Photovoltaic (PV) Modeling with Bias Correction Using Demonstration Site Data in the Southeastern US

Jonathan O. Allen
Allen Analytics LLC
3444 N Country Club Rd. Suite 100
Tucson, AZ 85716-1200
jon@allen-analytics.com

Ranran Wang
Allen Analytics LLC
3444 N Country Club Rd. Suite 100
Tucson, AZ 85716-1200
ranran@allen-analytics.com

Will Hobbs
Southern Company Services
600 N. 18TH ST.
BIN 14N-8195
Birmingham, AL 35203
whobbs@southernco.com

ABSTRACT

PV modules were deployed at a demonstration site in metro Atlanta in 2009. We implemented a suite of models in MATLAB and evaluated the performance of key radiation models and PV array models using onsite PV measurements. Radiation models had comparable performance in calculating the total incident radiation. The Sandia PV model out-performed the 5-parameter PV model in accurately calculating cell temperature and power. Modeled PV power output was corrected for bias. Residual analysis indicated that an apparent source of bias in the PV models was the underprediction of incident total radiation associated with cloudy days in the Southeastern US. We calculated bias-corrected power output for 14 years for each PV module using historical data, and examined the temporal variability of PV outputs. Power output calculated using TMY data appeared to be biased slightly high relative to the mean of the 14 year time series.

1. INTRODUCTION

Photovoltaic (PV) output depends on the local solar insolation and weather, and so is highly variable. In order to make PV investment decisions, one needs to predict the PV output and how this variable output will contribute to system-wide generation.

Meteorological data sets and engineering models of PV systems have been assembled by Department of Energy laboratories and private firms in order to predict PV output. Various studies have been undertaken to implement these models and validate their performance. Typical meteorological year (TMY) aggregated weather data are widely used for solar power prediction. Satellite-based solar data have also been validated and used in the PV modeling [1]. The performance of various radiation models for computing the solar radiation on inclined solar panel surfaces were studied (e.g. [2]), as well as various PV system performance models (e.g. [3]). A standardized approach was proposed to validate the PV performance models [4].

The System Advisory Model (SAM) created by the National Renewable Energy Laboratory (NREL) has been widely used in PV value estimation, PV technology comparisons, and system monitoring. A MATLAB toolbox was recently developed [5]. NREL SAM and its component models have been tested extensively at sites in the Southwestern US, and the accuracy of predictions for sites in the Southeastern US is unknown.

In this paper, we evaluated the deployment of photovoltaic (PV) generation capacity at a demonstration site in metro Atlanta, GA. The installed PV modules included a range of technologies: monocrystalline silicon, polycrystalline silicon, and several different thin film technologies. Multiple suites of radiation models and PV performance models were compared. Residual analysis was carried out on multiple levels in order to improve model accuracy and to further explore the sources of model inaccuracy. Finally, the long-term variability of PV outputs was analyzed using simulations.

2. DATA AND MODEL SPECIFICATIONS

2.1 PV Demonstration Sites

Seven PV modules were deployed at a demonstration site located in metro Atlanta, GA (TABLE 1). The nominal capacity of each PV array was approximately 4 kW. Arrays were south-facing with a tilt of approximately 10 degrees. Extensive data were collected including PV performance and meteorological data for over 2 years. The primary goals of the demonstration project were to evaluate the output of prospective PV installations and data-model combinations to predict PV output in the Southeastern US.

The PV data included time series of output variables of photovoltaic panels, strings and some key meteorological

measurements from an on-site weather station (WS). For each of 7 PV modules average DC power, DC voltage, and average AC power were measured at the inverter. Cell temperature was recorded by probes on the back of modules. The main WS recorded key meteorological measurements including solar insolation, ambient temperature, and wind speed. A reference cell was installed to provide calibrated electrical output. The plane-of-array (POA) incident radiation from the reference cell was used for model comparisons. The original sampling frequency for all variables was 15 minutes. Hourly averages were calculated and used for model assessment.

TABLE 1: DEMONSTRATION PV ARRAYS

Array	Solar Cell Technology	PV Model
A1	Monocrystal	Sandia
A2	Monocrystal	Sandia
A3	Polycrystal	Sandia
A4	Monocrystal	Sandia
A5	Thin Film Amorphous	Sandia
A6	Thin Film	CEC
A7	Thin Film	CEC

2.2 Solar and Meteorological Data

Inputs to PV models generally require hourly data for Global Horizontal Irradiance (GHI), Direct Normal Irradiance (DNI), Diffuse Horizontal Irradiance (DHI), ambient temperature, and wind speed. Insolation of PV cells was calculated from GHI, DNI, and DHI. In addition, temperature and wind speed are used to estimate the cell temperature of the PV modules.

Multiple sources of input solar and meteorological data including the satellite data, ground level weather station, and onsite measurements were used in this paper. Solar and meteorological data from 1998 through 2011 at Atlanta site were provided by a commercial source for solar irradiance time series data. These solar irradiance data were generated from satellite data,

Hourly meteorological measurements and modeled solar values from TMY3, “typical year” data created by the National Renewable Energy Laboratory (NREL), were used in this study. In addition, satellite-based TMY data for “DNI average months” (TMY-DNI) and “GHI average months” (TMY-GHI). The months for these TMYs were chosen as the month with GHI or DNI irradiance closest to the average over the period 1998 through 2011.

In this study, GHI ground measurements were provided for a nearby weather station. Eight months of the GHI ground measurement at the central Atlanta area overlapping with the PV demonstration site were used in the validation study.

2.3 PV Output Modeling

The System Advisory Model (SAM) created by NREL integrates state-of-art meteorological data sets and engineering models of PV systems in order to model power outputs. We encoded the key component models in NREL SAM as MATLAB functions (“SAM in MATLAB”) in order to perform computations not readily available in the NREL SAM system; these included simulations using multiple years of actual meteorological data, and large numbers of component model runs.

Multiple radiation models were used to calculate POA radiation, including the isotropic sky model [6], the Hay-Davis-Klutcher-Reindl (HDKR) model [7-9] and the Perez model [10]. These 3 radiation models were implemented in SAM in MATLAB and compared. All incident radiation data on the PV array were adjusted by shading and soling factors. Two major PV system performance models, the Sandia [11] and the CEC "Five Parameter" [12] PV models were implemented and compared. The Sandia inverter performance model [13] was used to calculate AC power output. Both DC and AC power outputs were adjusted by the SAM default derate factors.

3. RESULTS

3.1 Model Selection

Results from SAM in MATLAB were compared with NREL SAM implementations to verify the agreement on hourly and monthly time scales. The model outputs from SAM in MATLAB were in close agreement with those from NREL SAM. The relative errors of AC power were below 1% for Sandia PV array model and approximately 1.5% for the CEC PV model. The NREL SAM version used here was 2011.12.2.

We examined the total radiation incident to a PV panel (E_T) to evaluate three radiation models: the isotropic sky, HDKR, and Perez models. The incident radiation estimations from three radiation models were compared with the reference cell at the demonstration site. Root

mean square errors (RMSE) were calculated, and the rank for three radiation models was Perez < Isotropic < HDKR (TABLE 2).

TABLE 2: RMSE VALUES OF ARRAY A1 COMPARING RADIATION AND PV MODELS.

Model Set	Hourly Data			
	E_T (kW/m^2)	T_C (C)	P_{DC} (kW)	P_{AC} (kW)
IsoSky+Sandia	0.107	5.093	0.361	0.347
HDKR+Sandia	0.108	5.198	0.361	0.347
Perez+Sandia	0.104	4.968	0.362	0.348
IsoSky+CEC	0.107	8.563	0.351	0.34
HDKR+CEC	0.108	8.881	0.347	0.335
Perez+CEC	0.104	8.461	0.345	0.334
	Monthly Data			
IsoSky+Sandia	0.041	2.917	0.068	0.068
HDKR+Sandia	0.038	3.06	0.057	0.056
Perez+Sandia	0.034	2.772	0.055	0.054
IsoSky+CEC	0.041	6.159	0.117	0.116
HDKR+CEC	0.038	6.659	0.105	0.105
Perez+CEC	0.034	6.072	0.099	0.099

The relative mean absolute errors (RMAE) were also calculated as

$$RMAE = \frac{\frac{1}{n}\sum_{i=1}^{n}\left|X_{Model,i} - X_{Demo,i}\right|}{\frac{1}{n}\sum_{i=1}^{n}X_{Demo,i}} \quad (1)$$

where $X_{Model,i}$ and $X_{Demo,i}$ were modeled and measured values respectively. The rank of RMAE was HDKR < Perez < Isotropic (TABLE 3). However, the differences in error rates among the three radiation models were within 1% of total radiation. Comparing the error rates for monthly output data, the Perez model consistently returned more accurate estimates of total radiation. Therefore, the Perez model was preferred in total radiation calculation as suggested in the NREL SAM documentation.

TABLE 3: RMAE VALUES OF ARRAY A1 COMPARING RADIATION AND PV MODELS.

Model Set	Hourly Data			
	E_T (%)	T_C (%)	P_{DC} (%)	P_{AC} (%)
IsoSky+Sandia	18.50	9.27	15.86	15.96
HDKR+Sandia	17.47	8.73	15.78	15.84
Perez+Sandia	17.72	9.02	15.85	15.89
IsoSky+CEC	18.50	15.87	17.74	17.97
HDKR+CEC	17.47	15.35	17.28	17.50
Perez+CEC	17.72	15.67	17.07	17.29
	Monthly Data			
IsoSky+Sandia	9.60	9.77	4.28	4.42
HDKR+Sandia	8.25	9.71	3.53	3.68
Perez+Sandia	8.12	9.25	3.40	3.51
IsoSky+CEC	9.60	20.89	8.98	9.37
HDKR+CEC	8.25	21.54	8.10	8.50
Perez+CEC	8.12	20.57	7.62	8.02

We compared measured cell temperatures (T_C) of PV modules at the Atlanta site with cell temperatures calculated from the Sandia and CEC thermal models. The Sandia thermal model performed better than the CEC thermal model in terms of both RMSE and RMAE (TABLE 2 and TABLE 3). The estimates of cell temperature of both Sandia and CEC thermal models were in general biased low.

The performance of the Sandia and CEC PV models were evaluated by comparing DC and AC power (P_{DC} and P_{AC}) to the measurements from the Atlanta site. Since measured DNI and DHI were not available for PV array model validation, all radiation and weather input variables were from the satellite time series data. With the same radiation and weather inputs, the Sandia PV model performed better than CEC PV model in terms of

producing smaller RMSE and RMAE errors (TABLE 2 and TABLE 3).

Comparisons of modeled and measured values for other modules at the same site also supported the conclusion that Sandia models outperformed the CEC models in estimating cell temperature and power output. We selected Perez, Sandia Thermal, Sandia PV as the preferred model set. For modules for which Sandia model parameters were not available, the selected model set was Perez, CEC Thermal, and CEC PV.

3.2 Module Performance

We generated PV output predictions for the 2-year study period, and compared the modeled power outputs with the onsite measurements. Five out of 7 modules at the demonstration site had negative mean bias errors (MBE) ranging from -0.57% to -10% (see TABLE 4); in these cases the model underestimated the power output. Two modules had positive bias errors of 17% and 29%.

TABLE 4: MODEL PERFORMANCE EVALUATION WITH MBE, RMSE AND RMAE (HOURLY AND MONTHLY).

PV Array	MBE (%)	RMSE (kW)		RMAE (%)		d_0
		Hour	Mon.	Hour	Mon.	
A1	-0.57	0.34	0.05	15.92	3.51	1.01
A2	-10.2	0.38	0.14	19.42	10.51	1.11
A3	-5.67	0.36	0.08	18.17	6.05	1.06
A4	-6.85	0.37	0.11	18.01	7.37	1.07
A5	17.22	0.42	0.20	26.02	17.14	0.85
A6	-0.71	0.34	0.05	16.45	3.52	1.01
A7	29.10	0.46	0.27	36.14	28.81	0.77

The bias in AC power for each module can be considered a combination of site-specific and module-specific effects. The main site-specific effect on AC power biases was the bias error for estimated total radiation, which was -7.31% at the Atlanta site. Thus the recommended radiation models appear to have systematically under-estimated solar radiation at the Atlanta site. The module-specific biases were likely due to inaccurate PV model coefficients. PV module manufacturers likely understate the nominal power in order to ensure that modules generate at least the specified power. This is consistent with the negative bias in AC power output observed for most modules. For Arrays A5 and A7, measured power outputs were systematic lower than predicted power outputs; this was likely due to misspecification of the PV capacity parameters. It was also later found that 1 of 8 DC strings on Array A7 had a blown fuse for the majority of the study period.

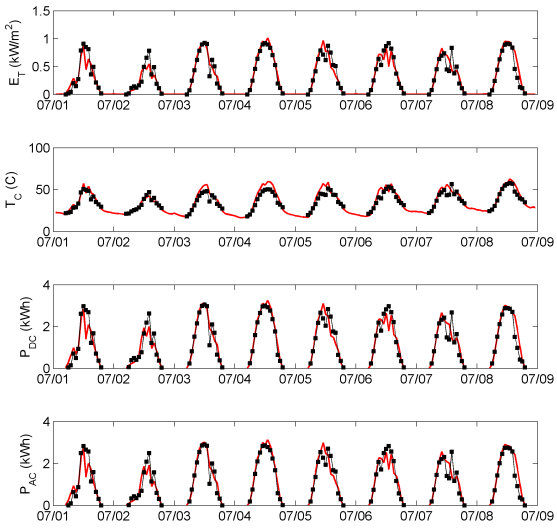

Fig 1: Measured (red line) and modeled (black line with dots) PV system outputs (E_T , T_C , P_{DC} and P_{AC}) for Array A4.

The comparison of measured and modeled intermediate variables for Array A4 (Fig 1) showed that the residual in PV power output tended to follow the same pattern as the residual in POA radiation. Cell temperature was systematically underestimated compared to actual cell temperature. In order to determine the source of differences between modeled and measured PV power, we analyzed the residuals of inputs and modeled intermediate values with respect to AC power output.

4. RESIDUAL ANALYSIS AND SIMULATIONS

4.1 PV Bias Correction

In order to remove the systematic bias in calculated PV output, we multiplied the calculated PV output by a scaling coefficient, d . The MBE of the calculated PV output was then determined for d ranging from 0.5 to 1.5. MBE for each PV array was a linear function of bias coefficient d (Fig 2). The recommended coefficient, δ_0 , was then selected for each PV array such that the MBE was zero. Coincident with the reduction in MBE, the

RMAE of the calculated PV output was reduced to approximately its minimum value for each PV array.

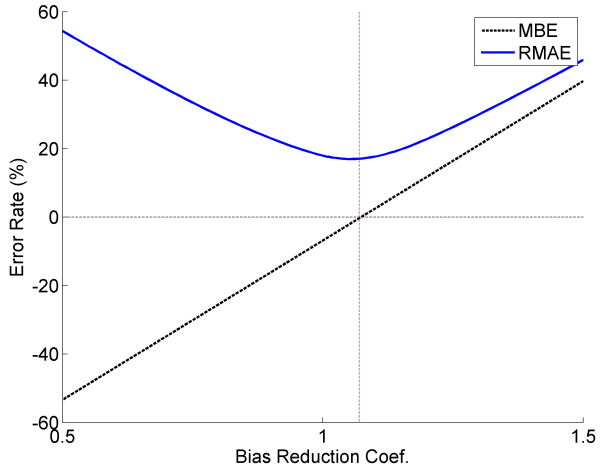

Fig 2: MBE (dashed line) and RMAE (solid line) of calculated PV output as functions of bias-correction coefficient d for Array A4

The residual plots indicated that bias correction improved model performance most significantly during clear days in summer and winter when PV energy production was relatively high. However, the modeling error inherent in the solar radiation inputs could not be reduced by correcting the PV array capacity bias. For Array A4, RMAE equaled 18% when MBE was zero, and this coincided with the RMAE of total radiation E_T. PV output RMAE was approximately equal to total radiation RMAE for all modules after the capacity bias corrections; this suggests that much of the error in PV models (separate from solar radiation models) can be removed by using a single scaling factor based on actual PV array performance.

4.2 Residual Analysis

In order to further understand the cause of differences between measurements and model calculations, we undertook detailed residuals analyses for Array A4, because the reference PV cell was located adjacent to it, and this module had typical MBE, RMAE, and RMSE values. A 6-month dataset was used which included complete PV and nearby weather station measurements. In this subsection, the modeled PV output has been scaled by d_0.

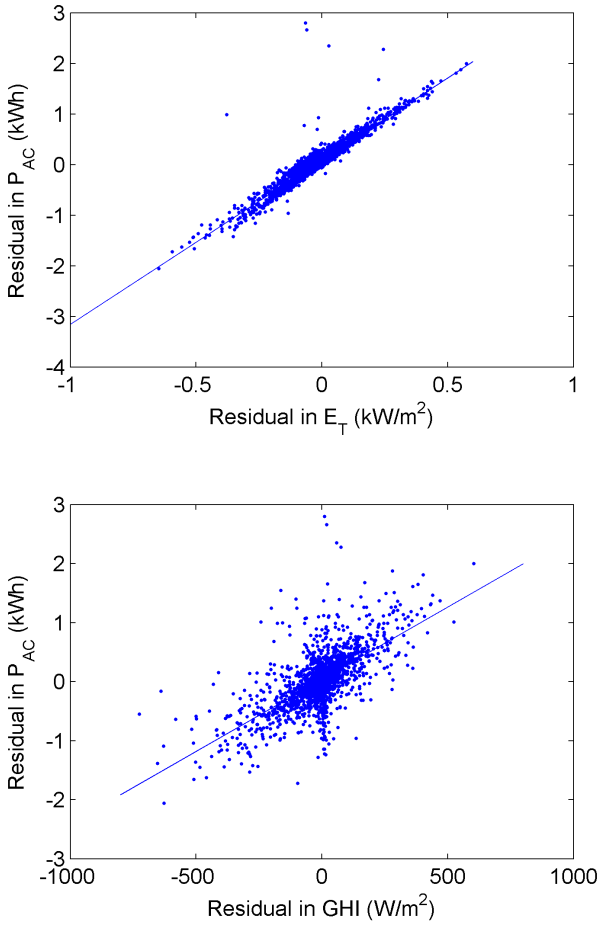

Fig 3: Residual comparisons for Array A4.

Residuals in P_{AC} had a strong positive linear relationship with residuals in E_T which suggests that the variability in modeled power output was mainly caused by errors in radiation estimation (Fig 3). Residuals in P_{AC} plotted against the residuals in GHI, the only measured component of solar radiation, also showed a strong positive correlation (Fig 3). Here, the residuals in GHI were calculated as the difference between satellite-based GHI and GHI measured at the local weather station. GHI and DNI data measured at the demonstration site were not available though. The correlation of GHI with E_T suggests that the accuracy in the calculation of E_T was largely affect by the accuracy of the input solar values.

We further investigated the relations between residuals in AC power and other variables, including zenith angle, sky clearness, ambient temperature, and relative humidity. Sky clearness indices were calculated as

$$e = \frac{\frac{I_{diff} + I_{dn}}{I_{diff}} + 1.041\left(\frac{\pi q_z}{180^0}\right)^3}{1 + 1.041\left(\frac{\pi q_z}{180^0}\right)^3} \quad (2)$$

where q_z is the zenith angle, I_{diff} and I_{dn} are the horizontal diffuse irradiance and direct normal irradiance respectively [10]. Higher values of sky clearness index indicate clearer skies.

Low zenith angles were associated with relatively high power residuals (Fig 4). Residuals tended to be relatively large and negative for overcast hours and the discrepancy in power was as much as 1.5 kW. Relatively high sky clearness values were associated with overestimated power. The largest residuals in power were associated with high ambient temperatures and moderate relative humidity (Fig 4). One explanation for high variability in power when ambient temperature was high and zenith angle was low is that power output itself was relatively large under these conditions, i.e. midday in summer.

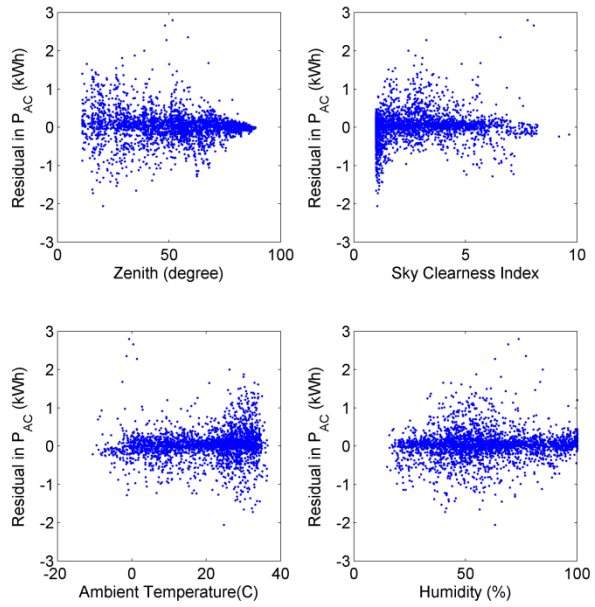

Fig 4: Residual comparisons for Array A4: residuals in power against zenith angle (upper left), sky clearness index (upper right), ambient temperature (lower left) and humidity (lower right).

In conclusion, the main source of variability in PV power output was the input radiation data. Power outputs were most inaccurate and variable during warm and clear days, when solar radiation was also high. These periods likely correspond to partly cloudy conditions during sunny seasons when solar radiation was highly variable spatially and the solar radiation modeled for an area by satellite data did not match exactly the solar radiation at the site. In addition, the satellite-based radiation data appears to underestimate GHI on overcast days at the demonstration site. However, without complete onsite solar measurements, including GHI, DNI and DHI, it was not possible to validate radiation models for local conditions.

4.3 Decomposition of Sandia PV model

The Sandia PV performance model [11] estimated the current and voltage as functions of the effective radiation (E_e) and cell temperature (T_C) at the maximum power point as:

$$I_{mp} = I_{mp,ref} \times (C_0 E_e + C_1 E_e^{\,2})[1 + a_{Imp}(T_c - T_0)] \quad (3)$$

$$V_{mp} = V_{mp,ref} + C_2 N_c \delta(T_c)\log(E_e) + C_3 N_c [\delta(T_c)\log(E_e)]^2 + b_{Vmp}(E_e)(T_c - T_0) \quad (4)$$

Here, C_0, C_1, C_2, C_3 were empirically determined and taken from the NREL SAM parameter library.

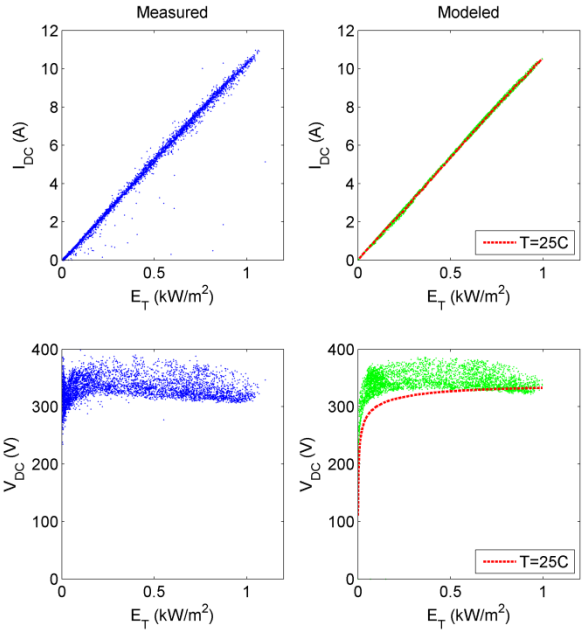

Fig 5: DC current and DC voltage as functions of POA radiation for Array A4 for the onsite measurements (left panels) and the model outputs after bias correction (right panels).

Both measured and modeled PV output showed that DC current scaled linearly with total irradiance; while the DC voltage scaled nonlinearly (Fig 5). Cell temperature had little impact on modeling DC current and coefficient a_{Imp} was close to zero. Onsite measurements showed larger variation than the model output, because the onsite solar

radiation measures were more variable than the satellite-based solar data. Applying the bias correction helped adjusting the fits between solar irradiance and current/voltage respectively. Using a large number of onsite measurements recorded on both clear and cloudy/overcast days, one could likely refit the Sandia PV model parameters accurately.

4.4 Simulations of Long-Term PV Output

We calculated PV output from the entire satellite solar and meteorological time series from 1998 through 2011, as well as the satellite TMY-GHI data, satellite TMY-DNI data and NSRDB TMY3 data. The Perez model was used to convert GHI, DNI, and DHI into solar irradiance on tilted PV panels. The Sandia thermal and Sandia PV model were used to predict cell temperature, output DC voltage and DC power if parameters for the Sandia PV model were available. The CEC thermal and CEC PV models were used if parameters for the Sandia PV model were not available. The Sandia inverter model was applied to calculate AC power. The 14-year time series of PV output predictions along with intermediate predictions including total irradiance, effective irradiance, cell temperature, DC voltage, DC power, were generated. The PV output predictions for TMY-GHI, TMY-DNI and TMY3 input data were also calculated. For each PV array, the PV power output has been scaled by its corresponding bias-correction coefficient.

Estimated annual PV power output for the 7 PV arrays over the years 1998 through 2011 showed a relative range of approximately 15% (Fig 6). The variability of annual energy productions over 14 years was similar among 7 PV arrays at the demonstration site, because the main source of variability was the input solar radiation and the nominal capacity of each module assembly was the same (Array A7 was the only exception).

We compared the annual energy predicted from long-term solar and meteorological data with the annual energy predicted using TMY-GHI, TMY-DNI and TMY3 input data. The annual energy values predicted from TMY3 deviated from those predicted from TMY-GHI and TMY-DNI by 1.5 to 3.5% (Fig 6). It also appeared that TMY3 was not as representative as the TMY-GHI and TMY-DNI when comparing with long-term prediction data. However, this may be an artifact that the time series data and the satellite TMY data had the same source data, whereas NREL TMY3 were developed using a different algorithm and time period.

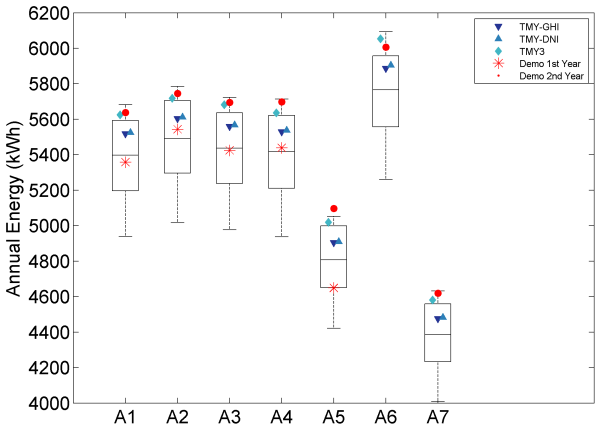

Fig 6: Distribution of annual energy production from the seven PV arrays estimated using the satellite time series (box plots) and typical meteorological years (triangles). Measured annual energy production values are shown as asterisks and dots.

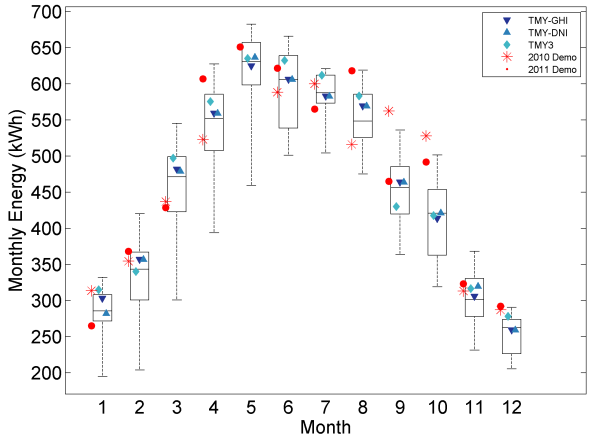

Fig 7: Distribution of monthly energy production from Array A4 simulated using the satellite time series (box plots) and typical meteorological years (triangles). Measured monthly energy production values are shown as asterisks and dots.

We also examined the predicted energy generated for each month for the years 1998 through 2011. For each of the 7 PV modules, we used the 14-year samples to assess the variability of monthly energy output, and compared them with actual energy output as well as energy output predicted based on TMY data (Fig 7).

We observed high variability in monthly energy predictions during the spring and fall, i.e., February

through May and September through October. The highest estimated solar energy production was during May. High levels of solar energy were predicted to have been generated during July and August with relatively low variability during these months.

We compared the monthly energy predicted using TMY-GHI, TMY-DNI and TMY3 input data and the actual energy measurements. The monthly energy predictions were close to the measurements for June through August. The component models appeared to underestimate monthly energy during September and October. These comparisons are site-specific but may demonstrate typical variance between TMY and long term averages.

5. CONCLUSIONS

Among the models considered here, we found that the combination of the Perez radiation model, Sandia PV, Sandia thermal, and Sandia inverter models return the most accurate PV output estimation. The CEC model may be used if parameters for the Sandia PV model are not available.

Bias correction using measured PV output resolved the underestimation of modeled PV output that was partially due to the mismatch between actual and nominal module capacities.

Another apparent source of bias in the PV models was the underestimation of incident total radiation associated with cloudy days. This bias may be systematically important for PV modeling in the Southeastern US. This bias may be corrected by either reparameterizing radiation models or by statistical post-processing of predictions from other models.

The usage of long-term satellite time series data was recommended to calculate historical hourly PV output and generate distributions of predicted PV output for PV siting decisions. TMY data may also be used for siting decisions, but the calculated PV output does not include variability in PV output and may be biased slightly high.

6. REFERENCE

1. Stein, J.S., R. Perez, and A. Parkins, *Validation of PV Performance Models Using Satellite-based Irradiance Measurements: A Case Study*, in *SOLAR 2010*2010: Phoenix, AZ.
2. Loutzenhiser, P.G., et al., *Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation.* Solar Energy, 2006(81): p. 254-267.
3. Cameron, C.P. and W.E. Boyson, *Comparison of PV system performance model predictions with measured PV system performance*, in *33rd IEEE PVSC*2008: San Diego, CA.
4. Stein, J.S., et al., *A standardized approach to PV system performance model validation*, in *37th IEEE PVSC*2011: Seattle, WA.
5. Stein, J.S., *The Photovoltaic performance modeling collaborative (PVPMC)*, 2012: Sandia Report SAND2012-4531.
6. Liu, B.Y.H. and R.C. Jordan, *The interrelationship and characteristic distribution of direct, diffuse and total solar radiation.* Solar Energy, 1960. **4**(3): p. 1-19.
7. Klutcher, T.M., *Evaluation of models to predict insolation on tilted surfaces.* Solar Energy, 1979. **23**(2): p. 111-114.
8. Hay , J.E., Davies, J. A. *Calculations of the solar radiation incident on an inclined surface*. Proc. of First Canadian Solar Radiation Data Workshop, 1980.
9. Reindl, D.T., Beckmann, W.A., Duffie, J.A., *Evaluation of hourly tilted surface radiation models.* 1990b(1): p. 9-17.
10. Perez, R., et al., *Modeling Daylight Availability And Irradiance Componants From Direct And Global Irradiance.* Solar Energy, 1990. **44**(5): p. 271-289.
11. King, D.L., W.E. Boyson, and J.A. Kratochvil, *Photovoltaic Array Performance Model*, 2004, Sandia National Laboratories: Photovoltaic System R&D Department, Albuquerque, NM. p. 41.
12. De Soto, W., *Improvement and Validation of a Model for Photovoltaic Array Performance*, in *Mechanical Engineering*2004, University of Wisconsin-Madison: Madison, Wisconsin. p. 124.
13. King, D.L., et al., *Performance Model for Grid-Connected Photovoltaic Inverters*, 2007, Sandia National Laboratories: Albuquerque, New Mexico. p. 47.