VAISALA ENERGY

3TIER by Vaisala

COMPARING PHOTOVOLTAIC YIELD PREDICTIONS FROM SYNTHETIC WEATHER YEARS TO THOSE FROM HOURLY TIME SERIES

By Gwendalyn Bender, Sophie Pelland Ph.D, and Louise Leahy Ph.D., Vaisala

Introduction

You're a developer or engineering firm who is working on a large solar project and you're deciding which solar resource dataset to use. You know you need to use a high-quality resource dataset to get the project across the line. But there is a tight deadline and no guarantee that the project will get a power purchase agreement, which means you don't want to spend a lot of money up front.

Solar project developers don't have many guidelines for what to use when it comes to resource data except those they impose on themselves and occasional guidance from the financial community. Also, there is not much information available that compares different resource datasets and allows developers to make informed choices. How do you make a good choice?

There are three key questions to ask when choosing your resource data.

1. Are hourly data available, or only monthly means?
2. How old is the dataset?
3. What is the dataset uncertainty?

If you can answer these questions about the resource data used in your projects, then you're already on the path to making better choices than many of your peers. Based on these answers and your risk appetite, you can make an informed tradeoff between schedule, accuracy, and cost. In this paper, we are going to focus on the first question.

Long-term monthly mean meteorological data is sometimes used as a basis for predicting the yield of photovoltaic systems, largely because this data is widely available at little or no cost. Software tools like PVsyst allow users to input monthly mean data, and then generate from this a year of synthetic hourly data (hereafter: PVsyst-SYN) that can be used as an input to photovoltaic yield modeling. But is this good enough?

Solar resource datasets come at different temporal resolutions – monthly, hourly, and sub-hourly. PVsyst and other solar energy modeling software allow you to model hourly data based on a monthly input. PVsyst and some other software programs will create an hourly energy profile for you based on monthly average values.

At very early stages of project design, this can be a quick way to get indicative numbers. If you have very good monthly data, then this approach may even be ok for fixed-tilt PV projects in the prospecting stages. The problem with this approach is it is not equivalent to using hourly data, even in the early prospecting stages of project development depending on the location and type of project. This is especially true for tracking PV plants.

Study Conditions

Vaisala performed a study to compare photovoltaic yield predictions based on PVsyst-SYN synthetic data to yield predictions based on the long-term hourly time series of meteorological data from which the monthly means in the PVsyst-SYN approach were derived and to TMY (Typical Meteorological Year) resource data generated from the same data source.

In this study, we used PVsyst to compare the P50 energy results from using multiple years of Vaisala's 3TIER Services hourly meteorological data (Vaisala, 2017) to those derived using two types of synthetic years based on the same data: a PVsyst-SYN synthetic year and a Typical Meteorological Year (TMY). The same underlying dataset was used and the only difference in the projects was the temporal resolution.

Modeling was performed for ten Megawatt-scale photovoltaic projects in six different countries: the United States., India, China, Mexico, Argentina, and Chile. For each project, simulations were run for two configurations: an equator-facing fixed-tilt orientation and one with horizontal single-axis East-West trackers with backtracking.

There is not much information available that compares different resource datasets and allows developers to make informed choices.

PVsyst was used to run all simulations, with only the meteorological data varying between the three approaches.

Full time series approach: The hourly time series for the ten project locations covered between 17 and 20 years, ranging from 1997 to 2017. A simulation based on each individual year was run separately in PVsyst to derive the corresponding annual yield. The full time series yield estimate (P50) was then taken to be the median of the individual annual yields.

PVsyst-SYN approach: In this approach, the long-term monthly means of global horizontal irradiance (GHI) are first used to stochastically derive synthetic hourly GHI data as described in Meteotest, 2017. Diffuse horizontal (DIF) irradiances are then derived from GHI using the DIRINT separation model (Perez etal., 1991). Hourly DIF values are then renormalized so that the monthly DIF means match the long-term means. Finally, hourly direct normal irradiance (DNI) values are then calculated from hourly GHI and hourly DIF.

TMY approach: 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 conserving the monthly and annual means of either GHI or DNI. The process is iterated until the annual means of all solar variables in the TMY dataset match the means of the full time series to within 1% or less.

Finally, the simulated yields from the PVsyst-SYN and TMY approaches were compared to the P50 yields from the full time series approach for each of the ten PV systems, for fixed and tracking configurations.

Study Results

As shown in Figure 1, differences with the P50 values from the full time series are significantly more pronounced in the PVsyst-SYN case than in the TMY case. The large differences in the PVsyst-SYN case suggest that this approach should not be considered as a proxy to the full time series approach, but as a different process with different uncertainties.

The larger differences associated with tracking systems using the monthly average PVsyst-SYN are because those systems are more sensitive than fixed systems to the direct-diffuse split of the global irradiance. Indeed, the PVsyst-SYN approach is designed to match GHI and DIF monthly means, but this does not guarantee a good match to DNI means. For instance, for the two projects which show the largest differences in energy, projects 4 and 8, the PVsyst-SYN DNI annual means are 5.0% and 11.7% higher, respectively, than the long-term mean DNI values .

The deviation between the energy yields from the long-term hourly time series and the TMY files is consistent with our previous studies (Pelland, S., et al., 2016) and relates to how well the TMY averages match the long-term dataset. If the TMY was not well matched for GHI we would expect these energy yield biases to be larger. Vaisala's TMY creation process is designed to minimize these deviations for that very reason.

Project number	Fixed TMY	Fixed PVsyst-SYN	Track TMY	Track PVsyst-SYN
1	-0.3	-1	-0.4	0.7
2	-0.1	-0.4	-0.2	1.4
3	0.5	-0.6	-0.7	0.3
4	0.7	1.8	0.9	4.8
5	0.2	0.1	-0.6	1.2
6	-0.2	-0.1	-0.5	-0.2
7	0.6	1.7	0.1	0.9
8	0.1	0.9	1.1	7.8
9	0	0.9	0.7	3.1
10	0.5	-0.6	-0.4	-1.6
Mean	0.2	0.3	0	1.8
Standard deviation	0.4	1	0.7	2.7
Minimum	-0.3	-1	-0.7	-1.6
Maximum	0.7	1.8	1.1	7.8

Figure 1: Percent differences between annual energy yields from TMY and PVsyst-SYN and P50 energy yields from the full time series for ten utility-scale PV systems for fixed and tracking configurations

For tracking plants, the average difference in the hourly versus monthly average approach was about 2% and the maximum difference was over 7%. For tracking plants, the monthly average PVsyst-SYN approach may introduce significant differences in energy estimates in later project development stages. A 7% difference in energy from predicted to actual production is enough to make a break a project; thus, we recommend using hourly data such as Vaisala's for which detailed validations exist and uncertainties are correspondingly well documented and understood, which is not necessarily the case when synthetic data is generated from monthly means

The news is better for fixed PV. The energy differences between the models was less than 1% and the maximum difference was around 2%. That is a deviation that's generally acceptable at early project stages. For project financing many banks require that hourly data be used and you do run the risk of a valuation change at a late stage if you're not using hourly data from the beginning of a project's development.

Proactive Approach

Since each of the three approaches considered in this study involves resource modeling, the ultimate test of each would be to compare its yield predictions to metered energy data from operational PV systems. We are pursuing this in upcoming work, but do not expect that the PVsyst-SYN monthly average method to be a better proxy to actual hourly production than hourly resource time series.

So as a proactive project developer or engineer how can you make sure you are making the best resource choices for your projects? We can say in situations where you have to choose between high quality monthly averages or low quality hourly time series, choosing the monthly averages is a reasonable choice, particularly for fixed PV plants.

However, high quality hourly data is currently accessible for nearly all locations globally at this time, for a nominal cost. Vaisala and other providers cover most geographies between 60 degrees north and 50 degrees south with satellite based datasets at an hourly or even sub-hourly resolution. For polar regions not covered by satellite data, numerical weather prediction modeling can be used in conjunction with ground station observations. Vaisala provides hourly time series data for as low as $50 U.S., which is a small price to pay when you consider the substantial increase to a project's energy yield accuracy and corresponding reduction in uncertainty.

It is important to take the time to understand the resource methodology and the energy estimate accuracy of

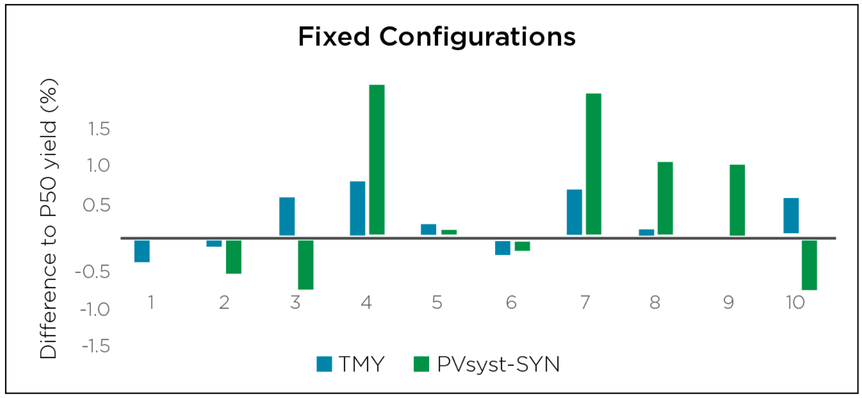

Figure 2: Difference between annual energy yields based on TMY and PVsyst-SYN synthetic approaches and median (P50) annual energy yield from full time series for 10 fixed-tilt PV systems

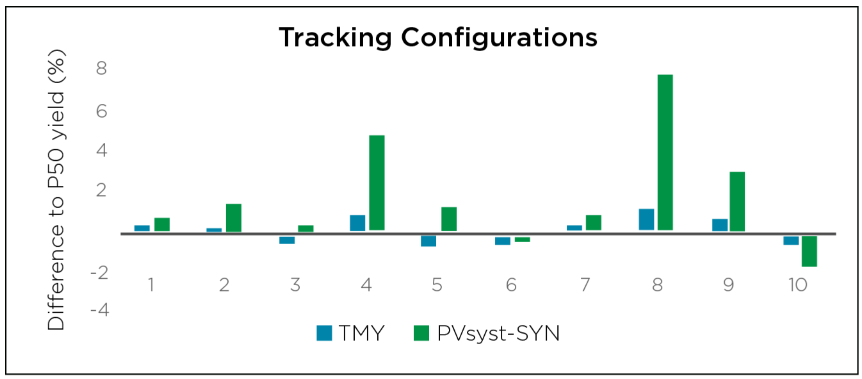

Figure 3: Difference between annual energy yields based on TMY and PVsyst-SYN synthetic approaches and median (P50) annual energy yield from full time series for 10 tracking PV systems

your engineering firm. Unfortunately, you cannot rely on a well-known name of a data provider being a sufficient indication of quality. Many prominent engineering firms get access to low cost monthly means from providers like Vaisala and others, but do not use the hourly time series in order to reduce their own costs.

If you have received a report from an engineering firm, there are key phrases to look for in the resource methodology sections to understand the approach used and its accuracy.

- "During the simulation the monthly meteorological data are processed to generate a synthetic year of hourly data."
- "Generated a times series of hourly values representing long-term average conditions at the project site."
- "On the basis of the long-term average values selected and presented before, a consistent hourly time series of GHI, DHI, ambient temperature, wind speed and precipitation has been generated using the well-established, climatic generator feature of Meteonorm 7.1."

Any time the words "synthetic" or "generated" are used this is an indication that any hourly values used are derived and not native to the resource time series.

If a synthetically generated hourly dataset was used, extra scrutiny should be applied when reviewing the uncertainty estimates of the project. In energy estimates produced by Vaisala we always use the hourly time series data available from our proprietary satellite based resource dataset to ensure the highest accuracy and lowest uncertainty for our customers. Given the increase in the availability and the decrease in costs for hourly time series, we would encourage others in the industry to evaluate their use of the synthetic data creation practice. By better understanding the implications on estimated energy output of using various resource data sources we can all build better projects together.

Acknowledgements

We would like to thank Jean-Marie Feppon from the PVsyst team for helpful comments and clarifications, and Mark Stoelinga and William Gustafson from Vaisala for useful feedback.

References

Meteotest, 2017. Meteonorm Handbook Part II: Theory, version 7.2 / March 2017, section 7.6.1, Accessed at http://www.meteonorm.com/en/downloads/documents/, accessed on July 13, 2017.

Pelland S, Maalouf C, Kenny R, Leahy L, Schneider B and Bender G, 2016. Solar Energy Assessments: When is a Typical Meteorological Year Good Enough?, Proceedings of the ASES 2016 conference, July 10-14, 2016, San Francisco, USA.

Perez R, Ineichen P, Maxwell E, Seals R and Zelenka A, 1991. Dynamic Models for hourly global to-direct irradiance conversion. Edited in: Solar World Congress 1991. Volume 1, Part II. Proceedings of the Biennial Congress of the International Solar Energy Society, Denver, Colorado, USA, 19-23 August 1991.

PVsyst photovoltaic software website, http://www.pvsyst.com/en, accessed July 13, 2017.

Vaisala 3TIER Services Global Solar Dataset: Methodology and Validation, April 27, 2017. http://www.vaisala.com/Vaisala%20Documents/Scientific%20papers/3TIER%20Solar%20Dataset%20Methodology%20and%20Validation.pdf, accessed July 13, 2017.