

Impact of different weather data sets on photovoltaic system performance evaluation

Chanikarn Yimprayoon and Mojtaba Navvab
University of Michigan, Ann Arbor, Michigan

ABSTRACT:

Building energy simulation plays an important role in decision makings involving energy conservation measures and choices of renewable energy systems in building designs. Traditional simulation tools rely on weather data sets called Typical Meteorological Year (TMY), representing a typical year of weather at ground weather stations throughout the United States. These data sets are constructed using an algorithm to select the “most typical” month of the many years in the database for each month. Some recent publications suggest that one-year TMY data is no longer sufficient to evaluate long-term performance of PV systems, because a typical year does not taken into account extreme weather, and thus does not address the meteorological uncertainty that might occur. Actual electricity outputs from photovoltaic systems vary from year to year. Having more accurate information about production performance should help facilitate system selections that match building designs and how to operate them. In this study, four sets of weather data, Detroit TMY2, Ann Arbor TMY3, Ann Arbor 15-year NSRDB, and Ann Arbor 13-year SolarAnywhere®, are used as inputs in PV system performance simulation. Their impacts on the PV system electricity output availability, variability and uncertainty are analyzed and compared. The magnitude and consequences of the analyses of different weather data sets are presented.

CONFERENCE THEME: On Measurement: What is performance? Approaches to energy, occupation, consumption and reuse.

KEYWORDS: Weather data, photovoltaic system output, availability, variability, uncertainty

INTRODUCTION

The performance of a building is a result of complex processes. A better building design can reduce energy use by 30% compared to a conventional building design, while still provide an equal or better environment for its occupants. To reach a 50% reduction or more, renewable energy system integration is needed (USGBC Research Committee, 2008). Barriers to achieve this goal is usually not technology constraints, but poor data to make informed decisions (Clarke, 2001). Building simulation tools are created to help provide real world replication and predict how buildings and systems will perform once they are constructed and implemented, thus providing information for decision making. Building energy performance prediction tools are a series of complex mathematical models that address the dynamic interaction of building and system performances with building geometry, plan, components, system choices, climate conditions and occupant use patterns. These computer based simulation tools usually require local weather data as main inputs for outside conditions. Performance simulation of solar energy systems such as a photovoltaic (PV) system relies heavily on specific time-location hourly weather data.

The availability of solar radiation at a specific site varies according to location latitude, topography, time of day, time of year, cloud cover and atmospheric aerosol condition. The amount of solar radiation and its temporal distribution at a specific location are essential information for determining if a PV system is suitable for that site. This information can be used to select the solar energy system's size and predict its performance and operation. Onsite measured solar radiation provides the most accurate information. However, measurement equipment and their maintenance are costly. Solar radiation data can also be obtained from the nearest ground weather stations which monitor weather data such as daylight hours, air temperature, humidity, pressure, wind direction, wind speed and other climate-related information. Protocols such as the International Daylight Measure Program (IDMP) have been developed as a guideline for these meteorological data measurements (CIE TC 3-07, 1994).

IDMP also provides quality assurance procedures so that data from weather stations following this protocol can be compared and utilized uniformly. The National Solar Radiation Data Base (NSRDB) maintains hourly weather data sets at various weather stations throughout the country from 1961-2005. To accommodate building simulations utilizing weather data, one-year Typical Meteorological Year (TMY) weather data sets are developed from information in the NSRDB database. TMY is an hourly weather data set that is usually used in building simulation tools. It represents a “typical year” of weather for ground weather stations. Results from simulation using TMY weather data sets will represent building performance in “typical weather.” Recent publications suggest that using one-year TMY data for PV performance evaluations might not be suitable (Storck et al., 2010, Dean, 2010), because a typical year weather data set omission of extreme weather leads to the inability to predict atypical weather that might occur.

At locations away from ground weather stations, data from the nearest station or, alternatively, estimates based on the interpolation of data between stations are used. The problem with using data generated in this way is that the accuracy of the data decreases with distance from or between ground weather stations. A method to estimate solar radiation based on data from meteorological satellites has been developed. An example is the State University of New York at Albany (SUNYA) model developed by Perez et al. (Perez et al., 2002). In the SUNYA model, satellite images are used to derive solar radiation data that is time and place specific. These satellites are geostationary, staying fixed over one spot directly above the equator to monitor the earth’s atmosphere. The geostationary satellite data offer the advantages of wider geographic coverage with high-resolution images typically at one to ten square kilometers per pixel. They repeatedly scan earth’s image, typically at 30 to 60 minute intervals. Mathematical models are developed to generate high-resolution solar radiation resource maps based on this data.

This study compares results from PV performance simulation using different weather data sets. Within these data sets, ambient temperature, global horizontal irradiance (GHI), direct normal irradiance (DNI) and wind speed are the most importance parameters. A building located in the northern part of Ann Arbor, Michigan was used to provide context for this experiment. PV outputs generated from satellite-derived weather data set were used in comparison to the outputs from three other weather data sets. Their impacts on the PV system output availability, variability and uncertainty were quantified. The availability of PV output is how much electricity is expected to be generated from the system in one year. Electricity output variability was obtained by calculating the daily average standard deviation of each month. A histogram was used to quantify the frequency of the daily electricity output to show the uncertainty that might occur.

I. BACKGROUND

I.1 NATIONAL SOLAR RADIATION DATABASE (NSRDB) WEATHER DATA

Solar radiation data in the United States are available from the NSRDB. The database is developed by the National Renewable Energy Laboratory (NREL) and the National Climatic Data Center (NCDC), and provides hourly solar irradiance and other climate parameters for public use. Data from 1961 to 1990 are available for 239 locations, and data from 1990 to 2005 are available for 1,454 locations (Figure 1). Among these locations, only 40 stations have solar radiation measurements. At other locations, solar radiation data are estimated using calculations based on other weather parameters measured at ground weather stations such as barometric pressure and the amount of cloud cover. Several solar radiation calculation models have been developed to predict the available solar radiation at ground weather stations without solar radiation measurement equipment such as the METSTAT model which was developed for the NSRDB (Maxwell, 1998). The available solar radiation measurements from 40 stations are used to validate these solar radiation models.

I.2 TYPICAL METEOROLOGICAL YEAR (TMY)

Weather data at each location show variation from year to year. To obtain average building performance from a simulation, several years of weather data should be used. However, using several years of weather data is time consuming. Therefore, a Typical Meteorological Year (TMY) is

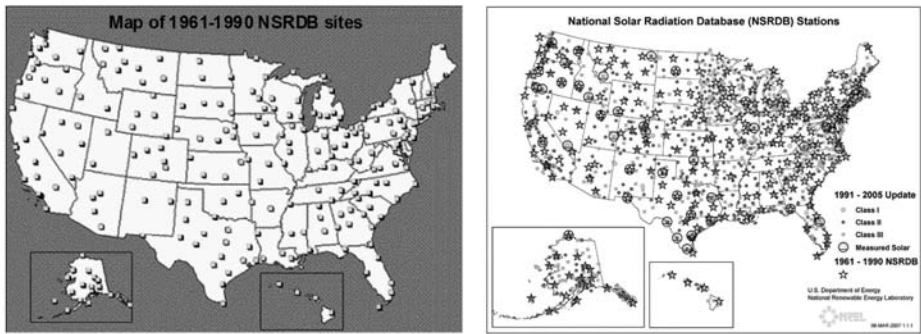


Figure 1: NSRDB sites (source: NREL)

developed and normally used as a representative of weather conditions at that location in building simulation programs. TMY is composed of hourly weather data such as temperature, humidity, solar radiation, wind speed and wind direction for 12 months. Each month was selected from multi-year database using statistical methods, with the condition that it represents the most typical weather pattern of that month. TMY2 (Typical Meteorological Year version 2) provided from the NSRDB is generated from 30 years of data from 1961 to 1990 and available for 239 locations in the United States. Statistical methods were used to select the typical month based on nine daily weather values such as daily maximum, minimum, and mean dry bulb temperature, the maximum and mean wind velocity and the total global horizontal solar radiation. The latest data set called TMY3 (Typical Meteorological Year version 3) contains weather information including solar radiation data at 1,454 ground weather stations throughout the country. Typical month weather data were selected from a 15-year database from 1991 to 2005.

Some simulation tools have embedded or built-in weather data that are derived from TMY2, for example, PVWATTS 1.0. When using these tools, the nearest available location in the tool is normally used to represent the site weather condition. TMY2 is not available for Ann Arbor. Therefore, Detroit TMY2 is always used to represent Ann Arbor typical weather. Some programs allow weather data inputs. In this case, TMY3 which is a newer version of TMY2 and is available for Ann Arbor can be used. However, the Ann Arbor TMY3 data are collected at Ann Arbor municipal airport which is located in the southern area of the city.

1.3 SOLARANYWHERE® SATELLITE-DERIVED WEATHER DATA

Recently, specific location weather data sets derived from satellite images have become available. Examples of satellite derived weather data are available from SolarAnywhere® (Clean Power Research, 2010) and 3Tier (3TIER Inc., 2010). Weather data sets derived from satellite images, for example, SolarAnywhere® data sets includes hourly global horizontal irradiance (GHI), direct normal irradiance (DNI), wind speed, and ambient temperature estimated for the specified location. These data from 1998-2007 are available for free. More recent data, as well as seven-day forecast data, are available for a fee. The spatial resolution of the data is available at approximately 10 km x 10 km in the form of satellite grid tiles (Figure 2). Real-time data like those from SolarAnywhere® provide an ability to obtain real-time PV system efficiency.

Solar radiation data generated by the traditional models METSTAT and NRCC, and solar radiation data generated by the satellite based model, SUNYA, were evaluated in 2005 by Myers et al. (Myers et al., 2005). The results show that the performance of these models was remarkably similar. However, when the distance between the site and the ground weather station is more than 34 kilometers, the solar radiation data derived from satellite images using algorithms like SUNYA are more accurate than using the nearest weather station data or the interpolation data between stations (Perez et al., 1997).

SolarAnywhere®

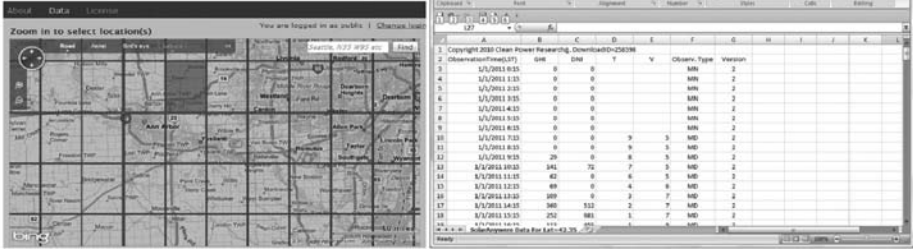


Figure 2: SolarAnywhere® data set grid tiles and outputs

1.4 SOLAR ADVISORY MODEL (SAM)

SAM is a program available free of charge from the National Renewable Energy Laboratory (NREL) (National Renewable Energy Laboratory, 2010). It is a standalone renewable energy system performance and economics simulation program. This tool was developed by the NREL in collaboration with Sandia National Laboratories and in partnership with the Solar Energy Technologies Program (SETP), U.S. Department of Energy. The full version was first available in 2006. For PV system performance, SAM can model a range of solar energy technologies including crystalline silicon (cSi), thin film (CdTe, CiS and aSi) concentration photovoltaic (CPV), multijunction concentrator photovoltaic (mj-CPV), and heterojunction with an intrinsic thin layer (HIT). Within SAM, there are options of sub- simulation models to choose from (Figure 3). Plane of array (POA) solar radiation models available in SAM are isotropic sky, Hay and Davies, Reindl, Perez 1998 and Perez 1990. Array performance models, implemented using the TRNSYS program as a simulation engine, are the Sandia model, CEC performance model, simple efficiency model and concentrating PV model. The result from the array performance model is direct current (DC) electricity output produced from PV arrays. There are two inverter models available: the Sandia model and single-point efficiency model. The result from the inverter model simulation is alternate current (AC) electricity output that the inverter converted from DC output.

The models selected in this study are the Perez 1990 solar radiation model, the Sandia array model and Sandia inverter model. The Perez 1990 model is the update of the Perez 1988 model and is widely used in simulation programs to calculate solar irradiance falling onto surfaces. With the global horizontal and direct normal irradiance data from weather data sets, total solar irradiance falling on a tilted surface can be computed. The Perez model is well validated by many researchers

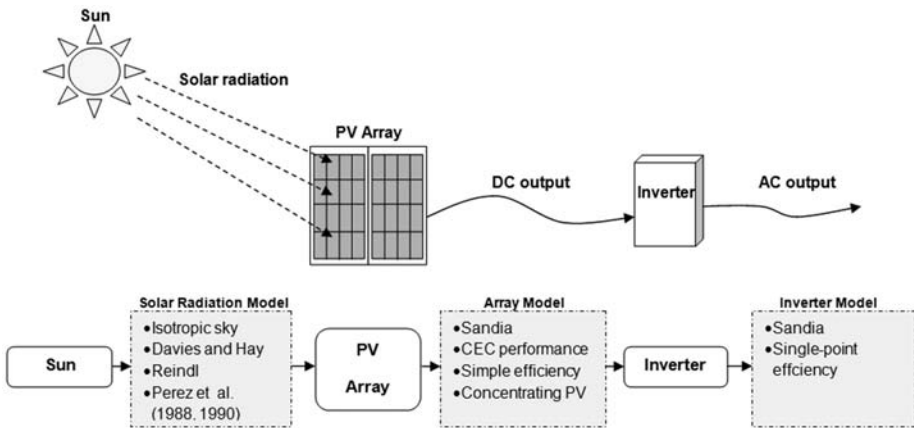


Figure 3: Diagram of mathematic models available in SAM

(Gueymard and Myers, 2009, Loutzenhiser et al., 2007). The Sandia PV Array performance model and the Sandia Inverter performance model use theoretical and semi-empirical methods. They utilize databases of empirically derived parameters developed by testing commercial PV modules and inverters in actual conditions. Validation of the models has been tested against measurement data and other models (Fannee et al., 2009). Figure 4 shows inputs needed for each model, their outputs and their relationships.

Weather data are needed in the solar radiation model and array model. Global horizontal irradiance and direct normal irradiance are used to compute the amount of total solar irradiance falling onto PV modules, while ambient temperature and wind speed are used to compute PV surface temperature, which will affect the amount of DC electricity output. Derate factors for the Sandia array model are efficiency reduction due to mismatch loss, wiring, diodes and connections, soiling, and module degradation. Derate factors for the Sandia inverter model are from the wiring and transformer.

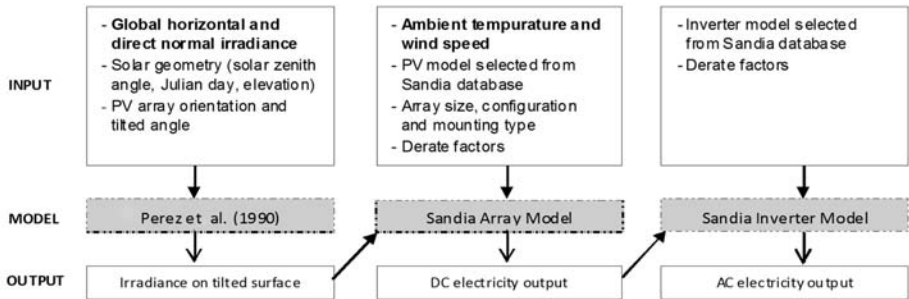


Figure 4: Inputs and output in PV system performance simulation

2.METHODS

An experimental study has been carried out to compare PV performance prediction using SAM as a simulation program and four different weather data sets for Ann Arbor, Michigan. These weather data sets are Detroit TMY2, Ann Arbor TMY3, Ann Arbor 15-year NSRDB (1991-2005) and Ann Arbor 13-year SolarAnywhere® (1998-2010).

2.1 PV OUTPUT PERFORMANCE PREDICTION VALIDATION

Uncertainties exist in every step of the calculations for obtaining a solar constant outside of the atmosphere, calculating the irradiance amount at the earth's surface, translating that amount into the irradiance falling on tilted surfaces, simulating how much DC electricity is generated from PV systems with available irradiance, and simulating how much AC electricity is converted from DC electricity output (Figure 5). In this study, a PV system installed on the roof of the Nature House at the Leslie Science and Nature Center (LSNC) was used as a validation system (Figure 6). LSNC is located in the northern area of Ann Arbor. SAM was used to simulate electricity output from the 2.5 kW DC First Solar Cadmium Telluride (CdTe) photovoltaic system. The DC electricity output was fed into SWR 2500U inverter. The simulation period was from May 2009 to April 2010 when the data of actual hourly PV output from the system was available for validation. The weather data set used in the simulation was SolarAnywhere® satellite derived weather data. The simulated PV electricity output from SAM compared to actual output shows only a 3% error rate. The root mean square error (RMSE) of the simulated output is 0.35 kW. Examples of PV simulated output compared to actual output are shown in Figure 6.

2.2 PV PERFORMANCE PREDICTION EXPERIMENT USING VARIOUS WEATHER DATA SETS

PV output prediction simulated using SAM and SolarAnywhere® was used as a base case to compare with three other weather data sets in terms of their availability, variability and uncertainty.

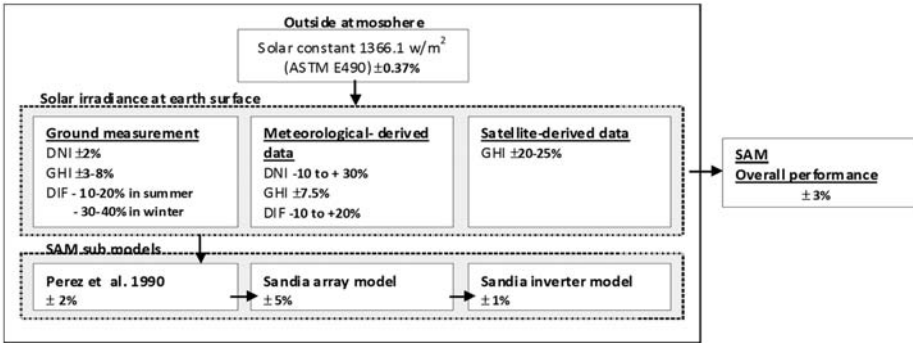


Figure 5: Uncertainties in each step of PV output calculations, solar constant (ASTM Standard, 2006), ground measurement (Gueymard and Myers, 2009), meteorological-derived data (Maxwell, 1998), satellite-derived data (Perez et al., 1997) and SAM sub-models (Cameron et al., 2008) were compared to overall performance in this study.

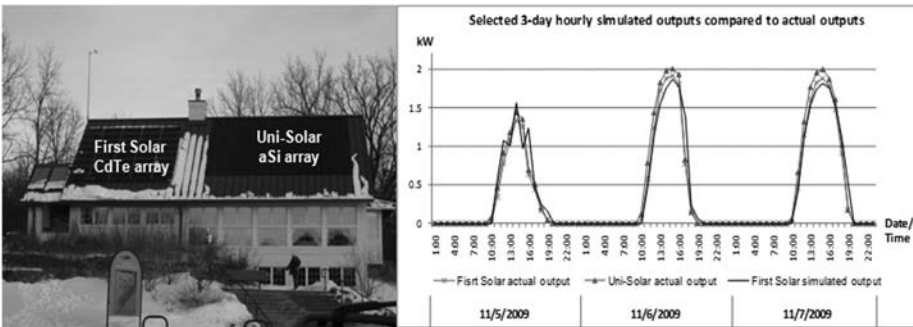


Figure 6: PV systems installed at LNC. SAM predicts PV output within a 3% error compared to actual output

3.RESULTS AND ANALYSIS

3.1 AVAILABILITY

Availability can be defined as the amount of electricity generated from PV systems. This information can be used to select the PV system size as well as the electrical equipment that will be used in the building utilizing the PV system. Using the Ann Arbor TMY3 weather data set, PV electricity output would be 2,684kWh/year. If this number is used, the actual PV yearly output using other weather data sets can vary from -28% (1997 NSRDB data set) to +17% (2002 SolarAnywhere® data set) (Figure 7).

It is obvious from the results that extreme weather that occurs occasionally such as volcano eruptions or unusual weather patterns, is not recognized by TMY weather data sets.

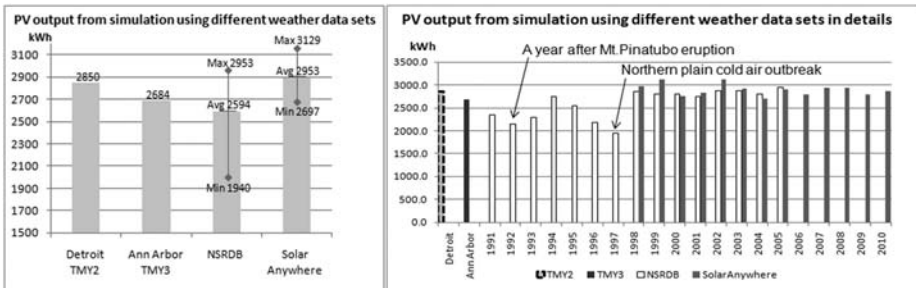


Figure 7: PV yearly electric output from simulation using different weather data sets.

3.2 VARIABILITY

Variability can be defined as how spread out or clustered the data is. In PV output prediction, this can be viewed as how much the PV output per day in each month can vary. Since the NSRDB and SolarAnywhere® are comprised of multi-year data, the maximum and minimum daily average of each month is available. The standard deviation is normally used to indicate variability. It refers to the average distance of data points from the data set average value. The higher the value, the more variable the data. The NSRDB and SolarAnywhere® have similar trends while TMY data sets have more variations. Figure 8 shows daily average PV output in each month, as well as daily PV output standard deviation of each month, from all four data sets.

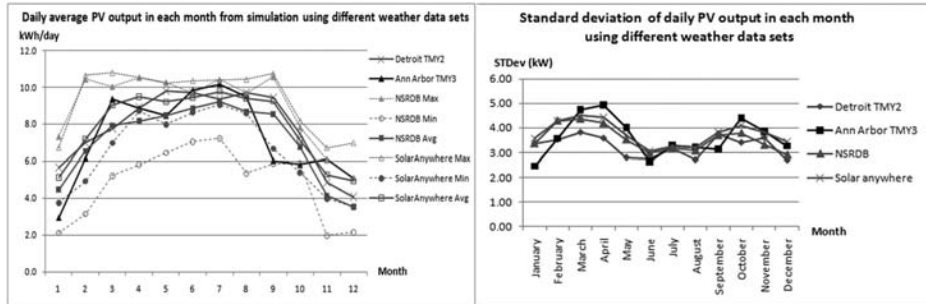


Figure 8: Daily average PV electricity output in each month using different weather data sets

3.3 UNCERTAINTY

Uncertainty can be defined as the chances of each possible event. In this study, each possible PV electricity output from simulations is presented using histograms. The x-axis represents the amount of electricity in kW from a smaller value on the left hand side the larger value on the right hand side. The y-axis represents the frequency of each electricity output value. Yearly histograms of PV electricity output from the NSRDB and SolarAnywhere® in Figure 9 show that shape of the histograms differ from year to year. It can also be seen that the shapes of the histograms are different depending on the weather data sets used. The normal distribution curves show dynamic and variations of weather pattern at the very same locations. The use of average weather data sets or Typical Meteorological Year weather data sets cannot capture these variations. This can indicate that a typical year weather data set might not be proper for predicting PV system electricity output.

Uncertainty can also be defined as an estimate of error. Root mean square error (RMSE) is typically used to evaluate the goodness of the prediction against actual data, especially in the atmospheric field (Yorukoglu and Celik, 2006). RMSE is used to calculate an average magnitude of error. When comparing PV predicted output using TMY weather data sets with PV output from 1998-2010, generated using SolarAnywhere® weather data sets, the difference between the former prediction with the latter yearly prediction is squared. The average of total squared values is then square rooted to find RMSE. The percent RMSE of PV predicted outputs using Ann Arbor TMY3 and Detroit TMY2 compared to those using SolarAnywhere weather data sets were calculated and are shown in Table 1. Even though the yearly RMSE is 4% in Ann Arbor TMY2 and 8% in Ann Arbor TMY3, the monthly percent RMSE values are quite high. They are ranging from 6% to 31% for Detroit TMY2 and 5% to 79% for Ann Arbor TMY3.

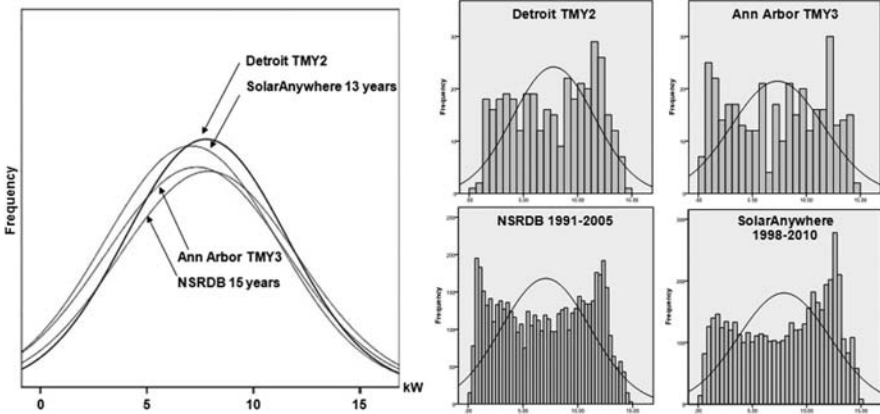
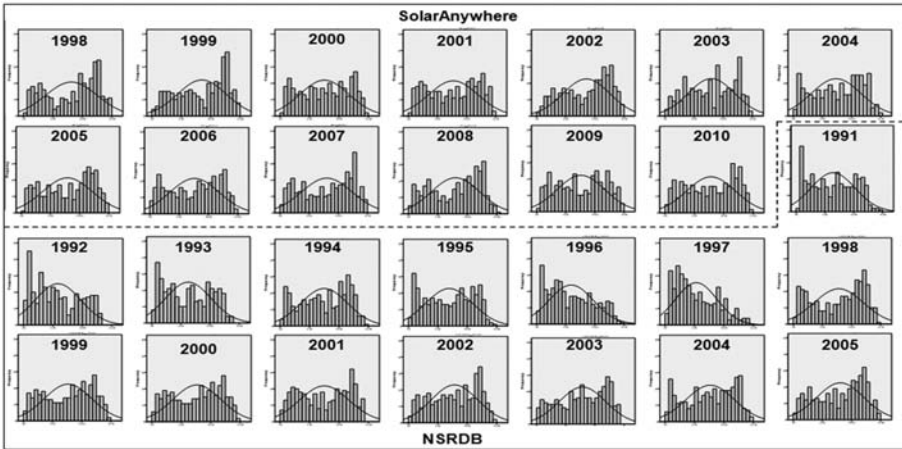


Figure 9: Normal distribution curves and histograms of simulated PV electricity outputs using four weather data sets

Month	Detroit TMY2				Ann Arbor TMY3			
	Total output		Daily average		Total output		Daily average	
	kWh	%RMSE	kWh	%RMSE	kWh	%RMSE	kWh	%RMSE
January	175.3	18%	5.7	16%	91.4	79%	2.9	79%
February	195.4	20%	7.0	21%	172.1	28%	6.1	28%
March	240.8	21%	7.8	20%	289.7	11%	9.3	11%
April	266.3	9%	8.9	9%	265.2	9%	8.8	9%
May	303.7	9%	9.8	10%	261.6	12%	8.4	12%
June	290.7	6%	9.7	6%	294.6	7%	9.8	7%
July	289.7	6%	9.3	6%	314.4	6%	10.1	6%
August	300.9	6%	9.7	6%	294.2	5%	9.5	5%
September	283.2	13%	9.4	18%	180.8	57%	6.0	57%
October	231.7	12%	7.5	13%	180.3	27%	5.8	27%
November	145.8	18%	4.9	18%	182.6	18%	6.1	18%
December	126.1	31%	4.1	29%	157.7	19%	5.1	19%
Average	237.5	13%	7.8	14%	223.7	21%	7.3	21%
Year	3,087.0	4%			2,908.3	8%		

Table 1: RSME for total electricity output and daily average output predicted using TMY weather data sets

CONCLUSIONS

The TMY weather data set is widely used in building simulation tools. However, for particular system performance predictions such as electricity output from PV systems, the TMY weather data format might not be suitable because it cannot capture the availability, variability and uncertainty of solar power that can vary from day to day and from year to year. For Ann Arbor, the availability of actual PV system outputs can vary from -28% to +17% compared to traditional use of TMY3 weather data set. Multi-year weather data sets can capture variation and uncertainty of PV outputs from year to year, while TMY data sets can give only typical performance without a hint of other possibilities. Moreover, both the NSRDB and SolarAnywhere® multi-year data sets show a similar trend in variation and uncertainty even though they are available at different periods. This information is essential for PV system design and operation. When behavior and magnitude of availability, variability, and uncertainty are known, management and operation options can be properly prepared. If multi-year weather data sets are available, they should be used to provide detailed information and better understanding of system performances related to that specific location. The study methodology in this particular location can be used to evaluate the impact of using different weather data sets to predict electricity output from PV systems at different locations and climate zones.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Richard Perez for the Ann Arbor SolarAnywhere® weather data sets, Leslie Science and Nature Center, and PowerDash LLC for access to the PV system actual performance data of the systems that are installed at the LSNP.

REFERENCES

- 3TIER Inc. 2010. 3TIER :: *Renewable energy information services :: Wind, solar, hydro data* [Online]. Available: <http://www.3tier.com/en/> [Accessed December 2010].
- ASTM Standard 2006. *ASTM E490-00a solar constant and zero air mass solar spectral irradiance tables*. West Conshohocken, PA: ASTM International.
- Cameron, C. P., et al. 2008. *Comparison of pv system performance-model predictions with measured pv system performance*. 33rd IEEE PVSC. San Diego, CA.
- CIE TC 3-07 1994. *CIE 108 - 1994 guide to recommended practice of daylight measurement*: International Commission on illumination.
- Clarke, J. A. 2001. *Energy simulation in building design*, Oxford ; Woburn, Mass., Butterworth-Heinemann.
- Clean Power Research. 2010. *SolarAnywhere* [Online]. Available: <https://www.solaranywhere.com> [Accessed December 2010].
- Dean, S. R. 2010. *Quantifying the variability of solar pv production forecasts*. In: The SOLAR 2010 National Conference Phoenix, Arizona.
- Fanney, A. H., et al. 2009. Comparison of predicted to measured photovoltaic module performance. *Journal of Solar Energy Engineering*, 131, 021011-10.
- Gueymard, C. A., et al. 2009. Evaluation of conventional and high-performance routine solar radiation measurements for improved solar resource, climatological trends, and radiative modeling. *Solar Energy*, 83, 171-185.
- Loutzenhiser, P. G., et al. 2007. Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation. *Solar Energy*, 81, 254-267.
- Maxwell, E. L. 1998. METSTAT--the solar radiation model used in the production of the national solar radiation data base (nsrdb). *Solar Energy*, 62, 263-279.
- Myers, D. R., et al. 2005. *Broadband model performance for an updated national solar radiation database in the united states of america*. In: The ISES 2005 Solar World Congress, Orlando, Florida.
- National Renewable Energy Laboratory. 2010. *NREL: Solar advisor model (SAM)* [Online]. Available: <https://www.nrel.gov/analysis/sam/> [Accessed December 2010].
- Perez, R., et al. 2002. A new operational model for satellite-derived irradiances: Description and validation. *Solar Energy*, 73, 307-317.

- Perez, R., et al. 1997. Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data. *Solar Energy*, 60, 89-96.
- Storck, P., et al. 2010. *Integration of short-duration ground based measurements with long-duration satellite based measurement and numerical weather simulations to provide a more accurate assessment of solar energy availability and variability*. In: The SOLAR 2010 National Conference Phoenix, Arizona.
- USGBC Research Committee 2008. A national green building research agenda.
- Yorukoglu, M., et al. 2006. A critical review on the estimation of daily global solar radiation from sunshine duration. *Energy Conversion and Management*, 47, 2441-2450.