Received: 4 April 2023 | Revised: 20 June 2023 | Accepted: 29 June 2023

DOI: 10.1002/pip.3729

RESEARCH ARTICLE

PROGRESS IN PHOTOVOLTAICS WILEY

Blind photovoltaic modeling intercomparison: A multidimensional data analysis and lessons learned

Marios Theristis[1] | Nicholas Riedel-Lyngskær[2] | Joshua S. Stein[1] | Lelia Deville[1] | Leonardo Micheli[3] | Anton Driesse[4] | William B. Hobbs[5] | Silvana Ovaitt[6] | Rajiv Daxini[7] | David Barrie[8] | Mark Campanelli[9] | Heather Hodges[10] | Javier R. Ledesma[11] | Ismael Lokhat[12] | Brendan McCormick[13] | Bin Meng[14] | Bill Miller[10] | Ricardo Motta[15] | Emma Noirault[16] | Megan Parker[10] | Jesús Polo[17] | Daniel Powell[18] | Rodrigo Moretón[19] | Matthew Prilliman[6] | Steve Ransome[20] | Martin Schneider[21] | Branislav Schnierer[22] | Bowen Tian[14] | Frederick Warner[23] | Robert Williams[24] | Bruno Wittmer[25] | Changrui Zhao[26]

Correspondence
Marios Theristis, Sandia National Laboratories, Albuquerque, NM, USA.
Email: mtheris@sandia.gov

Funding information
U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), Grant/Award Number: 38267

Abstract

The Photovoltaic (PV) Performance Modeling Collaborative (PVPMC) organized a blind PV performance modeling intercomparison to allow PV modelers to blindly test their models and modeling ability against real system data. Measured weather and irradiance data were provided along with detailed descriptions of PV systems from two locations (Albuquerque, New Mexico, USA, and Roskilde, Denmark). Participants were asked to simulate the plane-of-array irradiance, module temperature, and DC power output from six systems and submit their results to Sandia for processing. The results showed overall *median* mean bias (i.e., the average error per participant) of 0.6% in annual irradiation and −3.3% in annual energy yield. While most PV performance modeling results seem to exhibit higher precision and accuracy as compared to an earlier blind PV modeling study in 2010, human errors, modeling skills, and derates were found to still cause significant errors in the estimates.

KEYWORDS
blind comparison, modeling, performance, photovoltaic, yield modeling

1 | INTRODUCTION

Photovoltaic (PV) energy yield estimates are used to assign value to and obtain financing for a PV plant. Such predictions inform power plant designs, investor decisions, and cash flow in financial models and therefore can affect the viability of a project.[1] Although the performance models used for these predictions are generally presumed to be relatively simple, energy yield estimations are not straightforward

For affiliations refer to page 1155.

because their accuracy depends on the analysis and modeling pipeline, which commonly include irradiance transposition, module temperature, and power output modeling. Different models and their combinations may result in varying accuracies, and different assumptions for performance loss factors (derates) will significantly affect the energy yield estimations. These may also depend on the PV plant configuration, module type and geographical location. Furthermore, the modeler's skills and experience can affect the resulting accuracy.

New PV performance models are continuously being developed whereas existing models are frequently updated. However, only a limited number of them have been evaluated independently from multiple aspects against high-quality field datasets (e.g., in previous studies[2–7]). When an approach is tested against known datasets, the modeler might introduce a bias, which directly influences the approach's validity, reproducibility, and applicability for different systems. In such cases, blind intercomparisons are useful for benchmarking analysis pipelines and establishing the state of the art in PV performance modeling. The PV Performance Modeling Collaborative (PVPMC) was founded based on the outcomes of the blind PV modeling study in 2010.[8,9]

Previous intercomparisons of PV modeling approaches include that of Friesen et al.,[10] wherein time-series plane-of-array irradiance (Gpoa) and module temperature (Tmod) data were circulated to eight European institutions. These participants were asked to simulate the module-level performance of five PV technologies in seven climates, and it was found that the group's energy yield predictions agreed with ±5%. Moser et al. analyzed the long-term yield predictions of six expert modelers for a PV system in an Italian and an Australian site.[11] These modelers were required to independently obtain meteorological data for their simulations, which, for the Italian site, led to 6% differences in global horizontal irradiance (GHI), 20% differences in Gpoa, and ultimately nearly 30% differences in AC energy. And most recently, Vogt et al.[12] conducted an intercomparison of energy rating calculations per IEC 61853-3[13] with nine European institutes. Energy rating differences of 14% were found in the first blind comparison round. It took five rounds of calculations—and discussions among the participants—for the nine participants' calculations to agree within 0.1%. Ultimately, Vogt et al.[12] demonstrated how user-induced variability can be reduced when modelers have clear procedures for implementing key steps of the PV model chain.

To provide an opportunity for PV modelers to test their models and modeling ability against high-quality, real system data and to help provide a baseline quantifying the variability of different models and modelers, PVPMC organized a new blind PV performance modeling comparison in 2021. Measured weather and irradiance data and detailed descriptions of six PV systems from two locations (Albuquerque, New Mexico, USA, and Roskilde, Denmark) were provided. Participants were asked to simulate the systems' plane-of-array irradiance, module temperature, and DC power output and submit their results back to Sandia for processing. This work compares system-level performance modeling considering all DC-side loss factors. Rather than independently obtaining meteorological data for the simulations, participants were provided with measured meteorological and irradiance data as a starting point. This provision enabled the propagation of sources of uncertainty within the modeling pipeline instead of the results being affected by the uncertainty of the input data. Furthermore, this study was open to anyone (i.e., industry, research, and academia) to participate, rather than inviting specific individuals. As such, this article presents the multidimensional data analysis of the PVPMC blind modeling intercomparison, providing the results for each modeling step. Finally, it summarizes the lessons learned and areas where improvements are needed.

2 | METHODOLOGY

The blind PV modeling comparison was announced in July 2021 through the PVPMC email list (https://public.govdelivery.com/accounts/USDOESNLEC/subscriber/new?topic_id=USDOESNLEC_185). The data and document describing the exercise were downloaded >600 times. Sandia received 29 submissions from 28 participants with various modeling pipelines, including new commercial software. This effort represents 26 institutions from 12 different countries.

2.1 | Scenarios and data

For this comparison, six scenarios of practical interest to the community were identified and included (a) fixed and tracking systems, (b) monofacial and bifacial modules, (c) modules representative of the current PV market and upcoming technologies, and (d) distinctively different geographical locations/climates (see Table 1).

In Albuquerque (S1, S2), GHI was measured using a Kipp and Zonen CMP-21 pyranometer. Kipp and Zonen CH1 and Eppley normal incidence pyrheliometers (NIP) were used to measure DNI. To measure the diffuse horizontal irradiance (DHI), two Eppley precision spectral pyranometers (PSP) were used, one having a shade disk and the other having a shade band. The Gpoa was measured using a Kipp and Zonen CMP-11 pyranometer. Wind speed was measured at 10 m above ground level using a Climatronics Wind Mark III Wind Sensor. Air temperature was measured using two Climatronics Aspirated Shield Temperature Sensors. Module temperature was measured using back-of-module resistance temperature detectors (RTDs), on one module of each string. Voltage and current were measured at the string level for all systems using voltage dividers and Manganin shunts. The Roskilde systems and measurements setup in scenarios 3–6 are described by Riedel-Lyngskær et al.[14]

The participants had access to general instructions, hourly averaged weather data from the locations (Gpoa was not included), module and inverter spec sheets, system designs, and test reports. The test reports were only available for the systems in Albuquerque (i.e., S1 and S2) and included IEC 61853-1[15] matrix data, IEC 61853-2[16] incidence angle modifier (IAM), nominal module operating temperature (NMOT), and PAN (*Panneau Solaire*) files. A

TABLE 1 Characteristics of the six scenarios used in this blind modeling intercomparison. These were selected to include (a) fixed and tracking systems, (b) monofacial and bifacial modules, (c) modules representative of the current PV market and upcoming technologies, and (d) distinctively different geographical locations/climates.

	S1	S2	S3	S4	S5	S6
Resolution	Hourly[a]	Hourly[a]	Hourly[a]	Hourly[a]	Hourly[a]	Hourly[a]
Duration	1 year	1 year	1 year	1 year	1 year	1 year
Year of data	2020	2020	2019–2020	2019–2020	2019–2020	2019–2020
Capacity	3.9 kW	3.3 kW	26.84 kW	25.96 kW	26.84 kW	25.96 kW
Module type	Monofacial	Monofacial	Monofacial	Bifacial	Monofacial	Bifacial
Module manufacturer	Panasonic	Canadian Solar	Trina	Trina	Trina	Trina
Module technology	Silicon hetero-junction	Mono-crystalline	Mono-crystalline PERC	Mono-crystalline PERC	Mono-crystalline PERC	Mono-crystalline PERC
Module nominal power	325 W	275 W	305 W	295 W	305 W	295 W
Power temperature coefficient	−0.258%/°C	−0.41%/°C	−0.39%/°C	−0.39%/°C	−0.39%/°C	−0.39%/°C
DC/AC ratio	0.76	0.86	1.07	1.04	1.07	1.04
Tracking	Fixed tilt	Fixed tilt	Single-axis tracking	Single-axis tracking	Fixed tilt	Fixed tilt
Location	Albuquerque, NM, USA	Albuquerque, NM, USA	Roskilde, Denmark	Roskilde, Denmark	Roskilde, Denmark	Roskilde, Denmark
Commissioning date	June 2018	October 2017	August 2018	August 2018	August 2018	August 2018

Abbreviations: PERC, passivated emitter and rear contact.
[a]Hourly averages reported at the end of the hour.

frequently asked questions (FAQs) section was regularly updated on the PVPMC website to ensure everyone had access to the same information. Modeling results were collected and handled by Sandia, ensuring anonymity. The participants know their "participant number" only, and they had the option to exclude their name from any publication. This paper's author numbers and order are unrelated to the participant numbers in the figures.

The participants were asked to simulate Gpoa (in W/m^2), module temperature (in °C), and system DC power output (Pmp in W). Some participants resubmitted their estimates to correct "minor" mistakes such as modeling 48 modules instead of 12, submitting incorrect units (e.g., kWh/kWp or kW instead of W), reporting direct irradiance instead of global, and so forth.

To ensure non-physical irradiance values (i.e., sun below the horizon, below/above physical minimum/maximum, static measurements, and inconsistent irradiance components) are ignored, the datasets were filtered based on version 2 of the Baseline Surface Radiation Network (BSRN) Global Network-recommended quality control tests.[17] Furthermore, datapoints during days with snow were filtered out from both locations. The data from Roskilde include additional filters to ensure the proper operation of single-axis tracking (by comparing the tilt angles of neighboring trackers) and that the data acquisition system is online. All values lower than 100 W/m^2 in front-side Gpoa, 50 W in output DC power, and beyond −5°C and 45°C in ambient temperature were also filtered out.

The validation datasets are available online in open access at two locations. The first is on the website of the PV Performance Modeling Collaborative at https://pvpmc.sandia.gov/. The second is on the Duramat Data Hub at https://datahub.duramat.org/dataset/pv-performance-modeling-data (doi: https://doi.org/10.21948/1970772).[18]

2.2 | Statistics

Based on the participants' affiliations, they were grouped into the following categories: (1) Commercial, (2) Research, (3) Software, and (4) Student. The commercial category includes consulting and engineering companies, independent engineers, owners, utilities, and producers. Figure 1 shows the percentage breakdown per category.

Some models/software can reveal who the participant is when only used once and other models did not achieve an adequate statistical sample. To ensure anonymity and focus on approaches with significant participation, the following categories were created:

1. "Other model" refers to known models used by a single participant (e.g., *pvlib-pvwatts* from *pvlib-python*[19]);
2. "Custom model" refers to "in-house" customized models that are not available to the public (these are models developed and used by some independent engineers);
3. "Other software" includes software used by a single participant (e.g., Archelios,[20] PVSol,[21] and Sisifo[22]);
4. "SDM" includes single-diode models (e.g., *pvlib.pvsystem.calcparams_cec*, *pvlib.pvsystem.calcparams_pvsyst* models from *pvlib-python*).

Figure 2 shows statistics on the models used in this study. With respect to the transposition models, the majority used the Perez

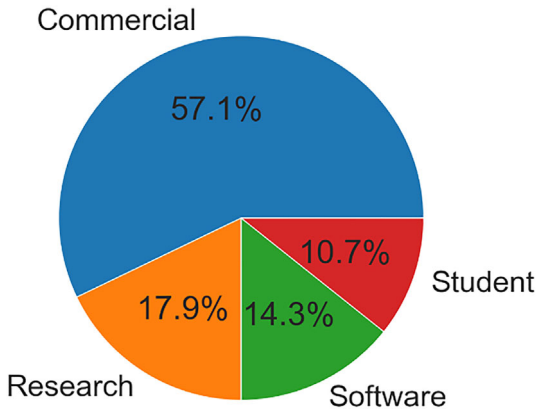

FIGURE 1 Categorization of participants. It includes commercial entities, researchers, software companies, and students. Commercial entities have the following subcategories: consulting and engineering, independent engineer, owner, utility, and producer.

model[23] whereas in the case of temperature modeling, more than 60% of the participants used the PVsyst (Tcell),[24] Sandia Array Performance Model (SAPM),[25] and the Nominal Operating Cell Temperature (NOCT) model.[26] It should be noted here that "PVsyst (Tmod)" refers to the temperature model in PlantPredict[27]; this model is the same as in PVsyst, but it then gets converted to module temperature using the equation developed by Sandia.[25] Finally, close to 50% of the participants used the PVsyst[28] and System Advisor Model (SAM)[26] software packages for PV performance modeling.

As mentioned in subsection 2.1, the S1 and S2 systems included PAN files, IEC 61853-1 matrix, IAM, and NMOT data. As seen in Figure 3, most participants used the PAN files, mainly due to the high percentage of PVsyst users. Only 24.6% of the participants used the provided IEC 61853-1 data. The IAM and NMOT data were used by half of the participants. The percentages in Figure 3 are also categorized as a function of individual models and categories.

3 | RESULTS

To rank the participants, the mean absolute percentage error (MAPE) was used to compare the annual irradiation (Figure 4A) and energy yield (Figure 4B) estimates. The median MAPE in annual irradiation and energy yield estimations were close to 2% and 5.5%, respectively. Interestingly, the participants with the lowest MAPE (<<1%) in the annual irradiation estimation (i.e., P23, P2, and P22) exhibited high MAPE in annual energy yield estimation (ranging from 8.2% to 68.7%; the y-axis limits were truncated to 7% for clarity). Note that not all participants modeled all six scenarios.

3.1 | Irradiance modeling

A PV performance modeling pipeline always begins with irradiance. In this study, the participants had the measured global horizontal, direct normal, and diffuse horizontal irradiances and were asked to apply transposition models to estimate Gpoa.

Figure 5 shows the diurnal front (top row) and rear (bottom row) Gpoa estimates by all participants. One of the participants in S3 appears to have simulated a fixed tilt system instead of tracking. As expected, front Gpoa is not as difficult to predict whereas problems arise when modeling the rear Gpoa, where minimum and maximum differences above 100% were observed. It is worth mentioning that despite these high differences in rear Gpoa, this component represents <10% of the total irradiance.

Another observation is an apparent time-shift in the estimates of some participants. It seems that there is confusion between instantaneous and time-averaged measurements especially when involving sun position. This study reported the hourly averaged irradiance data at the end of the hour. Therefore, most models should assume a sun position calculated 30 min before the hourly timestamp as being the most representative. On the other hand, other data sources commonly place timestamps at the beginning of the interval. As such, some software properly account for this by calculating the solar position and other time sensitive values at the center of the interval (i.e., +30 min). Modeling software, such as PVsyst, make an exception for timestamps that span sunrise or sunset to pick a sun position halfway between the horizon and the sun position at the neighboring daylight timestamp. In this study, some participants seemed to adjust by shifting their time-series by 30 min back, while others kept it at the end of the hour. This is clearly an area where procedures could be standardized.

Empirical cumulative distribution functions (ECDF) present residuals in an ascending order to observe how they are distributed across the datasets. The ECDF plots in Figure 6 show the residuals between modeled and measured Gpoa grouped by the transposition models. The off-pink lines are the individual participants, and the black dashed lines indicate the median residuals per model. A steep rise near zero suggests that there are mostly small model errors and relatively few large ones. Most models except the isotropic indicated good accuracy (median values close to zero). The isotropic model underestimated Gpoa by 11.25 W/m^2. Although the distributions of residuals for most Perez users cluster together, some outlying distributions still exist indicating errors in the implementation of solar position algorithms, system configuration, and the possibility of applying different Perez model coefficient sets, other than the coefficients of the most commonly used set "All sites composite 1990."[29] When comparing residuals against system configuration, there was a slight over-estimation in the single-axis tracking system in S3 (median residual of 6.5 W/m^2) as compared to the fixed tilt systems in S1 and S5 with −1.7 W/m^2 and 0.77 W/m^2, respectively.

3.2 | Module temperature modeling

First, it should be mentioned that although the accuracy of resistance temperature detector (RTD) sensors is typically within 0.1°C,

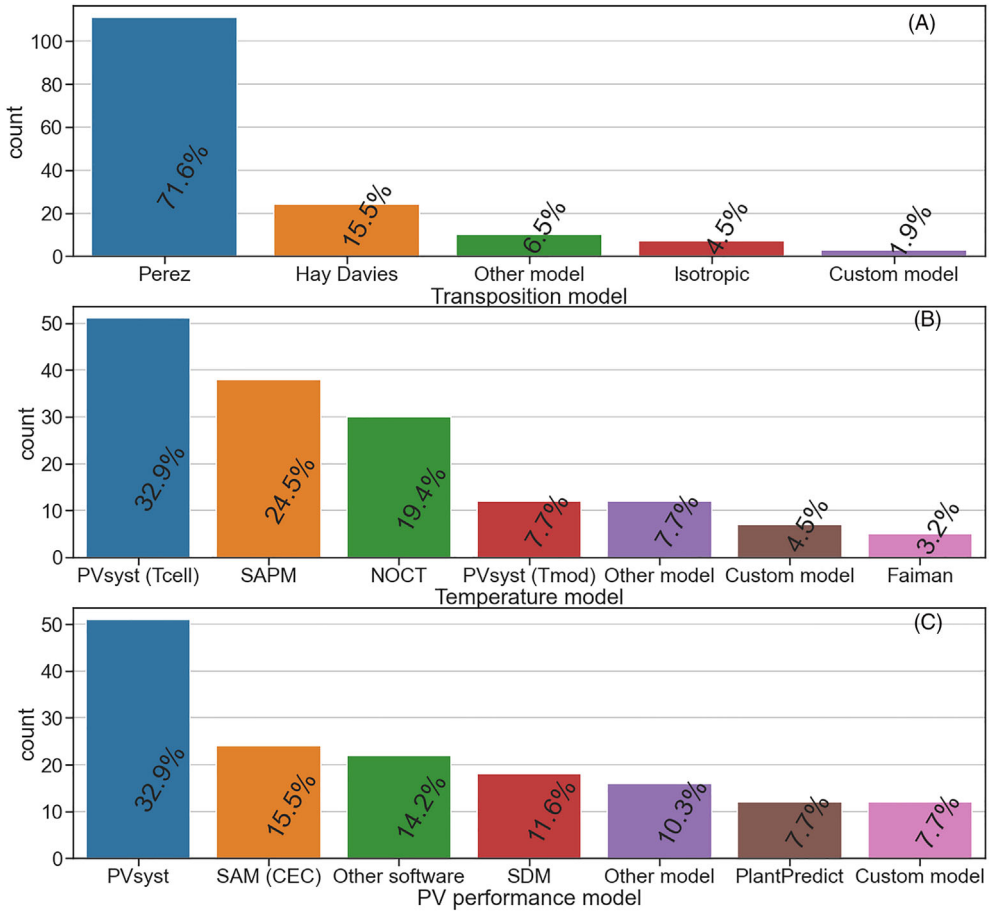

FIGURE 2 Frequency plots displaying the models and categories usage in (A) transposition, (B) temperature, and (C) power modeling. "Other model" refers to known models, which were only used once (e.g., PVWatts); "Custom model" refers to "in-house" customized models that are not available to the public (these are models developed and used by independent engineers); "Other software" includes software only used once (e.g., Archelios, PVSol, Sisifo); "SDM" includes single-diode models (e.g., *pvlib-cec* and *pvlib-pvsyst*).

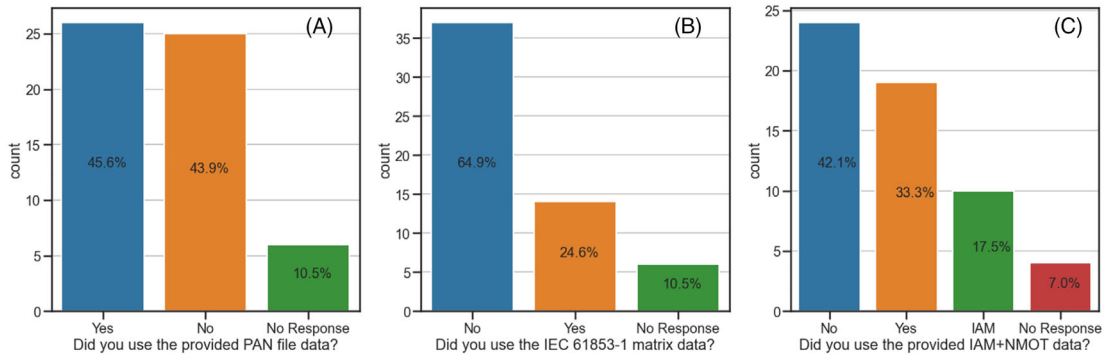

FIGURE 3 Percentage of participants that used (A) PAN file data, (B) IEC 61853-1 matrix data, and (C) IAM + NMOT data in S1 and S2.

it is still not possible to know what the representative temperature for a PV array is, unless an array is equipped with multiple sensors (e.g., one for each solar cell). This is practically and economically not feasible. Therefore, although this study compares with an average module temperature value from four different sensors, the differences reported in this work should not be taken in a strictly quantitative manner.

Figure 7 shows the ECDF plots of the module temperature residuals for three out of six scenarios (due to availability), grouped by all models. It should be noted that some models, such as the one from PVsyst, calculate only cell but not module temperature. The SAPM and PVsyst (Tcell) models exhibited median residuals relatively close to zero. In contrast, the PVsyst (Tmod), NOCT, and "Other model" underestimated the module temperature over 2°C. To put this into perspective, for a module with a typical temperature coefficient of power of −0.4%/°C, a 2°C error leads to a 0.8% error in output power. The median underestimation exhibited by PVsyst (Tmod) was due to the wrong U values the participants used: Instead of the PVsyst U values, the Faiman U0 and U1 values were used; it is speculated that one PVsyst (Tcell)

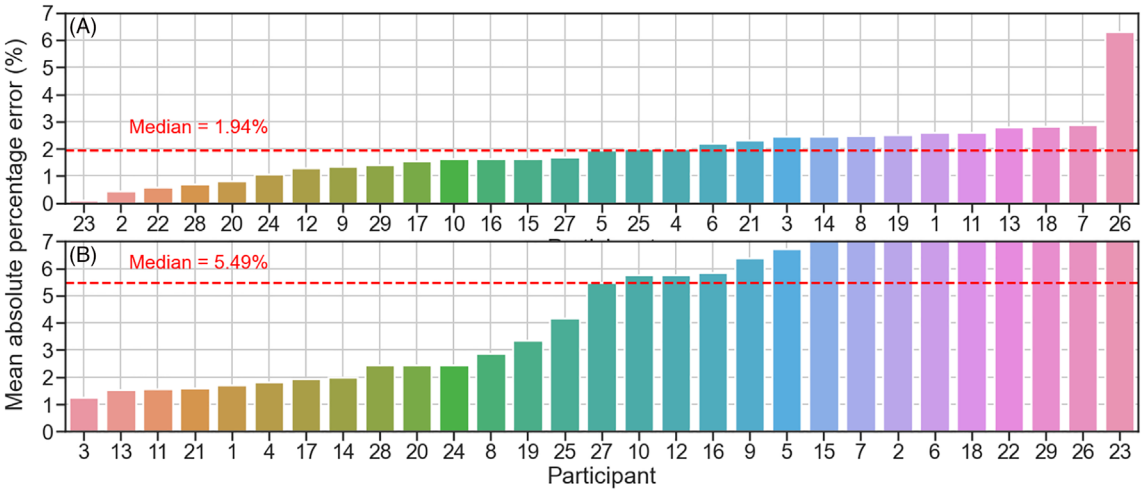

FIGURE 4 Mean (of scenarios) absolute percentage error (MAPE) of annual (A) irradiation and (B) energy yield estimates by each participant in ascending order. The red dashed horizontal lines display the overall median MAPE. The y-axis limits were truncated to 7% for clarity. The absolute percentage error is calculated as $100 \times |(model - measurement)/measurement|$.

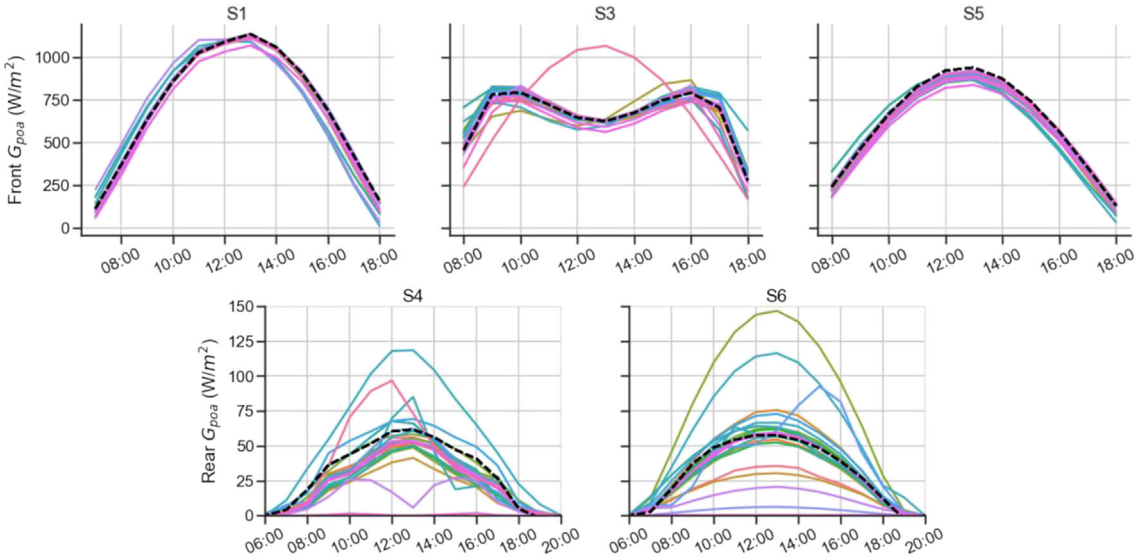

FIGURE 5 Diurnal plane-of-array irradiance measurements against all simulations submitted by the participants during a clear sky day in Albuquerque (April 14: S1) and Roskilde (March 25: S3–S6). The top row shows the front plane-of-array irradiance while the bottom row shows the rear plane-of-array irradiance. The black dashed vertical lines represent the measurements whereas the different colors are the participants' estimations. Minimum and maximum percentage differences from the measured front Gpoa at noon ranged from −11% (S5) to +61.3% (S3); the latter was due to a participant who simulated fixed tilt rather than tracked Gpoa. If this participant is excluded, the minimum and maximum differences would range from −11% to 1.99%. Minimum and maximum percentage differences from the measured rear Gpoa at noon ranged from −99.7% (S4) to +149.4% (S6).

participant made the same mistake. It is worth noticing that all but one of the PVsyst (Tcell) users exhibit nearly identical distributions demonstrating consistent calculations within the most popular software. However, it should be noted that while the median bias of this model is small, the comparison is against module temperatures, while PVsyst (Tcell) only calculates cell temperature.

3.3 | PV performance modeling

Figure 8 shows the ECDF plots of the normalized power residuals for all scenarios, grouped by all model categories. The power residuals were normalized by the system's nominal capacities to allow a meaningful comparison among the different scenarios. Overall, all models

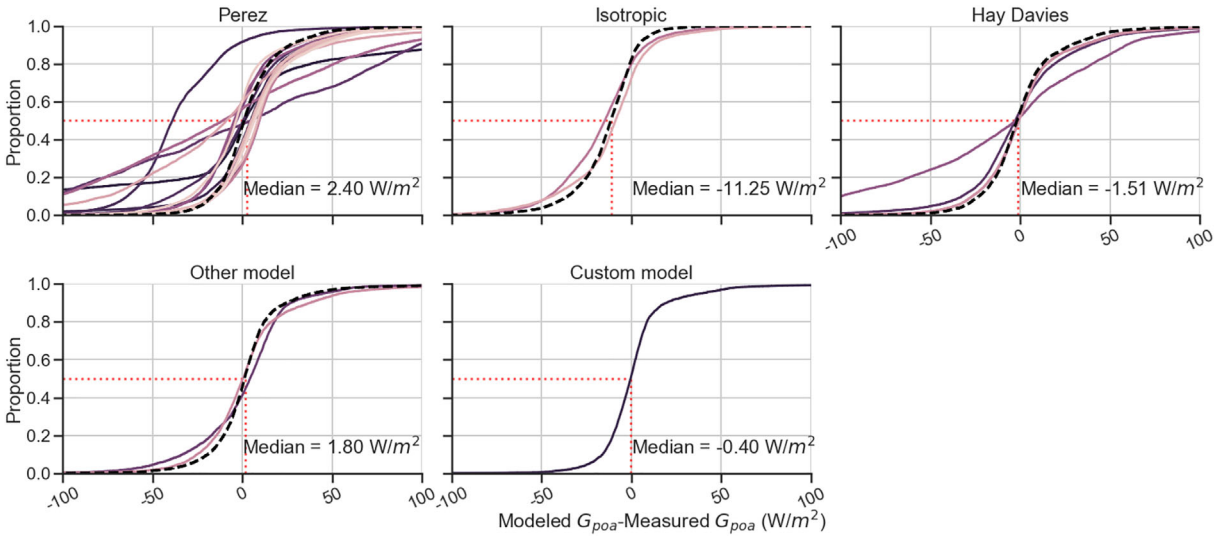

FIGURE 6 Empirical cumulative distributions of the plane-of-array irradiance residuals (in W/m^2) for all scenarios, grouped by the different transposition models. Participants are displayed in different colors, and the dashed black lines indicate the median residuals within the same modeling category. "Other" and "Custom" categories include models that differ within the same category.

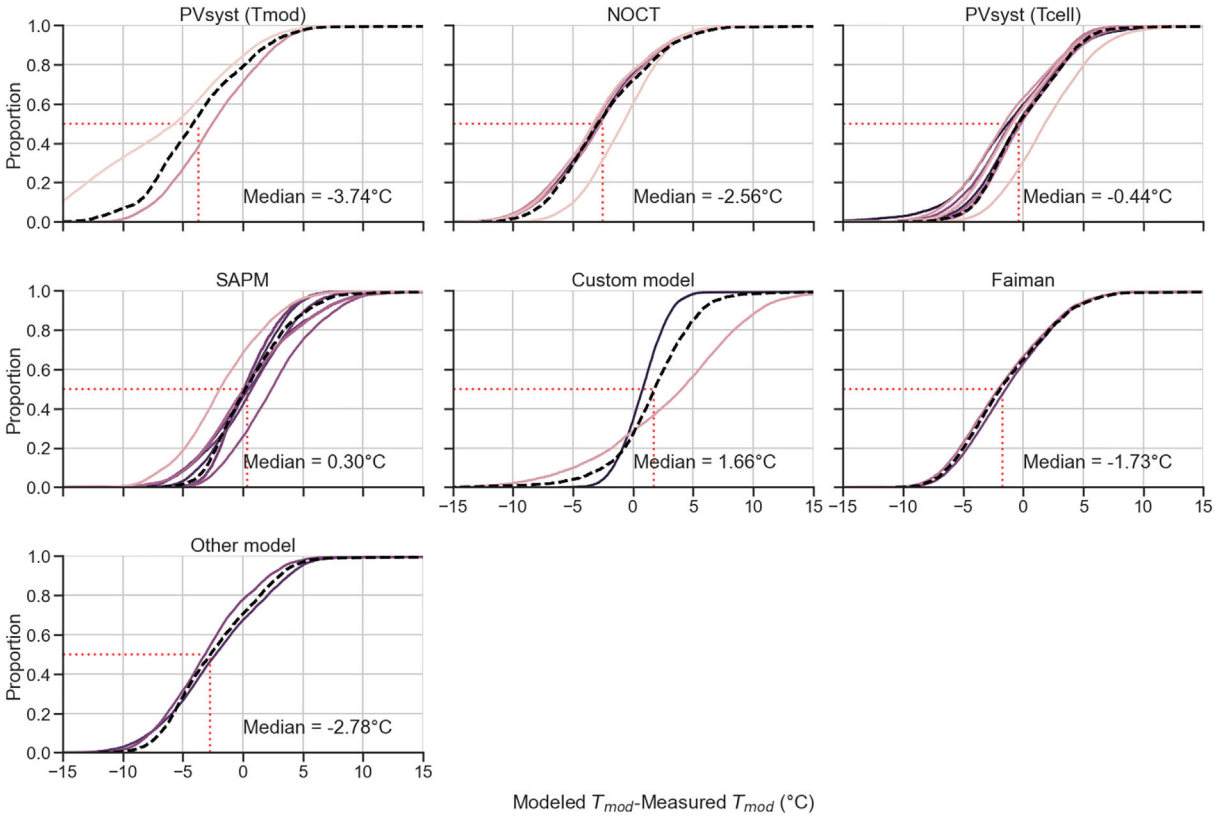

FIGURE 7 Empirical cumulative distributions of the module temperature residuals (in °C) for all scenarios, grouped by the different temperature models. Participants are displayed in different colors, and the dashed black lines indicate the median residuals within the same modeling category. "Other" and "Custom" categories include models that differ within the same category.

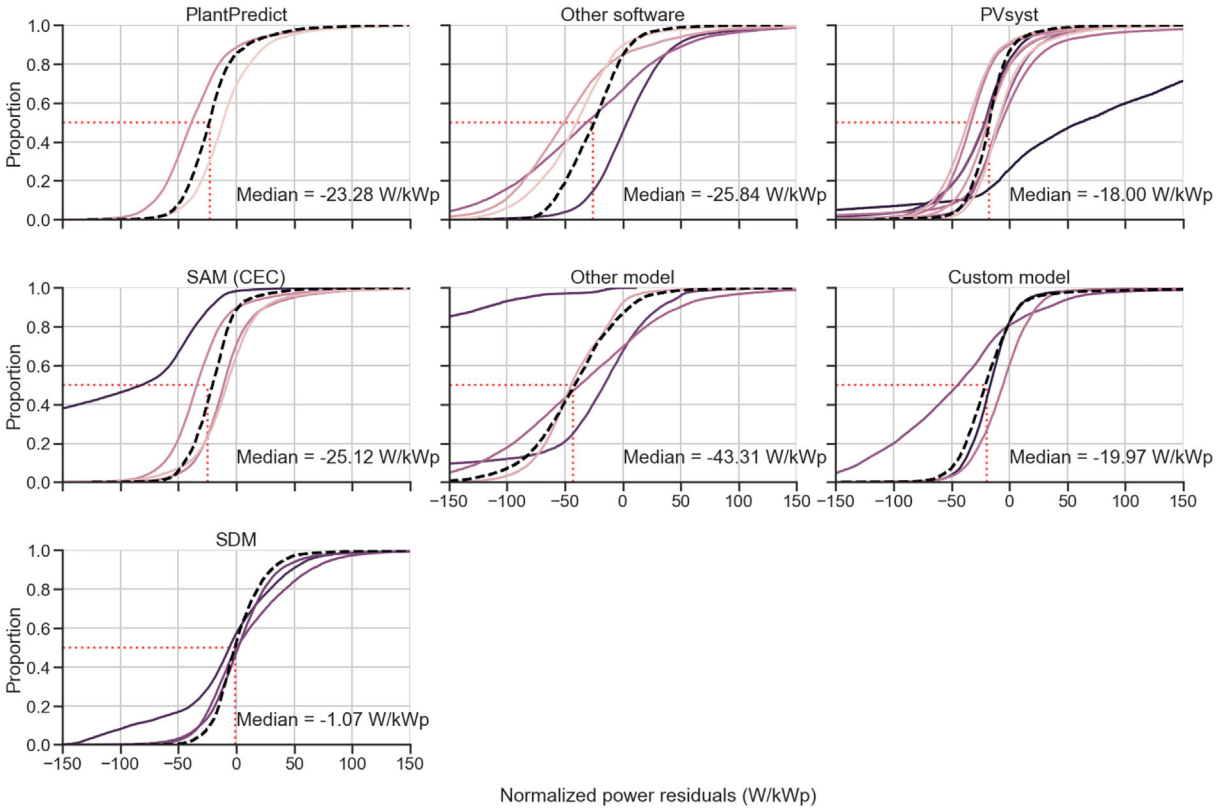

FIGURE 8 Empirical cumulative distributions of the normalized power residuals for all scenarios, grouped by the different models. The power residuals are normalized by the systems' nominal capacities (i.e. W/kWp). Participants are displayed in different color whereas the dashed black lines indicate the median residuals within the same modeling categories. "Other" and "Custom" categories include models/software that differ within the same category.

underestimated the normalized power by up to 43.3 W/kWp (or 4.33%), whereas the SDM category demonstrated superior performance with a bias close to zero (−1.07 W/kWp). PVsyst users, who comprise 33% of the participants, group well together except in one instance. This mass underestimation raises the question of whether this is because of a modeling issue, input selection, or any other assumptions involved within the pipeline. This is further examined in Section 4.

4 | OBSERVATIONS AND LESSONS LEARNED

For the systems in Albuquerque (i.e., S1 and S2), the participants had module information that is not commonly available. PAN files might be available in databases but, in this study, the PAN files were specific to the modules in S1 and S2, that is, not generic representative PAN files. The objective here is to observe how the various assumptions or usage of additional information affected the results and to describe the lessons learned from this study.

It should be expected that as module temperature increases, efficiency will decrease. To examine whether this holds true for all participants' temperature coefficient inputs, these trends were reverse calculated. This was done by taking a subset of data for modeled Gpoa from 800 to 1200 W/m^2 and wind speed lower than 10 m/s and fitting a regression model for modeled power against the module temperature by each participant (see Figure 9). Qualitatively, it can be observed that some participants assumed lower temperature dependency, while others assumed positive temperature coefficients. The latter might be due to an error in applying the sign in the equation; another speculation could be that the participant(s) may have used the temperature coefficient for current, instead of power. Furthermore, some participants miscalculated the system size by either over- or under-sizing the number of PV modules. Therefore, human errors are not uncommon in PV performance modeling. Another common confusion observed during this blind PV modeling comparison was that many participants interchangeably used the U0 and U1 values of the Faiman model with the Uc and Uv values of the PVsyst (Tcell) model. Although these models are similar, they are not the same, and therefore, the U parameters should not be used interchangeably. Methods for parameter translation between temperature models (e.g., translate U0, U1 to Uc, Uv) have been recently published elsewhere,[30] and functions are available in *pvlib-python* under the *pvlib.temperature. GenericLinearModel*.

The modeled temperature rise (i.e., the difference between modeled module temperature and measured ambient temperature)

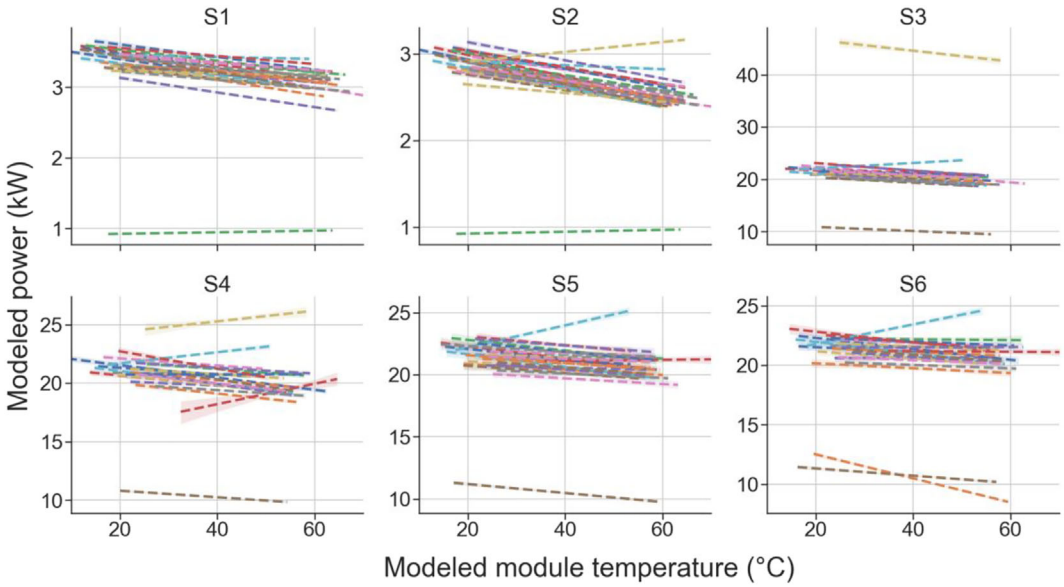

FIGURE 9 Regression model fits for modeled power versus module temperature. The scatter was removed to improve clarity. The regression model fits considered datapoints for modeled Gpoa from 800 to 1200 W/m^2 and wind speed <10 m/s. Participants are shown in different colors.

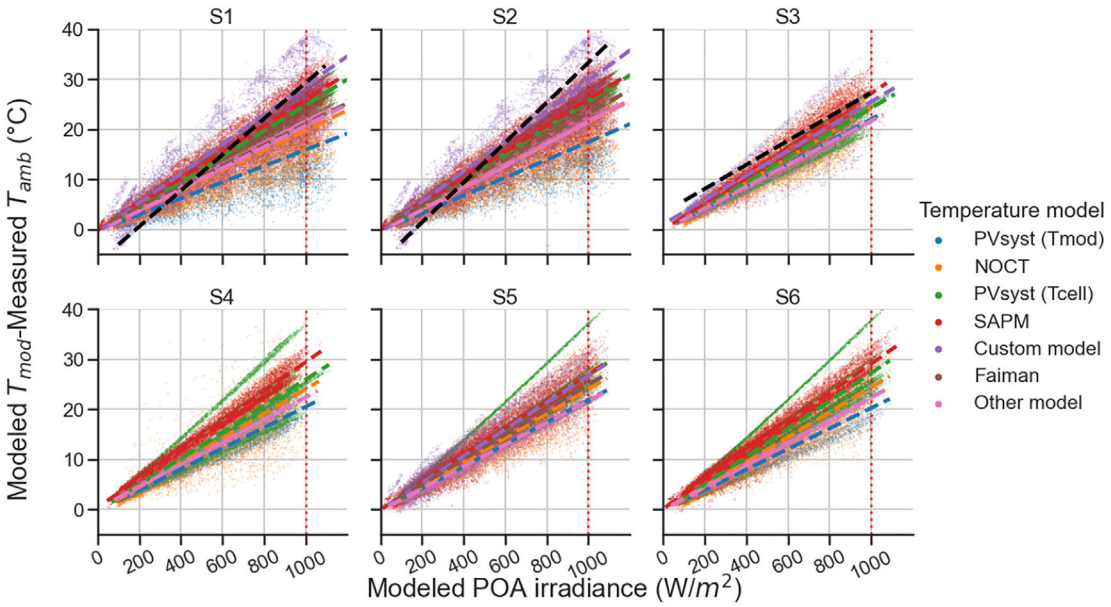

FIGURE 10 Modeled temperature rise (i.e., the difference between modeled module temperature and measured ambient temperature) as a function of modeled Gpoa. Robust regression was used to de-weight outliers (dashed lines). The black dashed lines correspond to module measurements (only available for S1–S3) for wind speed <2 m/s and module temperature >0°C.

as a function of modeled Gpoa is shown in Figure 10. Robust regression[31] (colored dashed lines) was used to de-weight outliers. The black dashed lines correspond to module temperature measurements (only available for S1–S3) for wind speed <2 m/s and module temperature >0°C. Negative temperature differences were measured in Albuquerque where low sky temperatures led to significant radiative cooling of the modules. Only one custom model separately accounts for radiative losses and correctly predicts such negative values. All other models lump radiative losses together with convective losses and represent the combined heat loss with one or two empirical heat loss coefficients. In Driesse et al.,[30] it was shown that all of the named models in Figures 7 and 10 are essentially equivalent; therefore, underprediction of module temperature in the simulations is a clear indication that the heat loss coefficients for the

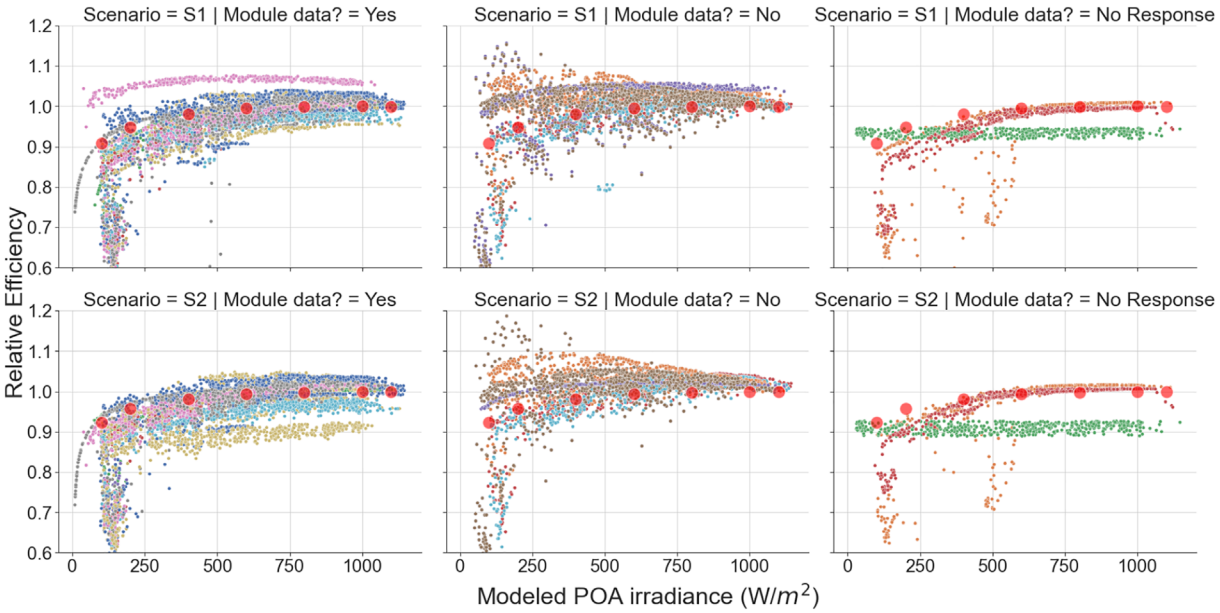

FIGURE 11 Relative efficiency (normalized to median efficiency for modeled Gpoa from 950 to 1050 W/m^2) across modeled plane-of-array irradiance in S1 (top row) and S2 (bottom row). The participants are grouped by whether they have used any of the provided module data (i.e., IEC 61853-1 matrix data or PAN files) or not or have not responded. The red circles represent the measured values, and the participants are shown in different colors. The data represent conditions where AOI < 70° and the modeled Tmod is from 20°C to 30°C.

respective models were frequently set too high. This resulted in maximum temperature difference between participants and field data of up to ~15°C at 1000 W/m^2 in Albuquerque and ~10°C in Roskilde, which would produce an error in simulated module power reaching ~6% at those times. Recent work[32] shows that all these named temperature models can be improved to account for radiative losses, and a modified Faiman model was made available in the open-source *pvlib-python* library as *pvlib.temperature.faiman_rad* (after this study took place). Nevertheless, the improved models still require the appropriate empirical heat loss coefficients for the system being simulated.

The relative efficiency across the modeled Gpoa plots (see Figure 11) can provide information on the electrical modeling assumptions made by the participants based on whether the provided module data were used or not. The relative efficiency was calculated by taking the ratio of modeled efficiency and each participant's "nominal" efficiency. This was calculated by taking a subset of modeled Gpoa from 950 to 1050 W/m^2 and assuming the median temperature corrected (to 25°C) efficiency as the "nominal" efficiency for each participant. The plots in Figure 11 were categorized based on the participants' responses on PAN or IEC 61853-1 matrix data usage; the top and bottom rows correspond to S1 and S2, respectively, while the red circles are the measured values. The data in Figure 11 represent conditions where the AOI < 70° and the modeled Tmod is from 20°C to 30°C. IEC 61853-1 data for all modules show lower efficiencies at low irradiance. Many participants' results matched these data trends, whereas others calculated flat efficiencies with Gpoa or showed higher efficiency at lower irradiance values or did not consider temperature effects.

The plots in Figure 12 show the bias in annual irradiation (A) and energy yield (B) for all participants. Although the irradiation bias for most participants was positive (i.e., showing overestimation) and the overall median was very close to zero (see red dashed lines), the energy yield was underestimated by most of the participants with a median value of −3.3%. This behavior raised the question about the derate assumptions made by the participants. Quantifying or setting the derates is a critical step in PV performance modeling. Derates (or performance loss factors) describe the losses that can occur within a system, for example, due to conductor resistance, soiling, module degradation, and so forth. After comparing derate assumptions by individual participants, it was found that the highest underestimations were exhibited by participants that over-budgeted for derates. In contrast, the participants applying modest derate assumptions achieved biases much closer to zero. This is interesting because the modeling community is often concerned with the accuracy of model equations and their parameter values,[33,34] whereas in this study, the errors were driven largely by the initial assumptions made by the modelers. It should be mentioned, however, that these scenarios include data from lab-scale systems that were built for research purposes. As such, these are likely to experience lower losses than utility scale power plants. To further examine the impact of the derate assumptions, the annual bias by each participant, for each scenario (i.e., Figure 12B), was applied as an adjustment to their corresponding hourly time-series (i.e., by multiplying the hourly modeled power time-series by *one*

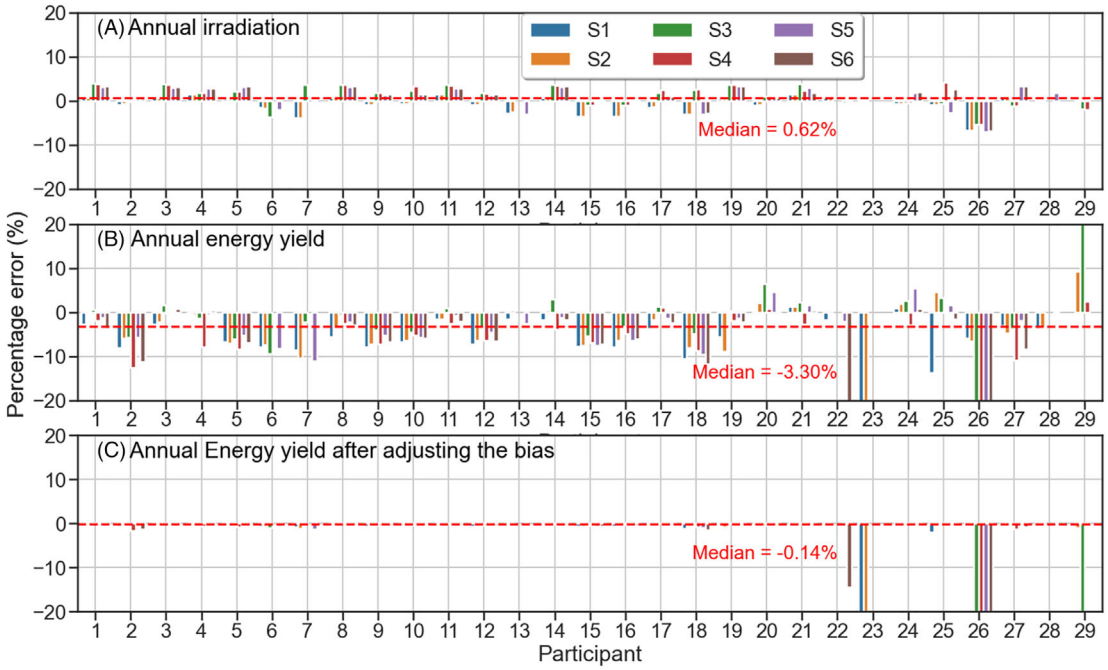

FIGURE 12 Percentage error or bias observed in annual (A) irradiation, (B) energy yield, and (C) energy yield after adjusting the bias. The percentage error is calculated as $100 \times (model - measurement)/measurement$. Dashed horizontal lines indicate median values. Participants that assumed low derates achieved higher accuracies than those that over-budgeted for derates. Overall, all derate assumptions were applied in a linear manner since the bias in (C) was minimized after adjusting the time-series.

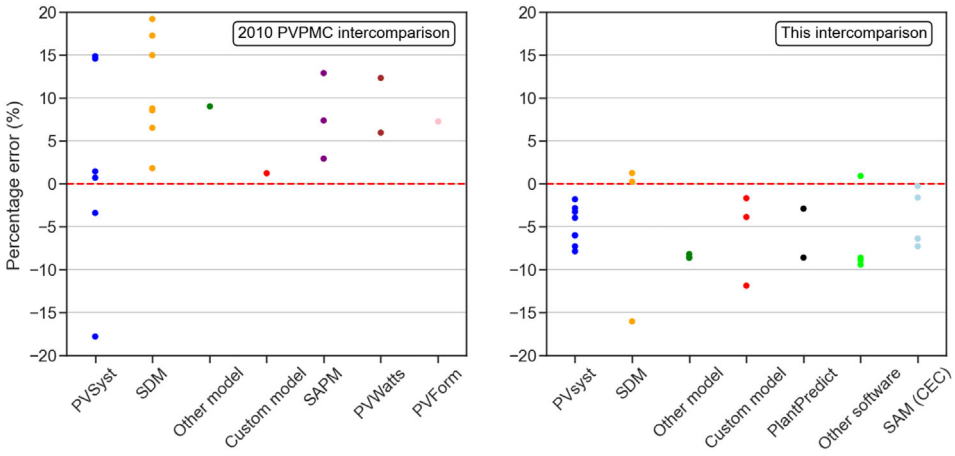

FIGURE 13 Comparing bias in annual energy yield predictions from Sandia's 2010 blind modeling study (left) against a scenario from this study (right). A shift from overestimation to underestimation is observed with a much tighter cluster among models indicating improvements in precision. Accuracy still depends on the modeler's skill.

minus bias). The adjusted annual energy yield bias is then plotted in Figure 12C, where the overall median bias is nearly zero indicating that the majority of bias errors in Figure 12B were indeed linear. This reinforces the potential conclusion that input assumptions matter more than the model, at least for the climates and systems investigated in this study.

How do these results compare to the original PVPMC blind modeling study in 2010[8,9]? Figure 13 (left) shows the bias in annual energy yield from the 2010 blind modeling study, whereas the plot on the right exhibits the bias from this study; these are categorized by the models. Overall, there is a large shift from overestimating energy yield (as high as ~20% in 2010), to being very conservative with the derates resulting in significant underestimation of energy yield (as low as −16%). This strongly indicates that the derate factors are still being guessed by the modelers introducing considerable uncertainty. On the other hand, the individual models, such as the popular PVsyst, seem to be more precise in this study (the maximum difference in energy yield among PVsyst participants was ~33% in 2010 whereas now the difference is ~6%) and tightly clustered. However, the accuracy still depends on the modeler's skill.

To examine whether there are differences among sectors, Figure 14 shows boxplots of the annual energy yield prediction bias grouped by the participants' affiliation category. Software companies exhibited the lowest errors with a median value close to zero. This

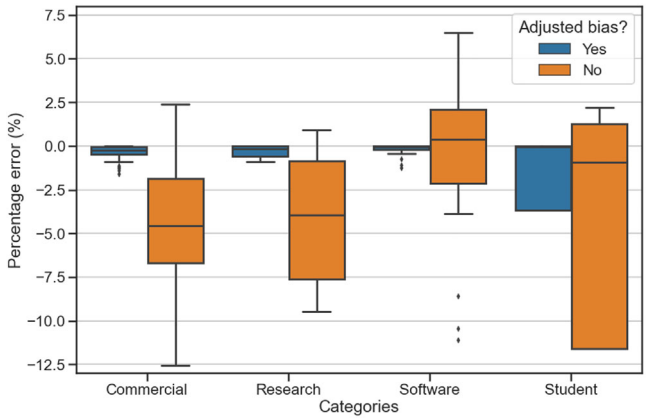

FIGURE 14 Boxplots of annual energy yield prediction bias for all scenarios grouped by the participants' affiliation category. The blue boxplots show the annual energy yield bias after adjusting the modeled power time-series.

could be expected since software companies know their products better than anyone else. It is also worth noting that the commercial sector, which deals mostly with larger power plants, assumed higher derate values resulting in a higher bias spread (see orange boxplot). This is another indication that modeling different system sizes will require appropriate derate budgeting. The adjusted annual energy yield bias shows that all sectors but the student category exhibited distributions very close to zero.

5 | CONCLUSIONS

PVPMC's blind PV modeling intercomparison found that:

1. The irradiance transposition models seem to perform well, except the isotropic one.
2. Modeling the rear Gpoa is still challenging with errors exceeding ∼±100%. However, it should be mentioned that rear Gpoa represents ∼10% or less of the total irradiance.
3. Standardization is needed for handling sun position calculations when using time-averaged irradiance measurements.
4. Incorporating a radiative loss term in module temperature modeling appears to improve accuracy.
5. There is confusion around the U values for Faiman and PVsyst temperature models. Uc and Uv (PVsyst) values should not be used in place of U0 and U1 (Faiman) values.
6. Most software and models showed similar results indicating good reproducibility among participants, especially when compared with the 2010 blind modeling study. For example, the spread in estimated energy yield among PVsyst participants is now ∼6% compared with ∼33% in 2010.
7. Uncertainty and large variation in derate factors between participants appear to explain most of the differences; it was observed that modelers overestimated the derates resulting in significant power underestimation.
8. Human errors are not uncommon. The intercomparison highlighted several errors related to the temperature coefficients and the efficiency across irradiance. There is an opportunity to develop screening tests that can detect such errors, thus assuring stakeholders of the accuracy of the modeling results.
9. Modeler skill at understanding, choosing, and using the models and their parameter correctly, and accumulated experience observing various derate mechanisms in operational systems seems to be more important than the PV model itself (see 7 and 8 above).

Unfortunately, the bifacial PV time-series in this study contained only a handful of rear Gpoa days. As such, no further analysis has been conducted to investigate the impact of their variations. Depending on data availability, future PVPMC blind modeling intercomparisons will include larger systems, subhourly time-series, investigations on rear Gpoa, and an iterative submission process that would enable a more detailed determination of the uncertainties involved at each step of a PV performance modeling pipeline.

AUTHOR CONTRIBUTIONS

Marios Theristis: Conceptualization; methodology; software; validation; formal analysis; investigation; data curation; writing—original draft; review and editing; visualization; project administration. **Nicholas Riedel-Lyngskær:** Data provision; software; validation; writing—review and editing; visualization. **Joshua S. Stein:** Conceptualization; writing—review and editing; supervision; funding. **Lelia Deville**: Data provision; writing—review and editing. **Leonardo Micheli**: Data provision; writing—review and editing. **Anton Driesse**: Data provision; writing—review and editing. **William B. Hobbs**: Data provision; writing—review and editing. **Silvana Ovaitt**: Data provision; writing—review and editing. **Rajiv Daxini**: Data provision; writing—review and editing. **David Barrie**: Data provision. **Mark Campanelli**: Data provision. **Heather Hodges**: Data provision. **Javier R. Ledesma**: Data provision. **Ismael Lokhat**: Data provision. **Brendan McCormick**: Data provision. **Bin Meng**: Data provision. **Bill Miller**: Data provision. **Ricardo Motta**: Data provision. **Emma Noirault**: Data provision. **Megan Parker**: Data provision. **Jesús Polo**: Data provision. **Daniel Powell**: Data provision. **Rodrigo Moretón**: Data provision. **Matthew Prilliman**: Data provision. **Steve Ransome**: Data provision. **Martin Schneider**: Data provision. **Branislav Schnierer**: Data provision. **Bowen Tian**: Data provision. **Frederick Warner**: Data provision. **Robert Williams**: Data provision. **Bruno Wittmer**: Data provision. **Changrui Zhao**: Data provision. All authors reviewed and approved the manuscript.

AFFILIATIONS

[1]Sandia National Laboratories, Albuquerque, New Mexico, USA
[2]Technical University of Denmark, Roskilde, Denmark
[3]Department of Astronautical, Electrical and Energy Engineering (DIAEE), Sapienza University of Rome, Rome, Italy

[4]PV Performance Labs, Freiburg, Germany
[5]Southern Company, Birmingham, Alabama, USA
[6]National Renewable Energy Laboratory, Golden, Colorado, USA
[7]University of Nottingham, Nottingham, UK
[8]Wood PLC, Calgary, Alberta, Canada
[9]Intelligent Measurement Systems, Bozeman, Montana, USA
[10]Southern Power, Birmingham, Alabama, USA
[11]Instituto de Energía Solar, Universidad Politécnica de Madrid, Madrid, Spain
[12]Cythelia Energy, La Motte-Servolex, France
[13]Solphi Engineering, Savannah, Georgia, USA
[14]Eindhoven University of Technology, Eindhoven, Netherlands
[15]Fotovoltec Solar Engineering, Londrina, Brazil
[16]K2 Management, Bristol, UK
[17]CIEMAT, Madrid, Spain
[18]Intersect Power, Oakland, California, USA
[19]Qualifying Photovoltaics, Madrid, Spain
[20]Steve Ransome Consulting Ltd, Kingston upon Thames, UK
[21]Terabase Energy Inc., Berkeley, California, USA
[22]Solargis, Bratislava, Slovakia
[23]Valentin Software GmbH, Berlin, Germany
[24]NEO Virtus Engineering Inc., Littleton, Massachusetts, USA
[25]PVsyst SA, Satigny, Switzerland
[26]CSI Solar Co. Ltd., Suzhou, China

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Anastasios (Tassos) Golnas of the U.S. Department of Energy, Solar Energy Technology Office (DOE-SETO) for the valuable inputs and suggestions on this work. The team at Sandia would like to thank the Technical University of Denmark for providing the data from Roskilde. This work was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 38267. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. L. Micheli's work was supported by Sole4PV, a project funded by the Italian Ministry of University and Research under the 2019 "Rita Levi Montalcini" Program for Young Researchers.

DATA AVAILABILITY STATEMENT

The validation datasets are available online in open access at two locations. The first is on the website of the PV Performance Modeling Collaborative at https://pvpmc.sandia.gov/. The second is on the Duramat Data Hub at https://datahub.duramat.org/dataset/pv-performance-modeling-data (doi: https://doi.org/10.21948/1970772).[18]

ORCID

Marios Theristis https://orcid.org/0000-0002-7265-4922
Nicholas Riedel-Lyngskær https://orcid.org/0000-0002-5654-4571
Lelia Deville https://orcid.org/0000-0002-5021-9743
William B. Hobbs https://orcid.org/0000-0002-3443-0848
Silvana Ovaitt https://orcid.org/0000-0003-0180-728X
Rajiv Daxini https://orcid.org/0000-0003-1993-9408
Mark Campanelli https://orcid.org/0000-0001-5154-2641
Javier R. Ledesma https://orcid.org/0000-0002-2276-0938
Matthew Prilliman https://orcid.org/0000-0001-8852-1318

REFERENCES

1. Andrews R, Yates T. PV performance modeling. SolarPro Issue 8.5; 2015. https://www.solarprofessional.com/account/archive-download
2. Stein JS, Klise GT. Models used to assess the performance of photovoltaic systems. Sandia Report No. SAND2009-8258; 2009.
3. Stein J, Hansen CW, King BH, Sutterlueti J, Ransome S. Outdoor PV performance evaluation of three different models: single-diode SAPM and loss factor model. Sandia Report No. SAND2013-7913C; 2013.
4. Freeman J, Whitmore J, Blair N, Dobos AP. Validation of multiple tools for flat plate photovoltaic modeling against measured data. In: *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*; 2014:1932-1937.
5. Roberts JJ, Mendiburu Zevallos AA, Cassula AM. Assessment of photovoltaic performance models for system simulation. *Renew Sustain Energy Rev*. 2017;72:1104-1123. doi:10.1016/j.rser.2016.10.022
6. Theristis M, Venizelou V, Makrides G, Georghiou GE. Chapter II-1-B—energy yield in photovoltaic systems A2. In: Kalogirou SA, ed. *McEvoy's Handbook of Photovoltaics*. Third ed. Academic Press; 2018:671-713. doi:10.1016/B978-0-12-809921-6.00017-3
7. Driesse A, Theristis M, Stein JS. A new photovoltaic module efficiency model for energy prediction and rating. *IEEE J Photovoltaics*. 2021; 11(2):527-534. doi:10.1109/JPHOTOV.2020.3045677
8. Stein JS. Results of model intercomparison: predicted vs. measured system performance. Sandia Report No. SAND2010-6845C; 2010.
9. Stein JS. The photovoltaic performance modeling collaborative (PVPMC). In: *38th IEEE Photovoltaic Specialists Conference (PVSC)*; 2012:003048-003052.
10. Friesen G, Gottschalg R, Beyer H, et al. Intercomparison of different energy prediction methods within the European project "performance"—results of the 1st round robin. In: *22nd European Photovoltaic Solar Energy Conference*; 2007:2659-2663.
11. Moser D, Herz M, Horvath IT, et al. Benchmarking yield assessment exercise in different climates within an international collaboration framework. In: *37th European Photovoltaic Solar Energy Conference and Exhibition*; 2020.
12. Vogt MR, Riechelmann S, Gracia-Amillo AM, et al. PV module energy rating standard IEC 61853-3 intercomparison and best practice guidelines for implementation and validation. *IEEE J Photovoltaics*. 2022;12(3):844-852. doi:10.1109/JPHOTOV.2021.3135258
13. Standard IEC 61853-3. Photovoltaic (PV) module performance testing and energy rating—part 3: energy rating of PV modules; 2018, Geneva, Switzerland.
14. Riedel-Lyngskær N, Berrian D, Alvarez Mira D, et al. Validation of bifacial photovoltaic simulation software against monitoring data from large-scale single-axis trackers and fixed tilt systems in Denmark. *Appl Sci*. 2020;10(23):8487. doi:10.3390/app10238487
15. Standard IEC 61853-1. Photovoltaic (PV) module performance testing and energy rating—part 1: irradiance and temperature performance measurements and power rating; 2011, Geneva, Switzerland.
16. Standard IEC 61853-2. Photovoltaic (PV) module performance testing and energy rating—part 2: spectral responsivity, incidence angle and

module operating temperature measurements; 2016, Geneva, Switzerland.
17. Long C, Dutton E. BSRN global network recommended QC tests, V2; 2010. https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf
18. Theristis M, Stein JS. PV performance modeling—data and resources, in DuraMat Data Hub. Accessed in April 24, 2023. doi:10.21948/1970772
19. Holmgren WF, Andrews RW, Lorenzo AT, Stein JS. PVLIB python 2015. In: *42nd IEEE Photovoltaic Specialists Conference (PVSC)*; 2015: 1-5.
20. Cythelia. archelios PRO software. La Motte-Servolex, France. https://www.trace-software.com/archelios-pro/solar-pv-design-software/
21. Valentin Software GmbH. PV*SOL software. Berlin, Germany. https://pvsol.software/en/
22. Carrillo JM, Muñoz J, Makibar A, Luna A, Narvarte L. SISIFO: the open-source simulation tool of PV systems developed in PVCROPS. In: *31st European Photovoltaic Solar Energy Conference*; 2015:14-18.
23. Perez R, Stewart R, Seals R, Guertin T. The development and verification of the Perez diffuse radiation model. Sandia Report No. SAND88-7030; 1988.
24. PVsyst SA. Project design > array and system losses > array thermal losses. PVsyst 7 Help. https://www.pvsyst.com/help/thermal_loss.htm
25. King DL, Boyson WE, Kratochvil JA. Photovoltaic array performance model. Sandia Report No SAND 2004-3535; 2004.
26. Gilman P. SAM photovoltaic model technical reference. NREL Report No. NREL/TP-6A20-64102; 2015.
27. Passow K, Ngan L, Rich G, Lee M, Kaplan S. PlantPredict: solar performance modeling made simple. In: *44th IEEE Photovoltaic Specialist Conference (PVSC)*; 2017:600-603.
28. PVsyst SA. PVsyst software. Satigny, Switzerland. https://www.pvsyst.com/
29. Perez R, Ineichen P, Seals R, Michalsky J, Stewart R. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*. 1990;44(5):271-289. doi:10.1016/0038-092X(90)90055-H
30. Driesse A, Theristis M, Stein JS. PV module operating temperature model equivalence and parameter translation. In: *IEEE 49th Photovoltaics Specialists Conference (PVSC)*; 2022:0172-0177.
31. Huber PJ. Robust regression: asymptotics, conjectures and Monte Carlo. *Ann Stat*. 1973;1(5):799-821. doi:10.1214/aos/1176342503
32. Driesse A, Stein J, Theristis M. Improving common PV module temperature models by incorporating radiative losses to the sky. Sandia Report No. SAND2022-11604; 2022.
33. Hansen CW, Pohl A. Which models matter: uncertainty and sensitivity analysis for photovoltaic power systems. In: *40th IEEE Photovoltaic Specialist Conference (PVSC)*; 2014:0175-0180.
34. Hansen CW, Martin CE. Photovoltaic system modeling: uncertainty and sensitivity analyses. Sandia Report No. SAND2015-6700; 2015.

How to cite this article: Theristis M, Riedel-Lyngskær N, Stein JS, et al. Blind photovoltaic modeling intercomparison: A multidimensional data analysis and lessons learned. *Prog Photovolt Res Appl*. 2023;31(11):1144-1157. doi:10.1002/pip.3729