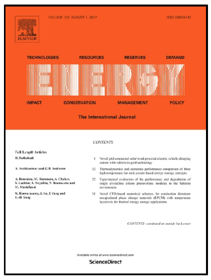
# Accepted Manuscript



Evaluation of Data-Driven Models for Predicting Solar Photovoltaics Power Output

Salim Moslehi, T. Agami Reddy, Srinivas Katipamula

PII: S0360-5442(17)31564-5

DOI: 10.1016/j.energy.2017.09.042

Reference: EGY 11536

To appear in: *Energy*

Received Date: 31 May 2017

Revised Date: 17 August 2017

Accepted Date: 10 September 2017

Please cite this article as: Salim Moslehi, T. Agami Reddy, Srinivas Katipamula, Evaluation of Data-Driven Models for Predicting Solar Photovoltaics Power Output, *Energy* (2017), doi: 10.1016/j.energy.2017.09.042

*Highlights:*

- Various data collection and modeling scenarios to predict PV power were evaluated.
- Inverse models of PV power output solely based on climatic data are very accurate.
- Inclusion of incidence angle modifier improves PV power model prediction accuracy.
- Wind velocity found to be statistically insignificant in PV power forecast models.
- PV power models are accurate even if only solar horizontal radiation is measured.

# Evaluation of Data-Driven Models for Predicting Solar Photovoltaics Power Output

Salim Moslehi [a,*]

T. Agami Reddy [a]

Srinivas Katipamula [b]

[a] *Arizona State University, School of Sustainable Engineering and the Built Environment, Tempe, AZ, 85281, USA*

[b] *Pacific Northwest National Laboratory, Richland, WA, 99354, USA*

## Abstract

This research was undertaken to evaluate different inverse models for predicting power output of solar photovoltaic (PV) systems under different practical scenarios. In particular, we have investigated whether PV power output prediction accuracy can be improved if module/cell temperature was measured in addition to climatic variables, and also the extent to which prediction accuracy degrades if solar irradiation is not measured on the plane of array but only on a horizontal surface. We have also investigated the significance of different independent or regressor variables, such as wind velocity and incident angle modifier in predicting PV power output and cell temperature. The inverse regression model forms have been evaluated both in terms of their goodness-of-fit, and their accuracy and robustness in terms of their predictive performance. Given the accuracy of the measurements, expected CV-RMSE of hourly power output prediction over the year varies between 3.2% and 8.6% when only climatic data are used. Depending on what type of measured climatic and PV performance data is available, different scenarios have been identified and the corresponding appropriate modeling pathways have been proposed. The corresponding models are to be implemented on a controller platform for optimum operational planning of microgrids and integrated energy systems.

*Keywords:* solar photovoltaics, PV power prediction, data-driven modeling, renewable energy, sustainable energy

* Corresponding author:
*Tel:* +1(480)799-8265, *E-mail address:* smoslehi@asu.edu
*Street address:* 660 S College Ave., CAVC building, Arizona State University (Tempe campus), Tempe, AZ, 85281

**Nomenclature**

| | |
|---|---|
| $A$ | PV panels surface area, $m^2$ |
| CV-RMSE | Coefficient of Variation of Root Mean Square Error |
| $I$ | Total horizontal irradiance, $W/m^2$ |
| $I_{diff}$ | Diffuse irradiance on horizontal surface, $W/m^2$ |
| $I_T$ | Total irradiance on tilted plane or POA irradiance, $W/m^2$ |
| $K_\eta$ | Incidence angle modifier |
| MAE | Mean Absolute Error, W |
| MBE | Mean Biased Error, W |
| NMBE | Normalized Mean Bias Error |
| NREL | National Renewable Energy Laboratory |
| NOCT | Nominal Operating Cell Temperature, ˚C |
| POA | Plane Of Array |
| PV | Photovoltaic |
| $P_{elec}$ | Solar PV power output (either DC or AC), W |
| RMSE | Root Mean Square Error, W or ˚C |
| $R^2$ | Coefficient of determination |
| $T_a$ | Ambient dry-bulb temperature, ˚C |
| $T_{cell}$ | PV module/cell temperature, ˚C |
| TMY | Typical Meteorological Year (TMY databases are collections of the selected meteorological data provided for many locations across the world by NREL) |
| U | Overall thermal heat loss coefficient, $W/m^2$˚C |
| $v_{wind}$ | Wind velocity, m/s |
| $\beta$ | Temperature degradation coefficient |
| $\theta_i$ | Solar incidence angle on POA, degrees |
| $\eta_{elec}$ | Electrical efficiency of PV module |
| $\eta_n$ | Optical efficiency at normal solar incidence |
| Subscripts | |
| $DC$ | direct current |
| $m$ | measured |
| $p$ | predicted |
| $rated$ | at rated condition |

## 1. Introduction

Sustainable development and planning of microgrids and integrated energy systems requires optimum management of various resources [1,2]; in this regard, forecasting power output of non-dispatchable resources, such as solar photovoltaic (PV) systems, wind, etc. is crucial. On the other hand, given the stochasticity in solar PV power generation due to randomness and variations in solar irradiation and other climatic factors, providing reliable electricity, especially in case of a high solar grid penetration, is a challenge which requires precise PV performance modeling and forecasting. Accurate forecasting of PV power output allows building managers to plan the operation and optimum control of their various energy systems most effectively, and also allows grid operators to manage solar power efficiently by economic dispatch and to ensure grid stability [3].

### *1.1. Background*

PV power prediction modeling methods can be classified into two general approaches: (i) deterministic methods which use physics-based models and require detailed design and rated performance information regarding the PV system; and (ii) data-driven approaches which require PV power output measurements [4]. The latter category includes statistical models, linear or time series models, and artificial intelligence techniques [5].

In the last few years, researchers have focused on developing models to forecast power generation of solar PV systems at different scales and over different forecasting horizons. Predictive performance of the proposed models varies depending upon the methodology and accuracy of the input data. Zhang et al. proposed a methodology for identifying the baseline and target values regarding accuracy of the solar PV power forecasting models [6]. Other relevant studies include: Picault et al. who evaluated forecasting models for crystalline silicon PV systems subject to mismatch losses [7]; Chen et al. who developed an on-line 24-hr ahead PV power forecasting model using artificial neural network [8]; Jimenez et al. who developed a power prediction model for grid connected PV plants comprised of weather and power forecasting modules [9]; Su et al. who used Gaussian equations to model grid-connected PV system outputs and performance efficiency [10]; Almeida et al. who used quantile regression forests to develop a nonparametric model to forecast AC power output of PV plants [11]; Chu et al. who created a real-time reforecasting method based on artificial neural network optimization which applied to the intra-hour PV power prediction models [12]; Nobre et al. who proposed a hybrid method for converting solar irradiation into PV power in a tropical and densely-built environment taking into consideration the effects of PV module temperature, shading effects, system degradation over time, and variation in incident irradiation caused by air pollution [13]; Rana at al. who developed a very short-term (5 to 60 minutes ahead) PV power forecasting model using neural network techniques [14]; and Baharin et al. who created artificial neural network models to predict PV system power generation in three typical weather conditions in tropical climates, i.e. clear, cloudy, and overcast sky [15]. Short-horizon solar power forecasting can also help building and industrial unit planners better integrate solar heat to energize their systems using a cleaner and cheaper energy [16,17].

### *1.2. Research Objectives*

The objective of this research is to identify the most effective data collection and model building scenario to support reasonably accurate prediction of solar PV power output given the uncertainties associated with field measurements. This broad goal is broken into more specific research objectives which are prioritized based on their importance. Different forms of inverse models were evaluated, using year-long measured data, in order to determine the most accurate ones. This analysis is also meant to provide range of errors that would be expected in power prediction depending on what type of data is being collected.

Primary research questions are as follows:

- To what extent does accuracy of the PV power prediction model improve if separate individual month models are constructed compared to fitting the data over the whole year data?
- Which of the model forms proposed in the literature to predict PV module power using climatic data performs the best? And how significant is it to include wind velocity as a regressor in the model?
- Which of the cell temperature prediction model forms has the least error? How accurate are the regression models compared to physics based models? Whether inclusion of the incidence angle modifier term will improve prediction accuracy?
- How much error is likely to be introduced to the power prediction as a function of cell temperature and POA irradiation if predicted cell temperatures are used instead of measured values?
- In identifying PV power prediction models, is there any benefit in predicting cell temperature and then deducing PV power output, instead of directly deducing it from climatic data?
- How much error is likely to be introduced in power prediction when only total horizontal radiation is measured and both diffuse and POA irradiation have to be predicted?

Proper answers to these questions will help energy system planners and operators to better define their data collection strategies so as to obtain most accurate power predictions.

## 2. Methodology and Modeling

In general, there are two scenarios associated with collection of required data and modeling of solar PV performance (Figure 1):

(i) Predicting PV module power output ($P_{elec}$) using measured cell/module temperature ($T_{cell}$) and climatic data: irradiation on POA ($I_T$), ambient temperature ($T_a$), and wind velocity ($v_{wind}$).
(ii) Predicting PV module power ($P_{elec}$), only using climatic data, i.e. solar irradiation on a horizontal surface ($I$) or on tilted surface ($I_T$), $T_a$, and $v_{wind}$.

Figure 1.

## *2.1. PV Power Prediction Models*

### *2.1.1. Inverse models for module power prediction using climatic data*

A black-box model needs monitored data to determine model coefficients appropriate for the specific PV system. In this case, module or array power output is directly correlated with the environmental parameters (Myers and Emery, [18]).

$$P_{elec} = A.I_T(a' + b'.T_a + c'.I_T + d'.v_{wind}) \tag{1}$$

where A is the collector area. This equation can be re-expressed in a form suitable for regression as:

$$P_{elec} = a.I_T + b.I_T.T_a + c.I_T{}^2 + d.I_T.v_{wind} \tag{2}$$

where a, b, c, and d are regression coefficients which include the effect of PV system surface area *A*. Though this model form includes the effects of wind, its statistical significance needs to be evaluated under the various operating conditions experienced by the PV systems over the year.

A more physical model (grey-box) of the power generated in terms of the climatic variables and the PV system characteristics is the one suggested by Gordon and Reddy, [19]:

$$P_{elec,DC} = A.I_T.\mathrm{K}_\eta.\eta_{rated}(a_1 - a_2.T_a - a_3 I_T) \tag{3}$$

where $K_\eta$ is the incidence angle modifier, discussed later in the paper. The parameters have some physical interpretation and characterize PV cell properties i.e. overall thermal heat loss coefficient, temperature degradation coefficient, and optical efficiency at normal solar incidence, as:
$a_1 = 1 + \beta.T_{cell,rated}$, $a_2 = \beta$ and $a_3 = \beta(\eta_n - \eta_{n,rated})/U(v_{wind})$

and $U(v_{wind})$ is equal to $a+b.v_{wind}$ where the coefficients should be determined from in-situ measurements and are dependent on the deployment type of the solar panel. Neglecting the effects of wind, this equation can be re-expressed as an inverse model:

$$P_{elec} = a.I_T.\mathrm{K}_\eta + b.I_T.\mathrm{K}_\eta.T_a + c.(I_T.\mathrm{K}_\eta)^2 \tag{4}$$

This equation can be considered to be akin to the Myers and Emery model (eq. 2) except that it accounts for the solar incidence angle by including the incidence angle modifier, $K_\eta$, calculated as:

$$K_\eta = 1 + b(1/\cos\theta_i - 1) \tag{5}$$

where the coefficient *b* can be assumed to be equal to -0.10 for single glazed collectors.

*2.1.2. Inverse models for module power prediction using cell temperature*

The power output can be predicted based on detailed physical models. However, a more convenient model for field applications is based on module cell temperature and the incident irradiation. The most widely used model is the one originally proposed by Evans [20]:

$$\eta_{elec,DC} = \eta_{elec,rated,DC}[1+\beta(T_{cell} - T_{cell,rated})] \tag{6}$$

Copper et al. [21] modified the Evans model (eq. 6) and presented it in terms of power output. This model, referred here as the modified Evans model, can be re-expressed as an inverse model:

$$P_{elec,DC} = a.I_T + \mathrm{b}(T_{cell} - T_{cell,rated}).I_T \tag{7}$$

## 2.2. Cell Temperature Models

A variety of approaches have been proposed for predicting solar PV module/cell temperature. The cell or module temperature is often not measured in field installations. Even if it is, there is some indeterminacy depending on where such point measurements are taken on the panel surface. Faiman experimentally found about a $2^{o}C$ root mean square error (RMSE) in cell temperature difference at different points of a PV module [22].

*2.2.1. Grey-box and physics-based cell temperature models*

The grey box modeling approach retains some of the physics-based interactions while allowing regression-based adjustments to the module form and associated model coefficients when measured data is available. This approach is perhaps the most widely used and most suitable for modeling actual PV systems and installations. Several papers (Gordon and Reddy [19], Evans [20], Copper et al. [21], Faiman [22], Koehl et al. [23], Mekhilef et al. [24], Schwingshackl et al. [25], and Kaldellis et al. [26]) have described this general approach and evaluated them against in-situ installations.

The physics-based models make use of many of the parameters provided in the specification sheets. Relevant papers are those by King et al. [27], De Soto et al. [28], Lo Brano et al., [29,30], Dolara et al. [31], and Razon [32]. Detailed PV system simulation software programs, such as PVsyst [33] or SAM [34], have modules which are based on this approach. This level of detail is not well suited for field application since model calibration at various stages is required which is somewhat tedious and requires user experience.

*2.2.2. Empirical cell temperature models*

A more empirical model to determine cell temperature, $T_{cell}$, is to use a thermal energy balance on the module or array:

$$I_T.\eta_n K_\eta = I_T.\eta_{elec} + \mathrm{U}.(\mathrm{T}_{cell} - T_a) \tag{8}$$

where U is the overall thermal heat loss coefficient of the module in $W/m^2.^{o}C$. It has been pointed out by some experimental studies that wind velocity representative of the value impacting module heat loss is not a well-defined variable and using actual values may not yield more accurate models (PVSyst, [33]). A few studies present detailed heat transfer equations to model the heat loss coefficient involving forced and natural convection effects as well as radiation heat loss (for example, Kaplani and Kaplanis, [35]), but these are not suitable for field models.

Several simplified approaches to predict cell temperature have been proposed based on the observation that the difference between cell and ambient air temperatures vary linearly with solar irradiation. One such formulation is that by Kurtz et al. [36] which was adopted from the King et al. model [27]:

$$T_{cell} = T_a + I_T.\exp(a + b.v_{wind}) \quad (9)$$

where a = -3.473 and b= -0.0594

A simpler black-box model was also found to capture the measured electric output of 22 PV systems of six different cell types monitored in the Phoenix area (Mani et al., [37]):

$$T_{cell} = a + b.I_T + c.T_a + d.v_{wind} \quad (10)$$

Perhaps the most popular forward-model to predict $T_{cell}$ is based on the NOCT value [38]. The NOCT is a measured performance value of the actual solar PV module under stipulated climatic conditions (solar irradiation of 800 W/m$^2$ and ambient temperature of 20 ˚C) and whose numerical value is provided by the manufacturer as part of the specification sheet. The actual cell temperature model is assumed to be independent of wind velocity and framed as:

$$T_{cell} = T_a + \left(\frac{NOCT - 20^o\ \mathrm{C}}{800\ \mathrm{W/m^2}}\right) \times I_T \quad (11)$$

The regression model will then assume the following form:

$$T_{cell} - T_a = a.I_T \quad (12)$$

Note that this is a simplified version of eq. 10 with coefficients *c*=1, *a*=0, and *d*=0.

## 3. Data Analysis and Results

### *3.1. Datasets and data channels*

Yearlong monitored data sets for five U.S. locations, collected and maintained by National Renewable Energy Laboratory (NREL), and University of Oregon were used for this analysis to develop/analyze regression models for solar PV cell temperature and PV power output prediction. The NREL database includes measured solar PV performance data for flat-plate PV modules along with meteorological data on solar irradiance in three locations: Cocoa, FL (from 2011/01 to 2012/03), Eugene, OR (from 2012/12 to 2014/01), and Golden, CO (from 2012/08 to 2013/09). These publicly available datasets are intended to facilitate the validation of existing models, as well as developing new ones for predicting the performance of PV modules. In addition to NREL database, solar PV performance data for two different locations, i.e. Portland, OR and Bend, OR (from 2004 to 2016) were obtained from Solar Radiation Monitoring Laboratory (SRML) of the University of Oregon. Note that these PV systems are rather small in capacity, and though not reflective of actual large PV systems, nevertheless, serve to evaluate/validate the modeling equations.

### *3.2. Analysis results*

After cleaning the datasets of erroneous data points and missing values, each of the datasets was divided into two sets, namely model training, including 70% of the entire dataset, and the remainder 30% served as the testing datasets. The cleaning process includes removing the data points for which the solar irradiation or power outputs from the panels were out of acceptable range, i.e. above the capacity of the system or negative values. Splitting the dataset to training and testing was done using a random number generator in order not to introduce seasonal biases. This was followed by a check using statistical measures of dispersion (mean and standard deviation) on the two data sets to reconfirm randomness. This enabled a proper evaluation of the predictive accuracy of identified regression models from the training datasets.

*3.2.1. Individual months vs whole year models for PV power predictions*

We have investigated the predictive accuracy of regression models identified from whole-year data when applied to different individual months and compared the results with that of models developed using individual month data. Portland dataset (SRML) was chosen to perform this analysis and the months of March, July and December were selected to represent moderate, warm and cold seasons. Table 1 summarizes performance of the power models built over whole year data versus those fitted over individual months.

Table 1.

Since the average solar radiation, and accordingly the power outputs are drastically different during different months, the best criterion to evaluate prediction performance of the models would be the CV-RMSE. It was found that the month of July has the lowest CV-RMSE and the highest values are associated with December. This could be due to the uncertainty and/or fluctuations introduced in the data as a result of partially cloudy sky in December. We can conclude that more uncertainty would be expected during seasons with variable weather conditions while model predictions tend to be more accurate during clear-sky seasons.

We noted that models identified from different months can be both better or poorer than that identified from whole-year data. Therefore, we can conclude that whole-year models are adequate for practical implementations. Whole-year models were identified and analyzed in the rest of this study.

*3.2.2. PV power models with climatic data measurements*

In order to evaluate existing models for PV power prediction using climatic data while investigating whether wind velocity is a significant regressor, we have developed regression models based on year-long data from all five locations using two different regression model forms, i.e. eqs. 2 and 4. The wind velocity variable is not measured in the NREL datasets, therefore eq. 2 was used without including this variable. Goodness-of-fit of the identified models are reported in Table 2 for both model fitting and testing. Neglecting the effects of wind would result in both models having the same form. However, the Gordon and Reddy model includes the incidence angle modifier ($K_\eta$) and this seems to improve prediction accuracy (range from 0.5 to 1.6% points in CV-RMSE) in all five locations as compared to the Myers and Emery model (eq. 2). Since the size of the investigated PV systems are different, CV-RMSE (normalized RMSE against the mean of the measurements) would be a better index for comparing the results across various systems and locations.

Table 2 also summarizes the performance of the power output model predictions over the testing datasets. Mean Biased Errors (MBE) are also reported along with other criteria (MAE, MBE, and RMSE are in Watts). It can be seen that models prediction errors are in the acceptable range for all five locations. It was also found that MBE values are uniformly negligible for all of the studied locations (all being less than 0.6% of the mean) which validates the predictive ability of the identified regression models.

Table 2.

Whether the inclusion of the wind velocity variable would improve model accuracy was also investigated. Since wind velocity data is not available for NREL datasets, this analysis could only be done for Bend and Portland (using SRML data). Results of the analysis (assembled in Table 3) suggests that inclusion of wind effects *does not* increase model prediction accuracy which is consistent with findings from previous studies in the literature (such as [5,6,13]). The issue is that it is difficult to measure the wind velocity representative of the actual wind velocity over all the solar PV panels. Measurements taken at a local weather station or even at a central location on site may not realistically reflect actual wind velocity over the PV arrays. Also, there is a time lag between wind velocity variations and changes in the cell temperature which affects the power outputs. These reasons can explain why the wind velocity is not a significant factor in the PV power prediction models.

Table 3.

*3.2.3. PV cell temperature models*

As mentioned earlier, an alternative approach to predict solar PV power output is based on using PV cell (or module) temperature along with climatic data. If cell temperature is not measured, it ought to be estimated from relevant models. Kurtz and NOCT models (eq. 9 and eq. 12 respectively) would essentially have the same form when applied to NREL datasets in the absence of wind data. In addition to black-box models, physics-based models can be adopted wherever the required data is available. In this regard, NOCT model performance was also evaluated for the SRML data since NOCT values of the panels were available. Model fitting and model validation results are summarized in Table 4.

The Mani black-box regression model (eq. 10) was found to be the most accurate among the various models evaluated. Parameters included in all models are the same and differences in predictions accuracies are due to differences in model forms; for example, in the absence of wind measurements (NREL datasets), Kurtz model is a zero-intercept model which relates the cell and ambient temperature differences to POA irradiations, while the Mani model includes the intercept term and also considers ambient and cell temperature as separate parameters which leads to a more flexible model; such model can capture variations in the data more effectively. The goodness-of-fit results for model training and model testing are in good agreement. This signifies that the identified models are robust, otherwise testing data fits would have been poorer. It was also found that the Kurtz model is more likely to result in biased predictions (about 7.5% of the mean) which can be due to the inclusion of the wind velocity variable. MBEs were found to be less than 3.2% of the mean wherever wind effects are not included.

Table 4.

Effects of including the incidence angle modifier on cell temperature predictions were investigated for the Mani model (which was earlier found to be the most accurate model). The results are assembled in Table 5. Generally speaking, including the incidence angle modifier correction term in the model for predicting $T_{cell}$ seem to result in small improvements (ranging from 0.2 to 0.5% points in CV-RMSE). It can be concluded that, compared to power models, cell temperature models are less sensitive to incidence angle modifier. The time lag between changes in the incident irradiation and the cell temperature response can justify this finding and also explain why cell temperature models tend to be less accurate than power prediction models.

Table 5.

*3.2.4. Effect of using measured vs modeled cell temperature on PV power models*

If solar cell temperature is not measured, one could predict cell temperature, denoted by $T_{cell,p}$, which could then be used to predict PV module power output. We are interested in determining the extent to which the accuracy of these predictions would degrade if predicted cell temperatures are used instead of measured ones. In order to perform this analysis, results of the most accurate regression model for predicting cell temperature, i.e. the Mani model form, was used to compare predicted power values against actual at-site measured ones. Likewise, measured cell temperatures, denoted by $T_{cell,m}$, were used to predict power outputs. The Evans model form was used in this analysis, and the effect of incidence angle modifier was also investigated. Results of model fitting and model validation are assembled in Table 6. We note that there is little or no improvement in predicting PV power from either $T_{cell,p}$ or $T_{cell,m}$.

Figure 2 is the scatter plot of predicted versus measured cell temperatures for two selected datasets, namely Cocoa and Bend. The differences are larger in the middle of the cell temperature range than the ends for Bend, while the discrepancies are larger at the higher end for Cocoa. The reason could be that the relatively few cell temperature sensors usually deployed may not be adequately capturing the average module temperatures under different operating conditions over the year.

Figure 2.

Table 6.

Results of this analysis suggest that predicted cell temperatures can be used in power prediction models without compromising the accuracy of the power prediction. In other words, the uncertainty associated with cell temperature prediction from the regression model seems to be of the same range as that from cell temperature measurements. Model validation errors are also consistent with fitted models errors.

*3.2.5. Comparing PV power models identified from climatic data with those using cell temperature*

Another pathway to PV power prediction would involve predicting cell temperatures from climatic data and then using these predictions to calculate the power output of the solar panels (rather than directly predicting PV power output from weather data). Hence, the power model prediction results have been compared for two possible scenarios:

i) using climatic data directly to predict power outputs (using the Gordon and Reddy model, eq. 4 with incidence angle modifier),
ii) using climatic data, first estimate the cell temperature and use that value to predict PV power outputs (using Evans model form, eq. 7).

Results of our analysis comparing the Gordon and Reddy model (eq. 4) and the two different model forms of the modified Evans model are shown graphically in Figure 3. We note that calculating the power outputs directly from climatic data would result in more accurate predictions(1-2% CV points) than using cell temperatures, either measured or predicted. Therefore, it can be concluded that (i) the Gordon and Reddy model approach is more accurate than the Evans model, and that (ii) using measured cell temperatures would not provide more accurate power predictions compared to those obtained directly from climatic data (unless in the future, a measurement strategy of determining more representative cell temperature can be developed).

Figure 3.

*3.2.6. Uncertainty in PV power models using global horizontal radiation only*

An important practical consideration in our research is that solar irradiation on the POA and/or the cell temperature may not be measured for the PV system. Copper et al. [21] investigated this issue with three PV systems in Australia. They found that when onsite measurements were available, NMBE values of PV power predictions were about 3.2% and CV-RMSE < 6%, which increases substantially (NMBE <13%) when such onsite measurements of tilted radiation and cell temperatures were not available.

In this section, we have analyzed the PV power output predictions when only global horizontal radiation and ambient temperature measurements are available and the POA irradiation values have to be estimated; this requires assessment of diffuse and beam portion of solar irradiation. The NREL datasets from Golden, CO, Eugene, OR, and Cocoa, FL were used to perform this evaluation. Following steps (also depicted in Figure 4) were taken to build the regression models required for predicting the solar PV power output:

i) The year-long dataset was divided into training and testing datasets based on variations in total horizontal irradiance throughout the year.
ii) TMY3 data on diffuse radiation, global horizontal radiation, and extraterrestrial radiation were used to built a regression model for calculating diffuse radiation values for both training and testing datasets
iii) The obtained values for diffuse radiation were used to evaluate the POA irradiance for both training and testing datasets using the Hay, Davies, Klucher and Reindl (HDKR) transposition model [39].
iv) Having POA irradiance, ambient temperature, incident angle modifier, and power for the training dataset, the Gordon and Reddy model (eq. 4) can be trained.
v) PV power outputs were predicted for the testing dataset using the obtained model.

Figure 4.

Due to unavailability of TMY3 files for Cocoa and Golden, TMY3 data from nearby locations were used. Results of this analysis are summarized in Table 7 for PV power output prediction for the testing dataset. We observed that the CV-RMSE values were almost doubled compared to the case when POA measurements were available at-site. We found that although the diffuse radiation predictions using the TMY3 data are relatively uncertain (CV-RMSE values of around 0.4), such error are unlikely to propagate to the POA predictions. Note that the CV-RMSE values were found to be slightly lower than those associated with PV power output predictions.

Table 7.

## 4. Conclusions

The most accurate predictions of field PV module power require measurements of solar irradiation on the plane of array (POA) as well as ambient air temperature. However, in practice, POA irradiation may not be available, and this quantity would need to be estimated from horizontal solar irradiation. We have investigated various pathways for predicting solar PV power output in order to identify the most accurate model forms, and also to provide insight on range of errors that should be expected. These broad goals were broken down into more specific research objectives for which the important conclusions and summary of results relevant to solar PV power prediction are stated below.

(a) PV power prediction models identified from different months can be both better or poorer compared with one model identified from whole-year data. Therefore, one simple annual model, as against 12 individual monthly models, are adequate due to convenience and simplification it offers for practical implementations.
(b) Among the various models investigated to predict PV module power output using climatic data, the Gordon and Reddy model form (eq. 4) was found to have the best predictive accuracy.
(c) During model prediction on testing datasets, residual error statistics confirm that both power prediction models, i.e. Gordon and Reddy model, and Myers and Emery model are unbiased (normalized MBE is less than 0.3% for all investigated location).
(d) Including the effects of wind did not increase the accuracy of power prediction models. This was attributed to the fact that wind velocity readings taken from a distant station do not accurately represent wind conditions across the installed PV module array.
(e) Calculating the power output directly from climatic data would result in more accurate predictions than using a two-step process where one would first predict cell temperatures from climatic variables and then predict power output. Expected CV-RMSE values of power output predictions vary between 3.2% and 8.6% when climatic data is used (using the Gordon and Reddy model) and ranges from 4.1% to 9.7% when cell temperature and POA irradiance was used (using the Evans model). Therefore, the Gordon and Reddy model form would be considered as the most accurate power prediction model among all investigated models.
(f) In the absence of diffuse and POA irradiation, the expected CV-RMSE of solar PV power prediction was found to be around 7.5%. Recall that the CV-RMSE of power prediction using the POA irradiation measured data for the same location was 5.3%; this implies that $I_{diff}$ and POA irradiation prediction errors have been propagated to the power prediction results.
(g) Inclusion of the incidence angle modifier term was found to improve power and cell temperature prediction accuracy by 0.5 – 1.5% and 0.2 – 0.4% respectively in terms of CV-RMSE values.
(h) Measuring cell temperatures would not make the power prediction models more accurate compared to those based on climatic data, unless more representative cell temperature measurements are available. The issue is that the spatial (and even temporal) changes in cell temperature distribution across the module may be impossible to capture with one (or even a few) temperature sensor placed on the module.

## Acknowledgements

The authors acknowledge the Buildings Technologies Office of the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy [under Contract DE-AC05-76RL01830 thru Pacific Northwest National Laboratory] for supporting this research and development effort. The authors thank Mr. Joseph Hagerman, Technology Development Managers for his guidance and strong support of this work. We acknowledge George Hernandez from PNNL for his technical guidance. We also thank Dr. Frank Vignola from the Solar Radiation Monitoring Laboratory at University of Oregon and Dr. Andy Walkers from NREL for supplying us with much of the monitored solar and PV module power data used in our analysis.

## References

[1] Moslehi S, Reddy TA. Sustainability Index of Community Eenergy Systems for Benchmarking and Multi-criteria Decision Analysis. Am. Soc. Mech. Eng., vol. 6B, American Society of Mechanical Engineers; 2016.

[2] Moslehi S, Arababadi R. Sustainability Assessment of Complex Energy Systems Using Life Cycle Approach-Case Study: Arizona State University Tempe Campus. Procedia Eng., vol. 145, Elsevier Ltd; 2016, p. 1096–103.

[3] Arababadi, R., Parrish K. Reducing the Need for Electrical Storage by Coupling Solar PVs and Precooling in Three Residential Building Types in the Phoenix Climate. ASHRAE Trans Press 2017.

[4] Pierro M, Bucci F, Felice M De, Maggioni E, Perotto A, Spada F, et al. Deterministic and Stochastic Approaches for Day-Ahead Solar Power Forecasting. J Sol Energy Eng 2017;139:1–12. doi:10.1115/1.4034823.

[5] Raza MQ, Nadarajah M, Ekanayake C. On recent advances in PV output power forecast. Sol Energy 2016;136:125–44. doi:10.1016/j.solener.2016.06.073.

[6] Zhang J, Hodge B, Lu S, Hamann HF, Lehman B, Simmons J, et al. Baseline and target values for regional and point PV power forecasts : Toward improved solar forecasting. Sol Energy 2015;122:804–19. doi:10.1016/j.solener.2015.09.047.

[7] Picault D, Raison B, Bacha S, Casa J De, Aguilera J. Forecasting photovoltaic array power production subject to mismatch losses. Sol Energy 2010;84:1301–9. doi:10.1016/j.solener.2010.04.009.

[8] Chen C, Duan S, Cai T, Liu B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol Energy 2011;85:2856–70. doi:10.1016/j.solener.2011.08.027.

[9] Fernandez-jimenez LA, Muñoz-jimenez A, Falces A, Garcia-garrido E, Lara-santillan PM, Zorzano-alba E, et al. Short-term power forecasting system for photovoltaic plants. Renew Energy 2012;44:311–7. doi:10.1016/j.renene.2012.01.108.

[10] Su Y, Chan L, Shu L, Tsui K. Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems. Appl Energy 2012;93:319–26. doi:10.1016/j.apenergy.2011.12.052.

[11] Narvarte L, Almeida MP, Perpin O. PV power forecast using a nonparametric PV model 2015;115:354–68. doi:10.1016/j.solener.2015.03.006.

[12] Chu Y, Urquhart B, Gohari SMI, Pedro HTC, Kleissl J, Coimbra CFM. ScienceDirect Short-term reforecasting of power output from a 48 MWe solar PV plant. Sol Energy 2015;112:68–77. doi:10.1016/j.solener.2014.11.017.

[13] Nobre M, Severiano CA, Karthik S, Kubis M, Zhao L, Martins FR, et al. PV power conversion and short-term forecasting in a tropical , densely-built environment in Singapore Andr e 2016;94:496–509.

doi:10.1016/j.renene.2016.03.075.

[14] Rana M, Koprinska I, Agelidis VG. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers Manag 2016;121:380–90. doi:10.1016/j.enconman.2016.05.025.

[15] Baharin KA, Rahman HA. Short-term forecasting of solar photovoltaic output power for tropical climate using ground-based measurement data 2016;53701. doi:10.1063/1.4962412.

[16] Moslehi S, Maerefat M, Arababadi R. Applicability of Radiant Heating-Cooling Ceiling Panels in Residential Buildings in Different Climates of Iran. Procedia Eng., vol. 145, Elsevier Ltd; 2016, p. 18–25.

[17] Baniassadi A, Momen M, Amidpour M. A new method for optimization of Solar Heat Integration and solar fraction targeting in low temperature process industries. Energy 2015;90:1674–81. doi:10.1016/j.energy.2015.06.128.

[18] Myers DR, Emery K, Renewable N, Blvd C, Co G. Terrestrial Solar Spectral Modeling Tools and Applications for Photovoltaic Devices 1976:1683–6.

[19] Gordon JM, Reddy TA, Solar A, Unit C. Generalized Capacity Factors for Grid-intertie Solar Photovoltaic Systems 1988;23:127–37.

[20] Evans DL. Simplified Method for Predicting Photovoltaic Array Output 1981;27:555–60.

[21] Copper JK, Sproul AB, Jarnason S. Photovoltaic ( PV ) performance modelling in the absence of onsite measured plane of array irradiance ( POA ) and module temperature. Renew Energy 2016;86:760–9. doi:10.1016/j.renene.2015.09.005.

[22] Faiman D. Assessing the Outdoor Operating Temperature of Photovoltaic Modules 2008:307–15. doi:10.1002/pip.

[23] Koehl M, Heck M, Wiesmeier S, Wirth J. Modeling of the nominal operating cell temperature based on outdoor weathering. Sol Energy Mater Sol Cells 2011;95:1638–46. doi:10.1016/j.solmat.2011.01.020.

[24] Mekhilef S, Saidur R, Kamalisarvestani M. Effect of dust , humidity and air velocity on efficiency of photovoltaic cells. Renew Sustain Energy Rev 2012;16:2920–5. doi:10.1016/j.rser.2012.02.012.

[25] Schwingshackl C, Petitta M, Wagner JE, Belluardo G, Moser D. Wind effect on PV module temperature : Analysis of different techniques for an accurate estimation. Energy Procedia 2013;40:77–86. doi:10.1016/j.egypro.2013.08.010.

[26] Kaldellis JK, Kapsali M, Kavadias KA. Temperature and wind speed impact on the ef fi ciency of PV installations . Experience obtained from outdoor measurements in Greece. Renew Energy 2014;66:612–24. doi:10.1016/j.renene.2013.12.041.

[27] King DL, Boyson WE, Kratochvill JA. Photovoltaic Array Performance Model 2004.

[28] Soto W De, Klein SA, Beckman WA. Improvement and validation of a model for photovoltaic array performance 2006;80:78–88. doi:10.1016/j.solener.2005.06.010.

[29] Lo Brano V, Orioli A, Ciulla G, Gangi A Di. An improved five-parameter model for photovoltaic modules. Sol Energy Mater Sol Cells 2010;94:1358–70. doi:10.1016/j.solmat.2010.04.003.

[30] Lo Brano V, Orioli A, Ciulla G. On the experimental validation of an improved five-parameter model for silicon photovoltaic modules. Sol Energy Mater Sol Cells 2012;105:27–39. doi:10.1016/j.solmat.2012.05.028.

[31] Dolara A, Leva S, Manzolini G. Comparison of different physical models for PV power output prediction. Sol Energy 2015;119:83–99. doi:10.1016/j.solener.2015.06.017.

[32] A R. Dynamic Model of a Three Layer Photovoltaic Panel and the Optimal Use of Spill-Over Light in a Concentrator Photovoltaic System. Ben Guirion University of the Negev, Sede Boqer campus, Israel., 2009.

[33] PVsyst Contextual Help: User's Guide 2012.

[34] Solar Advisor Model (SAM). Golden, Colorado: 2016.

[35] Kaplani E, Kaplanis S. Thermal modelling and experimental assessment of the dependence of PV module temperature on wind velocity and direction , module orientation and inclination. Sol Energy 2014;107:443–60. doi:10.1016/j.solener.2014.05.037.

[36] Kurtz S, Miller D, Kempe M, Bosco N, Wohlgemuth J. Evaluation of High-Temperature Exposure of Photovoltaic Modules Preprint 2009.

[37] Tamizhmani G, Ji L, Tang Y, Petacci L, Osterwald C. Photovoltaic Module Thermal / Wind Performance:: Long -Term Monitoring and Model Development For Energy Rating 2003:936–9.

[38] Myers DR, Emery K, Gueymard C. Revising and Validating Spectral Irradiance Reference Standards for Photovoltaic Performance Evaluation. J Sol Energy Eng 2004;126:567–74. doi:10.1115/1.1638784.

[39] Duffie, J.A., Beckman WA. Solar Engineering of Thermal Processes. 3rd ed. New York: John Wiley & Sons; 2006.

Table 1– Goodness-of-fit statistics of PV power models identified from whole year vs monthly data, Gordon and Reddy model- eq. 4 (MAE and RMSE values are in Watts)

| Criteria | models constructed from whole-year data and used to predict … | | | | models constructed from individual month data | | |
|---|---|---|---|---|---|---|---|
| | **Annual** | **Mar.** | **Jul.** | **Dec.** | **Mar.** | **Jul.** | **Dec.** |
| **MAE** | 12.9 | 8.4 | 19.1 | 7.9 | 11.6 | 16.3 | 9.8 |
| **RMSE** | 21.1 | 16.6 | 28.3 | 16.2 | 19.43 | 25.8 | 15.9 |
| **CV-RMSE** | 0.079 | 0.078 | 0.075 | 0.128 | 0.091 | 0.069 | 0.125 |

Table 2 – Results of the PV Power output models for both the training datasets (internal prediction) and testing datasets (validation). MAE and RMSE are in Watts

| Database | Location (time step) | Criteria | **Internal Prediction** | | **Model Validation** | |
|---|---|---|---|---|---|---|
| | | | **Gordon and Reddy (eq. 4)** | **Myers and Emery without wind (eq. 2)** | **Gordon and Reddy (eq. 4)** | **Myers and Emery (eq. 2)** |
| NREL | Cocoa (5-min) | MAE | **1.3** | 1.6 | **1.3** | 1.7 |
| | | MBE | **-** | - | **0.0** | -0.1 |
| | | RMSE | **1.8** | 2.1 | **1.7** | 2.2 |
| | | CV-RMSE | **0.032** | 0.040 | **0.032** | 0.040 |
| | Eugene (5-min) | MAE | **1.3** | 1.6 | **1.3** | 1.6 |
| | | MBE | **-** | - | **0.1** | 0.0 |
| | | RMSE | **2.0** | 2.3 | **1.9** | 2.2 |
| | | CV-RMSE | **0.046** | 0.053 | **0.044** | 0.052 |
| | Golden (15-min) | MAE | **1.7** | 2.0 | **1.7** | 2.0 |
| | | MBE | **-** | - | **0.0** | -0.2 |
| | | RMSE | **2.3** | 2.6 | **2.3** | 2.6 |
| | | CV-RMSE | **0.048** | 0.054 | **0.048** | 0.054 |
| SRML | Bend (5-min) | MAE | **74.5** | 83.3 | **74.9** | 83.8 |
| | | MBE | **-** | - | **-7.5** | -9.3 |
| | | RMSE | **125.7** | 136.6 | **127.7** | 138.3 |
| | | CV-RMSE | **0.086** | 0.093 | **0.087** | 0.094 |
| | Portland (1-hr) | MAE | **12.9** | 16.6 | **12.3** | 15.6 |
| | | MBE | **-** | - | **-0.6** | -0.2 |
| | | RMSE | **21.1** | 25.5 | **19.6** | 23.5 |
| | | CV-RMSE | **0.079** | 0.095 | **0.073** | 0.088 |

Table 3 – Model goodness-of-fit results of considering wind effects in power output prediction models (eq. 2). MAE, MBE and RMSE are in Watts

| Database | Location | Criteria | Myers and Emery | Myers and Emery without wind effects |
|---|---|---|---|---|
| SRML | Bend | **MAE** | 83.3 | 83.4 |
| | | **RMSE** | 136.6 | 136.6 |
| | | **CV-RMSE** | 0.093 | 0.093 |
| | Portland | **MAE** | 16.7 | 16.6 |
| | | **RMSE** | 25.5 | 25.5 |
| | | **CV-RMSE** | 0.095 | 0.095 |

Table 4 – Results of models for cell temperature models (MAE, MBE and RMSE values are in ˚C)

| Data base | Location | Criteria | Model Training | | | | Model Testing | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | Kurtz model form (eq. 9) | Mani model (eq. 10) | NOCT model Form (eq. 12) | NOCT (eq. 11) | Kurtz model form (eq. 9) | Mani model (eq. 10) | NOCT model Form (eq. 12) |
| NREL | Cocoa | **MAE** | 2.3 | **2.0** | 2.3 | - | 2.2 | **2.0** | 2.2 |
| | | **MBE** | - | **-** | - | - | 0.5 | **0.0** | 0.5 |
| | | **RMSE** | 3.2 | **2.8** | 3.2 | - | 3.2 | **2.8** | 3.2 |
| | | **CV-RMSE** | 0.087 | **0.077** | 0.087 | - | 0.087 | **0.077** | 0.087 |
| | Eugene | **MAE** | 2.1 | **1.9** | 2.1 | - | 2.1 | **1.9** | 2.1 |
| | | **MBE** | - | **-** | - | - | 0.7 | **0.0** | 0.7 |
| | | **RMSE** | 2.9 | **2.7** | 2.9 | - | 2.9 | **2.6** | 2.9 |
| | | **CV-RMSE** | 0.113 | **0.103** | 0.113 | - | 0.112 | **0.103** | 0.112 |
| | Golden | **MAE** | 2.8 | **2.7** | 2.8 | - | 2.7 | **2.6** | 2.7 |
| | | **MBE** | - | **-** | - | - | -0.4 | **-0.1** | -0.4 |
| | | **RMSE** | 3.6 | **3.6** | 3.6 | - | 3.61 | **3.5** | 3.6 |
| | | **CV-RMSE** | 0.111 | **0.110** | 0.111 | - | 0.110 | **0.108** | 0.110 |
| SRML | Bend | **MAE** | 5.2 | **4.2** | 4.6 | 4.6 | 5.3 | **4.2** | 4.7 |
| | | **MBE** | - | **-** | - | - | 2.8 | **0.1** | -0.4 |
| | | **RMSE** | 6.7 | **5.4** | 5.7 | 5.7 | 6.8 | **5.4** | 5.7 |
| | | **CV-RMSE** | 0.203 | **0.170** | 0.182 | 0.182 | 0.203 | **0.168** | 0.179 |
| | Portland | **MAE** | 2.8 | **1.5** | 1.8 | 1.8 | 2.8 | **1.6** | 1.9 |
| | | **MBE** | - | **-** | - | - | 2.1 | **0.0** | -0.4 |
| | | **RMSE** | 3.5 | **2.0** | 2.3 | 2.3 | 3.5 | **2.0** | 2.3 |
| | | **CV-RMSE** | 0.114 | **0.066** | 0.078 | 0.078 | 0.115 | **0.068** | 0.077 |

Table 5 – Results of the Mani black-box model (eq. 10) with and without the inclusion of incidence angle modifier $K_\eta$ effects (MAE and RMSE are in ºC)

| Database | Location | Criteria | Mani (black-box) | Mani (black-box) with $K_\eta$ |
|---|---|---|---|---|
| NREL | Cocoa (5-min) | **MAE** | 2.0 | **2.0** |
| | | **RMSE** | 2.8 | **2.8** |
| | | **CV-RMSE** | 0.077 | **0.075** |
| | Eugene (5-min) | **MAE** | 1.9 | **1.9** |
| | | **RMSE** | 2.7 | **2.6** |
| | | **CV-RMSE** | 0.103 | **0.101** |
| | Golden (15-min) | **MAE** | 2.7 | **2.5** |
| | | **RMSE** | 3.6 | **3.4** |
| | | **CV-RMSE** | 0.110 | **0.105** |
| SRML | Bend (5-min) | **MAE** | 4.2 | **4.1** |
| | | **RMSE** | 5.4 | **5.2** |
| | | **CV-RMSE** | 0.170 | **0.165** |
| | Portland (1-hr) | **MAE** | **1.5** | 1.6 |
| | | **RMSE** | **2.0** | 2.1 |
| | | **CV-RMSE** | **0.066** | 0.068 |

Table 6 – Power prediction model results using measured and predicted cell temperature (MAE, and RMSE values are reported in ˚C)

| Database | Location | Criteria | Model Training | | Model Testing | |
|---|---|---|---|---|---|---|
| | | | Evans Model based on $T_{cell,m}$ | Evans Model based on $T_{cell,p}$ | Evans Model based on $T_{cell,m}$ | Evans Model based on $T_{cell,p}$ |
| **NREL** | **Cocoa** | **MAE** | 1.7 | 1.6 | 1.7 | 1.6 |
| | | **MBE** | - | - | -0.2 | -0.1 |
| | | **RMSE** | 2.2 | 2.1 | 2.2 | 2.2 |
| | | **CV-RMSE** | 0.041 | 0.040 | 0.041 | 0.040 |
| | **Eugene** | **MAE** | 1.6 | 1.6 | 1.6 | 1.6 |
| | | **MBE** | - | - | -0.2 | -0.1 |
| | | **RMSE** | 2.4 | 2.3 | 2.3 | 2.3 |
| | | **CV-RMSE** | 0.055 | 0.054 | 0.054 | 0.053 |
| | **Golden** | **MAE** | 2.0 | 2.0 | 2.0 | 2.0 |
| | | **MBE** | - | - | -0.1 | -0.1 |
| | | **RMSE** | 2.6 | 2.6 | 2.7 | 2.6 |
| | | **CV-RMSE** | 0.055 | 0.054 | 0.055 | 0.054 |
| **SRML** | **Bend** | **MAE** | 78.3 | 81.4 | 78.1 | 81.1 |
| | | **MBE** | - | - | -22.9 | -22.6 |
| | | **RMSE** | 142.2 | 144.1 | 143.6 | 145.6 |
| | | **CV-RMSE** | 0.097 | 0.098 | 0.097 | 0.099 |
| | **Portland** | **MAE** | 16.0 | 16.0 | 15.2 | 15.2 |
| | | **MBE** | - | - | -1.3 | -1.6 |
| | | **RMSE** | 26.3 | 26.5 | 24.3 | 24.5 |
| | | **CV-RMSE** | 0.098 | 0.099 | 0.091 | 0.092 |

Table 7- PV power prediction results for two scenarios: when only global horizontal irradiance ($I$) is measured and when IT measurements are available

| Location | TMY3 data Location | Only $I$ is measured | | | $I_T$ is measured | | |
|---|---|---|---|---|---|---|---|
| | | MAE | RMSE | CV-RMSE | MAE | RMSE | CV-RMSE |
| **Cocoa, FL** | **Melbourne, FL** | 2.1 | 3.4 | 0.057 | 1.3 | 1.7 | 0.032 |
| **Eugene, OR** | **Eugene, OR** | 2.5 | 4.4 | 0.094 | 1.3 | 1.9 | 0.044 |
| **Golden, CO** | **Denver, CO** | 2.8 | 4.0 | 0.075 | 1.7 | 2.3 | 0.048 |

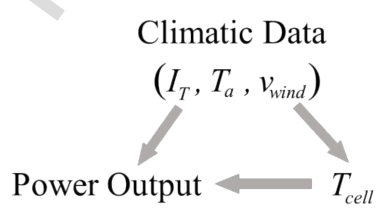



Figure 1- Schematic of data collection and modeling scenarios (2-column fitting image)

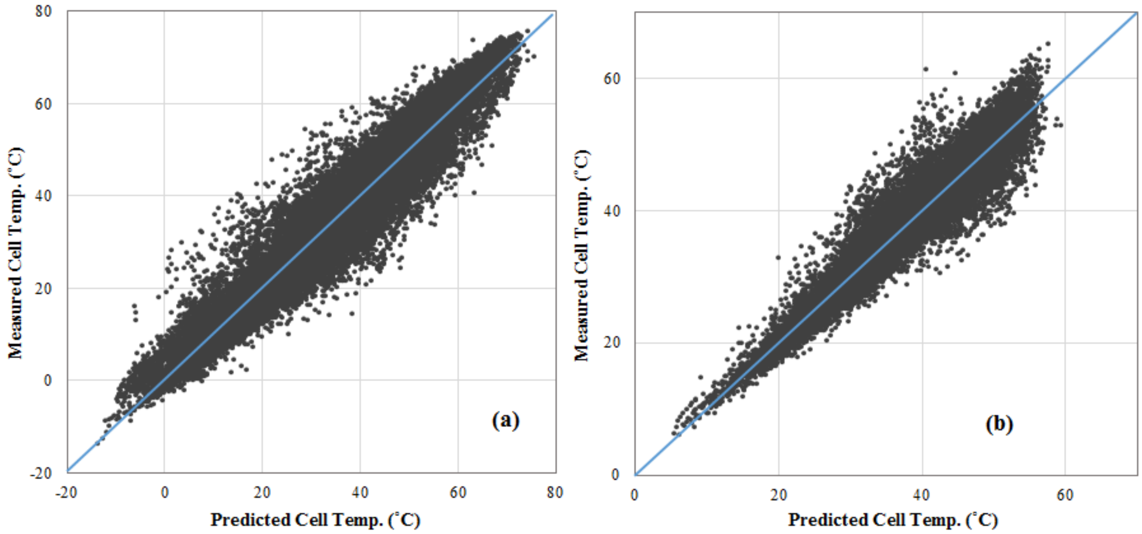



Figure 2- Measured versus predicted cell temperatures using Mani model (with $K_\eta$ effects) for (a) Bend and (b) Cocoa during model training (single column fitting image)

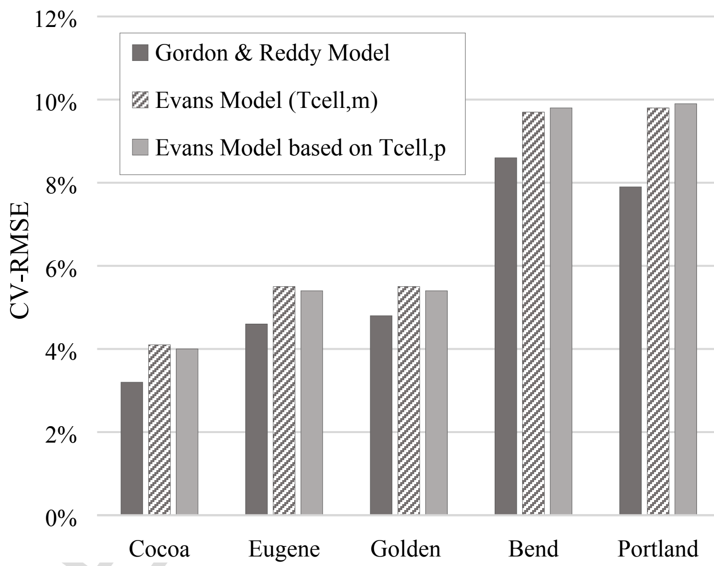



*Figure 3*– Comparison prediction of PV power output in different scenarios (2-column fitting image)

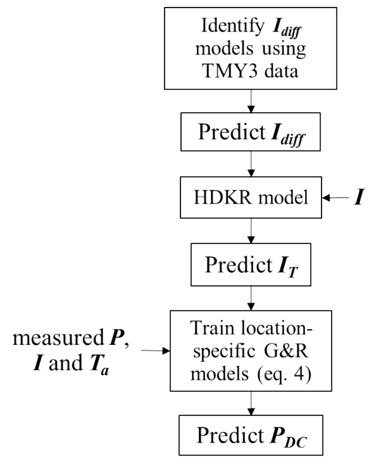



Figure 4- PV power output prediction flowchart in the absence of POA and diffuse radiation measurements (2-column fitting image)