Sophie Pelland[1], Charles Maalouf[2], Renée Kenny[1], Louise Leahy[1], Brad Schneider[1] and Gwendalyn Bender[1]

[1] Vaisala, Seattle (U.S.A.)

[2] Independent, Seattle (U.S.A.)

Abstract

This study compares probability-of-exceedance values (P-values) for photovoltaic systems derived using multiple years of Vaisala's 3TIER Services weather data to those derived using a Typical Meteorological Year (TMY) based on the same resource data. Both approaches were used to estimate the year-1 yield of eighteen Megawatt-scale photovoltaic projects located in different parts of North America, South America and Asia. P-values in the TMY case were derived by using the standard deviation of annual global horizontal insolation as a proxy for inter-annual variability. All other uncertainties were treated identically in both approaches. Since our analysis included only eighteen case studies, we supplemented it by examining extreme cases where the differences between the two approaches should be maximum, namely cases where inter-annual variability dominates all other uncertainties. P50 values derived from a 3TIER Services TMY are usually within 0.5% of those derived using a full time series. Meanwhile, other P-values derived using a TMY exhibited a positive bias, indicating that this approach systematically underestimates uncertainty. A simple method for removing this bias was developed using ten projects as a training data set, and tested on the remaining eight projects. Overall, differences in P90 and P99 values are typically less than 1%, but can reach up to 2-5% in extreme cases. These results can serve as benchmarks for deciding whether and when TMY analysis is good enough.

Keywords: *Photovoltaic system, energy modeling, Typical Meteorological Year, TMY, energy assessment*

1. Introduction

Typical Meteorological Year (TMY) files provide one year (8760 hours) of synthetic weather data that is meant to capture a typical year. They are constructed from an underlying long-term, multi-year time series of meteorological data, with the aim that the annual energy yield from a TMY simulation should match as closely as possible the mean yield obtained by running a simulation over the full time series.

Currently, many photovoltaic (PV) simulation tools make it fairly easy to run simulations over multiple years of weather data. Such multi-year simulations provide distributions of annual energy yields which give an indication of the inter-annual variability that can be expected for a given project.

Meanwhile, TMY simulations cannot strictly speaking provide any information about inter-annual variability. On the other hand, simple statistics from the underlying long-term time series can provide such information. For instance, the standard deviation of the annual global horizontal insolation (GHI) can be used as a proxy for estimating the standard deviation in energy.

This study addresses the question: when is it worthwhile to run a simulation over a full time series, and when is TMY analysis good enough? Specifically, we compare probability-of-exceedance values (P-values) derived

doi:10.18086/solar.2016.01.17 Available at http://proceedings.ises.org

using both approaches, where inter-annual variability in the TMY approach is approximated using the standard deviation in annual GHI. Since both approaches are still commonly used even for Megawatt-scale projects, the aim of our analysis is to help guide decisions about which approach to use in any given case.

2. Methodology

This study is based on eighteen energy assessments that Vaisala conducted for Megawatt-scale photovoltaic projects. Projects are located in various parts of North America, South America and Asia. They include fourteen projects with single-axis horizontal East-West trackers and four with fixed (or seasonally varying) orientations. Some projects are at the pre-construction stage, while others are already operational.

Probability-of-exceedance values were generated corresponding to the year-1 yield of the projects. This is the first year of operation for new projects and the upcoming year for operational projects. Specifically, P50, P75, P90 and P99 values were calculated. These indicate, respectively, the energy yield which a PV project has a 50%, 75%, 90% and 99% probability of exceeding during year-1.

2.1. Full Time Series Simulations

Full time series simulations were conducted using 3TIER Services meteorological data as input to the PVsyst simulation software (PVsyst, 2016). Vaisala's 3TIER Services derives irradiance using images from the visible channel of satellites in geosynchronous orbit. The ground resolution of the data is approximately 3 km and images are collected every 10-60 minutes depending on the region and time period. A proprietary algorithm is used to convert the satellite-observed images into ground irradiance (Vaisala, 2016). Wind and temperature data are constructed using the Weather Research and Forecasting (WRF) Numerical Weather Prediction model using reanalysis data for initial and boundary conditions. The time series for the eighteen project locations covered between 16 and 19 years, ranging from 1997 to 2016.

For each project, PVsyst was run in batch mode to generate annual energy yields for each year of weather data, returning 16 to 19 year-1 yields. These multiple year-1 yields were then used to construct a probability distribution of year-1 yields using kernel density estimation (KDE).

Kernel density estimation is a non-parametric method of estimating the probability density function of a random variable (Silverman, 1998). Kernel density estimators are a generalization over empirical histograms, which are often used. Estimating a density function with a histogram involves dividing the data into bins of equal width, then counting the number of observations falling within each bin. This leads to a density estimator that is not smooth and highly dependent on the end points of each bin, as well as the width of the bin. Kernel density estimators center a kernel function at each observation, averaging out the contribution of all observations over a local neighborhood of the given observation. Using a continuous kernel function also yields a smooth estimator. This alleviates the first two issues of the histogram above. Unfortunately, there remains the issue related to the bandwidth of the kernel function, similar to the width of a histogram's bins. A bandwidth that is too small will result in a highly variable density estimate, while a bandwidth that is too large will result in a biased one. It is very important to select an appropriate bandwidth value. In this analysis, we selected bandwidths using a cross-validation approach, where data points were withheld one at a time, and the bandwidth leading to the maximum total log-likelihood over withheld data points was selected.

2.2. Typical Meteorological Year Simulations

Vaisala creates TMY datasets using an empirical approach that selects four-day samples from the full time series to create a "typical year" of data with 8760 hours, while preserving the monthly and annual means of either global horizontal irradiance (GHI) or direct normal irradiance (DNI). The process is iterated until the monthly and annual means of both GHI and DNI in the TMY dataset match the means of the full time series to within roughly 0.5% or less.

The TMY datasets were used as inputs to PVsyst for each of the eighteen projects. The resulting year-1 yield was interpreted as the mean of a normal distribution of year-1 yields. In order to estimate the standard deviation of this distribution, different proxies for the standard deviation in year-1 yield were evaluated, the best of which was found to be the standard deviation in annual GHI.

Figure 1 shows the standard deviation in the year-1 energy yields obtained through the full time series simulations vs the standard deviation in GHI. As can be seen from this figure, the standard deviation in GHI tends to systematically underestimate the standard deviation in energy. We therefore considered two versions of the TMY approach: one in which the standard deviation in GHI was used directly and another, which we refer to as "TMY-adjusted", in which corrections to the standard deviation in GHI were made to partly compensate for biases. These corrections were developed on the first ten projects that we analyzed, which included eight tracking systems and two fixed tilt systems. These ten projects acted as our training data set, while the next eight projects were used as a testing data set on which to independently validate the corrections developed using the first ten projects.

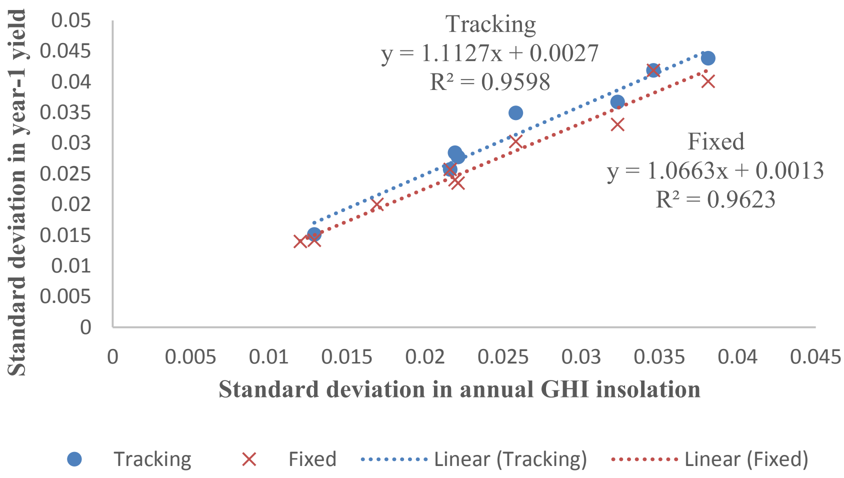

Fig. 1: Standard deviation in year-1 yield vs. standard deviation in annual GHI for fixed and tracking systems. All standard deviations are expressed are as fraction of the annual mean. Note that for the purposes of this fit, fixed orientation equivalents to the eight tracking projects in the training data set were modeled.

The first correction was to use the Figure 1 linear fits to estimate standard deviation in energy from standard deviation in GHI. Separate fits were performed for tracking systems and for fixed tilt systems, as shown in Figure 1. A second linear adjustment was made to account for the fact that, in the TMY approach, year-1 yields are treated as normally distributed, whereas in the multi-year time series approach the more general KDE distribution is used. It turns out that the KDE distributions tend to have fatter tails than the normal distributions. Obviously, since KDE distributions and normal distributions generally have different shapes, it's not possible to match these up completely. As a proxy to this, the P90 and P50 values from the KDE distributions were used to calculate the standard deviation of a normal distribution matching these two P-values. The resulting standard deviation was then fitted linearly against the standard deviation of the original normal distribution as shown in Figure 2.

The standard deviation of the year-1 yield in the adjusted-TMY approach was thus calculated as follows: the standard deviation in year-1 yield was estimated by using the standard deviation in GHI as input to the fit in Figure 1 (fixed or tracking), which was used as input to the fit in Figure 2. Additional uncertainties were also included in the adjusted-TMY approach to account for the standard error in the Figure 1 and Figure 2 fits, as well as for the uncertainty of about 0.5% or less on the P50 that comes from using a TMY. The overall uncertainty associated with inter-annual variability in the adjusted-TMY approach was calculated assuming that each of these sources of uncertainty are independent of each other, so that the overall standard deviation is given simply by the square root of the sum of their squares. Equation (1) gives this overall standard deviation, σ_{IAV}, as a function of the standard deviation in annual GHI, σ_{GHI}.

$$\sigma_{IAV} = \sqrt{a\sigma_{GHI}^2 + b\sigma_{GHI} + c} \qquad \text{(eq. 1)}$$

where a=1.0401, b=0.02036, c=0.0001482 for tracking systems, and a=0.9550, b=0.0170, c=0.0001238 for fixed systems.

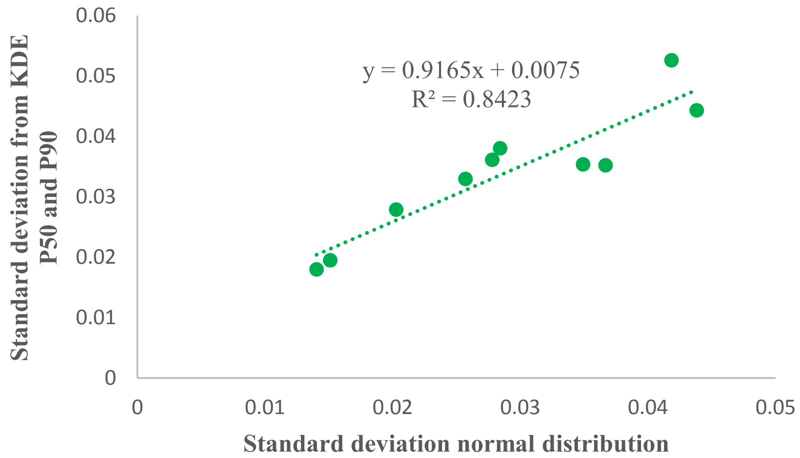

Fig. 2: Standard deviation of a normal distribution calculated from the P50 and P90 of the KDE distribution vs. standard deviation of the original normal distribution.

2.3. Overall Uncertainty in Year-1 Yield and P-values

Uncertainties not associated with the inter-annual variability in the weather were treated identically across all approaches. These uncertainties can be classified into the following categories:

- **Resource modeling:** Resource modeling uncertainty captures the uncertainties related to the accuracy of the satellite derived irradiance data utilized in the energy assessment, excluding uncertainties associated with climate variability. In some cases, this uncertainty is reduced by making adjustments to the satellite data based on comparisons with ground station measurements. Since projects can sometimes span more than one satellite pixel, a spatial component is included in the resource uncertainty to reflect the pixel-to-pixel variability in the solar resource.
- **Power modeling:** Power modeling uncertainty considers each step in converting solar irradiance estimates into energy estimates. This uncertainty captures the following: uncertainty in the transposition model used to derive irradiance in the plane of the array, bias in the simulation model itself and uncertainties in the inputs to the simulation model. Uncertainties in simulation model inputs include uncertainties in PV system specifications and uncertainties in the various losses that can reduce PV system output.
- **Aging:** The rate at which photovoltaic systems experience degradation is subject to uncertainty. Vaisala uses technology-specific median long-term degradation rates based on an extensive review of the existing scientific literature by NREL (Jordan and Kurtz, 2011). The difference between the degradation rate corresponding to the module manufacturer's 25-year minimum output warranty and the median degradation rate is used to determine the uncertainty in the annual degradation rate. Uncertainty in the annual degradation rate is then propagated over the time period of interest to yield an overall aging uncertainty. Note that for year-1 yields, this uncertainty is typically quite small compared to the other sources of uncertainty.

These uncertainties were combined with the inter-annual distributions in year-1 yield (either KDE or normal distributions) to generate an overall cumulative distribution function from which P-values were obtained.

2.4. Analysis of Extreme Cases

Since our analysis is based on a fairly small sample of eighteen projects, it may not pick up extreme cases where the difference between the TMY approach and the full time series approach is most pronounced. We conducted two analyses to try to expand our results to capture extreme cases. Since uncertainties other than inter-annual variability are treated identically in the TMY and full time series approaches, the differences between the two approaches should be most pronounced when inter-annual variability is large relative to other uncertainties. In order to explore this, two hypothetical projects (one tracking, one fixed) were simulated at a location near Pades, Romania, where inter-annual variability is high. All other uncertainties were set to realistic minimum values.

The second extreme case analysis consisted of calculating P-values for each project neglecting all uncertainties except inter-annual variability. This essentially mimics the case where other uncertainties are negligible compared to inter-annual variability.

3. Results and Discussion

The main results of our analysis are shown in Tables 1 and 2. These give percent differences between P-values calculated using the TMY and TMY-adjusted approaches and P-values calculated using the full time series, for the ten projects in the training data set (Tab. 1) and the eight projects in the testing data set (Tab. 2). If we consider first the P50 values, the mean difference is 0.1% in both cases, with standard deviations of 0.3-0.4% and a maximum of 0.8%. This is in line with the fact that 3TIER Services TMY means typically match long-term means of annual insolation to within 0.5%. These results can be compared to those of Ryberg et al. (2015), who simulated PV system yield at 239 locations in the United States by running simulations for representative fixed and tracking systems using both 30-year time series and TMYs. Their TMY results differed from their 30-year P50 values by up to ±4% in some cases. The cause of this difference is not clear: it is probably due in part to differences in how well the NREL TMY means match the means of the underlying time series, but could also be due to cases where the means and P50 values in the 30-year distributions differ substantially.

Considering other P-values in Tables 1 and 2, the TMY approach without adjustment tends to underestimate uncertainty, as expected, leading to P-values that are too high. This is reflected in the fact that the means for all P-values in this approach are positive. Meanwhile, the TMY-adjusted approach leads to P-values that are on average very close to those derived using the full time series. This is true both for the training and the testing data sets, showing that the bias correction works outside of the context in which it was developed. The largest differences in P-values between the TMY and full time series approaches are of the order of 2% in the unadjusted case, and less than 1% in the adjusted case.

In the extreme case scenarios discussed in Section 2.4., differences in the P90 reach 3.6% in the unadjusted case and 2.0% in the adjusted case, while differences in the P99 reach 5.3% in the unadjusted case and 3.2% in the adjusted case. Although their analysis differed from ours in a number of ways, it is still instructive to compare our results to those of Ryberg et al. (2015). Since their analysis considered only inter-annual variability as a source of uncertainty, it can be re-interpreted as an extreme case where this source of uncertainty dominates. Ryberg et al. (2015) provided TMY results as well as standard deviations in energy and P90 values for each location. We re-analyzed their results to compute P90 values for each location using the TMY mean and the standard deviation, and compared this to the P90 values they derived from empirical distribution functions based on 30-year simulations. Differences between the two P90 values reached up to 3-4%.

Tab. 1: Percent differences between P-values calculated using TMY approaches and full time series for the ten projects in the training data set

		TMY			TMY-adjusted		
Project #	**P50**	**P75**	**P90**	**P99**	**P75**	**P90**	**P99**
1	0.8%	1.0%	1.3%	1.5%	0.8%	0.7%	0.3%
2	-0.7%	0.0%	0.7%	1.7%	-0.4%	-0.2%	0.0%
3	0.1%	0.5%	0.9%	1.8%	0.1%	0.2%	0.4%
4	0.4%	0.7%	0.8%	0.9%	0.2%	0.0%	-0.8%
5	0.5%	0.9%	1.1%	1.1%	0.4%	0.1%	-0.9%
6	-0.2%	0.0%	0.3%	0.9%	-0.4%	-0.4%	-0.5%
7	0.2%	0.4%	0.5%	0.9%	0.0%	-0.2%	-0.5%
8	0.0%	0.0%	0.1%	0.2%	-0.2%	-0.3%	-0.6%
9	-0.2%	0.2%	0.6%	1.7%	-0.1%	0.1%	0.5%
10	-0.1%	0.2%	0.6%	1.2%	-0.1%	0.0%	-0.1%
Mean	**0.1%**	**0.4%**	**0.7%**	**1.2%**	**0.0%**	**0.0%**	**-0.2%**
Standard deviation	**0.4%**	**0.4%**	**0.4%**	**0.5%**	**0.4%**	**0.3%**	**0.5%**
Maximum	**0.8%**	**1.0%**	**1.3%**	**1.8%**	**0.8%**	**0.7%**	**0.5%**
Minimum	**-0.7%**	**0.0%**	**0.1%**	**0.2%**	**-0.4%**	**-0.4%**	**-0.9%**

Tab. 2: Percent differences between P-values calculated using TMY approaches and full time series for the eight projects in the testing data set

		TMY			TMY-adjusted		
Project #	**P50**	**P75**	**P90**	**P99**	**P75**	**P90**	**P99**
11	-0.3%	0.0%	0.3%	1.0%	-0.2%	-0.2%	0.0%
12	0.3%	0.4%	0.4%	0.5%	0.2%	0.1%	-0.2%
13	0.4%	0.7%	1.0%	1.7%	0.5%	0.5%	0.7%
14	0.2%	0.7%	1.0%	2.0%	0.3%	0.4%	0.7%
15	-0.3%	-0.2%	-0.1%	0.3%	-0.4%	-0.5%	-0.6%
16	0.0%	0.1%	0.2%	0.3%	-0.1%	-0.2%	-0.4%
17	0.5%	0.9%	1.2%	1.8%	0.6%	0.6%	0.7%
18	0.0%	0.2%	0.4%	0.7%	-0.1%	-0.3%	-0.6%
Mean	**0.1%**	**0.3%**	**0.6%**	**1.0%**	**0.1%**	**0.1%**	**0.0%**
Standard deviation	**0.3%**	**0.4%**	**0.5%**	**0.7%**	**0.4%**	**0.4%**	**0.6%**
Maximum	**0.5%**	**0.9%**	**1.2%**	**2.0%**	**0.6%**	**0.6%**	**0.7%**
Minimum	**-0.3%**	**-0.2%**	**-0.1%**	**0.3%**	**-0.4%**	**-0.5%**	**-0.6%**

4. Conclusion

This analysis started with the question: when is a TMY solar energy assessment good enough? Obviously, there is no hard-and-fast answer to this question. It will depend in particular on which TMY dataset is being considered, on user requirements as to what constitutes an acceptable difference between TMY and full time series analyses, as well as on specifics of the PV project, including the relative size of inter-annual variability and of other uncertainties.

Having said that, this study does provide some rough benchmarks to help address this question. First, it shows that differences in the P50 closely reflect differences between the TMY means and the long-term time series means of GHI and DNI. In te case of the 3TIER Services TMY, this difference is usually less than 0.5%. For other P-values, our analysis shows that using the standard deviation in GHI as a proxy for inter-annual variability in the yield tends to systematically underestimate uncertainty, but also that this bias can be removed through simple corrections. With this correction applied, differences in P-values between the TMY-adjusted and full time series simulations for the eighteen PV projects analyzed were all less than 1%. However, our analysis also suggests that differences in P-values can reach up to 2%-5% in extreme cases where inter-annual variability dominates all other uncertainties. Such cases could correspond for instance to operational reforecasts in regions with high inter-annual variability, since power modeling and resource modeling uncertainties can often be significantly reduced when past project performance and weather data are available.

One way to decide whether or not a TMY approach is appropriate is to ask whether or not these types of differences on P-values are acceptable for a given project. If P-values are being used to secure financing on Megawatt-scale projects, then the small added complexity involved in running a full time series will probably seem worth the effort! On the other hand, if a project is still at the pre-feasibility stage with inter-annual variability being relatively small compared to other uncertainties, then a TMY analysis will probably be good enough. Finally, while this study focused on the impact of TMY analysis on P-values, there can be other reasons for running a full time series analysis, for instance whenever there is a need to generate a realistic long-term time series of PV output power.

5. Acknowledgements

We would like to thank Mark Stoelinga, Gwendalyn Bender and William Gustafson for helpful feedback and suggestions.

6. References

Jordan, D.C. and Kurtz, S.R., 2013. Photovoltaic degradation rates – an analytical review. Progress in Photovoltaics: Research and Applications, Volume 21, Issue 1, 12–29.

PVsyst photovoltaic software website, http://www.pvsyst.com/en, accessed May 30, 2016.

Ryberg, D.S., Freeman, J., Blair, N., 2015. Quantifying Interannual Variability for Photovoltaic Systems in PVWatts, Technical Report NREL/TP-6A20-64880, available online at http://www.nrel.gov/docs/fy16osti/64880.pdf, accessed May 30, 2016.

Silverman, B.W., 1998. Density Estimation for Statistics and Data Analysis. Chapman & Hall/CRC, London, UK.

Vaisala, Vaisala 3TIER Services Global Solar Dataset/ Methodology and Validation, available online at http://www.vaisala.com/en/energy/Documents/WEA-ERG-3TIER-Global%20Solar%20Validation.pdf, accessed May 30, 2016.