Received: 25 April 2022 | Revised: 6 November 2022 | Accepted: 22 November 2022

IET Renewable Power Generation

DOI: 10.1049/rpg2.12651

ORIGINAL RESEARCH

Comparative analysis of photovoltaic performance metrics for reliable performance loss rate

HyunYong Lee | **Jungi Lee** | **Nac-Woo Kim** | **Byung-Tak Lee**

Honam Research Center (HRC), Electronics and Telecommunications Research Institute (ETRI), Buk-gu, Gwangju, Republic of Korea

Correspondence
HyunYong Lee, Honam Research Center (HRC), Electronics and Telecommunications Research Institute (ETRI), 176-11 Cheomdan Gwagi-ro, Buk-gu, Gwangju, Republic of Korea.
Email: hyunyonglee@etri.re.kr

Funding information
Electronics and Telecommunications Research Institute, Grant/Award Number: 22ZK1100

Abstract

A reliable performance loss rate of photovoltaic systems requires accurate and reliable performance metrics. This study proposes a systematic method for assessing the performance metrics, particularly predicted power models in terms of both accuracy and uncertainty. The gist of the proposed method is to examine how accurately a predicted power model predicts the manipulated degradation in a controlled environment. For this, the proposed method divides a given dataset evenly into base data (to generate reference performance) and test data (to generate test performance via manipulation) so that the two data have similar features. The proposed method also utilizes the bootstrap iteration to derive a reliable assessment. The novelty of this study is that the proposed method estimates both the accuracy and uncertainty of arbitrary predicted power models, which is missing in existing works. Extensive experiments using the proposed method with real-world datasets reveal the followings. One interesting observation is that a well-known machine learning prediction model, not considered in existing works, exhibits the best performance in terms of both accuracy and uncertainty. Existing predicted power models require different experiment settings to produce reliable performance. The number of test data is closely related to uncertainty, but not much related to accuracy.

1 | INTRODUCTION

As one of the fastest-growing renewable energy technologies, photovoltaic (PV) systems are increasing their share in the energy and power mix worldwide. For example, PV-generated power had a share of over 10% in eight countries, including Luxembourg, Chile, and Australia, in 2021 [1]. Owing to this reason, it is important to know how the PV power output decreases over time.

One popular option to predict and assess the long-term performance degradation of PV systems is performance loss rate (PLR). The PLR is a parameter that indicates the decline of the power output over time and is provided in %/year. The PLR does not only indicate the irreversible physical degradation of PV systems but also measures performance-reducing events, which can be reversible or preventable through good maintenance practices. The PLR calculation includes two main components: performance metrics (e.g. predicted power models [2–5]) to generate performance time series and statistical methods (e.g. Year-on-Year [6]) to estimate the PLR from the performance time series. Therefore, the quality of PLR is heavily affected by the quality of the performance metrics and statistical methods.

To obtain a reliable PLR, it is necessary to assess the performance metrics and statistical methods in terms of accuracy and uncertainty. Existing studies address this issue only partially. Some studies examine performance metrics only in terms of uncertainty [7–10]. Regarding the statistical methods, some studies try to understand the sources of uncertainties [9] and the impact of missing data [11]. Although some studies propose new performance metrics and statical methods for better PLRs [12–17], a comparison with existing well-known methods is missing.

To this end, this study proposes a systematic method to assess performance metrics in terms of both accuracy and uncertainty. This study focuses on the performance metrics solely because that is the basis of the PLR calculation pipeline. In particular, among various performance metrics, this study focuses on the predicted power models that are widely used. The main idea of the proposed method is to examine how accurately a predicted power model predicts the manipulated degradation in a controlled environment. For this, a particular original dataset that includes weather conditions and PV power output is divided into base data (used to produce reference performance) and test data (used to produce test performance after manipulating the PV power output). The intention of this approach is to calculate PV performance degradation by comparing the reference performance and test performance and to determine how close the calculated PV performance degradation is to the actual manipulated degradation. To realize the manipulation-based approach successfully, the original dataset is evenly divided (i.e. odd-numbered data are classified as base data and even-numbered data are classified as test data). The rationale behind this approach is that weather conditions do not change abruptly across a few data points. To get reliable outputs, the proposed method applies bootstrap iteration. The proposed method calculates the performance degradation (i.e. the average of the results) and uncertainty (i.e. the standard deviation of the results) through 1,000 experiments with 65% random sampling of the data.

This study also proposes two approaches to utilizing predicted power models in calculating performance degradation. The first approach is to use dual model. Base and test models are built with the base and test data, respectively. Thereafter, representative weather conditions are applied to the two models to determine the reference performance and test performance. The second approach is to use a single model. A base model is built with the base data, and the weather conditions of the test data are applied to the base model to get the reference performance. The PV power output of the test data is used as the test performance.

To examine the feasibility of the proposed method and assess the predicted power models, this study conducts extensive experiments with publicly available datasets, including National Renewable Energy Lab (NREL) datasets [18, 19] and Regional Test Center (RTC) datasets [20–22]. In addition to the well-known predicted power models, including XbX [4], PVUSA [2], and PVWatts [3], this study examines Light Gradient Boosting Machine (LGBM) [23], which is a well-known machine learning prediction model. This study first examines the effects of representative weather conditions in the dual model approach. For the representative weather conditions, this study considers standard test conditions (STC), nominal operating cell temperature (NOCT) conditions, and the mean value of the base data (MEAN). This study also considers the case where the average of the results of STC, NOCT, and MEAN is used (WSUM). The experiment results first show that models require different representative weather conditions to produce promising results. XbX, PVUSA, and LGBM show the best performance with MEAN, NOCT, and WSUM, respectively. XbX, PVUSA, and LGBM show the worst performance with STC, STC, and MEAN, respectively. Under the representative weather conditions that result in the best performance for the dual model approach, this study compares the models in the dual model approach and the single model approach. The experiment results show that LGBM and XbX show good performance in terms of both accuracy and uncertainty. More specifically, LGBM in the single model approach shows the best performance. Another observation is that models tend to show different performances in the dual model approach and single model approach. XbX shows similar performance in both approaches. On the contrary, PVUSA and LGBM show better performance in the dual model approach and single model approach, respectively. This study also examines the effect of test data to see how much data is required to produce a reliable performance time series. The experiment results first show that a small number of data points may be enough to estimate accurate performance degradation. For example, models with 18 days and with 182 days show similar accuracy. However, the more data is given, the more reliability (i.e. less uncertainty) can be achieved.

The main contributions of this study are as follows.

- This study proposes a systematic method for assessing the predicted power models in terms of both accuracy and uncertainty. Please note that most existing works handle accuracy or uncertainty partially.
- This study compares the predicted power models in terms of both accuracy and uncertainty through extensive experiments following the proposed methods.
- This study shows that a machine learning prediction model, not considered in existing works, can be used as a predicted power model for producing a reliable performance time series. This shows the possibility of using machine learning or deep learning prediction models as reliable performance metrics.

The rest of this paper is as follows. Section 2 discusses related works. Section 3 introduces background about PLR calculation. Section 4 proposes a systematic method for comparative analysis of the performance metrics in terms of both accuracy and uncertainty. Section 5 introduces the results of comparative analysis through extensive experiments. Section 6 finally concludes this paper.

2 | RELATED WORK

Few existing studies conduct comparative analysis regarding the performance metrics and the statistical methods in the calculation of PLRs. In [10], the authors propose and compared different performance metrics to extract reliable long-term performance indicators in terms of uncertainty. The metrics include the performance ratio, the performance ratio fitted to two sinusoidal functions (emulating the climatic influence and a decaying trend), time-series decomposition, and the metrics intended to utilize the physical properties of the material to correct for seasonal fluctuations. In [7], the authors assess IV curve-based performance metrics including DC performance ratio ad DC STC performance ratio in terms of uncertainty. In [8], the authors compare the predicted power models including XbX, PVUSA, and 6K in terms of uncertainty. Like the observations of this study, in this work, XbX leads to the most stable PLR. In [24], to identify areas of improvement of PV monitoring solutions, the authors test different performance metrics (including performance ratio, power performance index with physical and machine learning-based modeling) in terms of uncertainty. Most existing work examines the performance metrics just in terms of uncertainty. On the contrary, this study examines the performance metrics in terms of both accuracy and uncertainty.

In [25], the authors compare the statistical methods for various PV technologies using the results reported in the published papers. In [26], the authors examine various combinations of performance metrics (including performance ratio, XbX, PVUSA, 6K, and PVWatts) and statistical methods (including LR, CSD, HW, and YoY). However, they focus on PLR itself, not the performance metrics. In [9], the authors try to understand sources of uncertainties in PV degradation rates using physical degradation models. They show that there are three major sources of uncertainties: climate variables estimation, PV modules reliability, and statistical methods for calculating PLRs. In [11], the authors study the impact of the time period and missing data on PLR. Through experiments with emulated degradation levels and imputed missing data, they show that the accuracy of the statistical methods (including YoY, autoregressive integrated moving average (ARIMA) and robust principal component analysis) is largely affected by the time period. They also show that the estimated PLR is strongly affected by the amount of missing data.

Some work try to propose new performance metrics and statistical methods for better PLRs. In [12], the authors introduce the normalized efficiency of a PV system as an additional PV performance metric for analysis purposes. The normalized efficiency can be implemented on time scales ranging from seconds to days and longer. In [13], the authors try to quantify the effect of dynamic environmental stresses on the power degradation of the module. For this purpose, they explain the fusion of the physics-based material degradation mechanism with the statistics-based data modeling approach. They show that the degradation of PV modules is mainly associated with the module construction type and climatic conditions. In [14], the authors first argue that most existing degradation modeling approaches are susceptible to bias due to inverter clipping, module soiling, temporary outages, seasonality, and sensor degradation. Then, they propose a way of determining PV degradation through modeled clear-sky irradiance data and a robust year-over-year rate calculation. In [15], the authors propose a novel unsupervised machine learning approach that can be applied to PV system degradation estimation. The proposed approach just requires a measured power as an input. In [16], the authors propose a novel method for modeling PV system performance loss rate (PLR) through a self-regulated multi-step algorithm. The proposed method automatically detects the number and positions of breakpoints in nonlinear performance time series and divides the performance trend into an adequate number of linear segments. In [17], using the cumulative damage model, the authors establish a mathematical model between climatic stresses and performance degradation. They also propose regional clustering based on climatic stressors to predict the degradation at different locations. Unlike existing works that try to propose new performance metrics, this study proposes a systematic method to evaluate existing and newly proposed performance metrics in terms of both accuracy and uncertainty.

3 | BACKGROUND: PLR CALCULATION

PLR calculation consists of two steps: calculation of performance time series over a certain period of time (with performance metrics) and estimation of PLR (with statistical methods) (Figure 1). In this way, the quality of the estimated PLR heavily depends on the quality of the performance time series and the statistical methods. In other words, the accuracy and uncertainty of the performance metrics and the statistical methods are important for acquiring reliable PLR. This study is interested in assessing the performance metrics in terms of both accuracy and uncertainty, which is missing in most existing works.

3.1 | Performance metrics: Predicted power models

The term *performance metric* used in this manuscript and other related materials (e.g. [26]) is used to indicate a certain measure that provides information about the performance of a PV system. In other words, the performance metrics are data-driven or physics-based models that are used to estimate the predicted power output of a PV system. Common performance metrics used to calculate the PV performance can be grouped into (1) electrical parameters from IV curves [27, 28], (2) normalized and scaled ratings [29, 30], and (3) predicted power models [2–5, 23]. Among them, this study is interested in the predicted power models because they are widely used. The predicted power models work as follows. A power prediction model is built to predict power output as a function of weather conditions over a period of time. Then, representative weather conditions are applied to the built model. This produces a predicted power output at the given conditions. The followings are the well-known predicted power models. Throughout this paper, P_{pred} is used to indicate the predicted power output derived by the predicted power models.

3.1.1 | PVUSA

PVUSA model [2] is a physics-based model. The PVUSA model is described as follows.

$$P_{pred} = G_{POA}(\beta_0 + \beta_1 G_{POA} + \beta_2 T_{amb} + \beta_3 WS), \tag{1}$$

where G_{POA} is the plane of array (POA) irradiance, T_{amb} is the ambient temperature (in $^\circ C$), and WS is the wind speed (in m/s). β_i are the model's coefficients to be determined while building a model for given data. The assumption of this model is that the current of a PV panel is a function of G_{POA}, the voltage is a

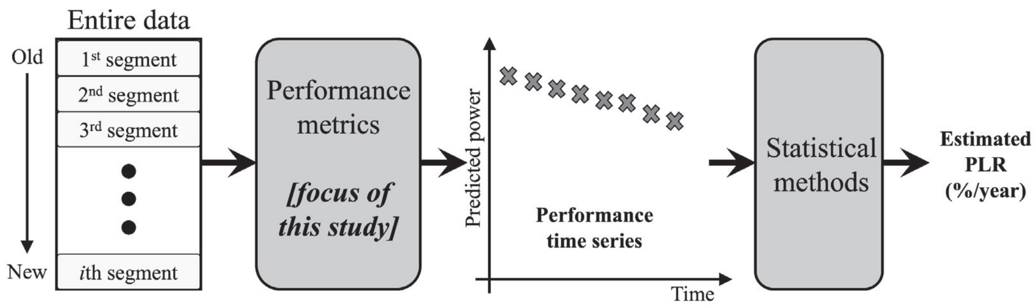


FIGURE 1 High-level illustration of PLR calculation

function of G_{POA}, and the module temperature is predicted by T_{amb} and WS.

3.1.2 | PVWatts

PVWatts model [3] is a simple model as follows:

$$P_{pred} = P_{mea}\frac{G_{POA}}{G_0}(1 + \gamma(T_{cell} - T_0)), \tag{2}$$

where G_0 and T_0 are the reference irradiance and cell temperature, respectively. 1,000 W/m^2, 25°C, and −0.004 are typically used for G_0, T_0, and γ. G_{POA} , P_{mea}, and T_{cell} are the POA irradiance, PV power output, and cell temperature, respectively. In this case, there are no model coefficients to be determined. Therefore, without building a model, the PVWatts model can be used directly.

3.1.3 | XbX

XbX model [4] is a data-driven multiple regression predictive model. The XbX model is as follows:

$$P_{pred} = \beta_0 + \beta_1 G + \beta_2 T + \epsilon, \tag{3}$$

where G is the irradiance term, T is the temperature term, β_i are the model's coefficients to be determined while building a model for given data and ϵ is the residual error between the model and the given data. Unlike the models above, XbX has the flexibility of using POA or Global Horizontal Irradiance (GHI) for G and air or module temperature for T. The X in the name refers to a given time step that the power prediction model is built over. For example, a model built on a day of data would be DbD. The time step is chosen based on the condition of the given data and what modeling will be performed on the given data.

3.1.4 | 6K

6K model [5] is a data-driven model. The name 6K refers to the coefficients fit by the model. The 6K model is as follows:

$$\begin{aligned} G' &= G_{POA}/G_{STC}, \\ T' &= T_{mod} - T_{STC}, \\ P_{pred} &= G'(P_{NP} + k_1 ln(G') + k_2 ln(G')^2 + k_3 T' \\ &\quad + k_4 T' ln(G') + k_5 T' ln(G')^2 + k_6 T'^2). \end{aligned} \tag{4}$$

This model uses POA irradiance (G_{POA}) and module temperature (T_{mod}) but models them as a fraction of standard irradiance (G_{STC}) and a difference from standard temperature (T_{STC}). Additionally, this model requires a nameplate power input (P_{NP}) and always predicts P_{NP} at STC conditions. This model has the model's coefficients, k_i to be determined while building a model for the given data.

3.1.5 | LGBM

In addition to the traditional prediction models above, machine learning models can also be used as the performance metric because it shows promising prediction performance in various areas. One of the popular machine learning prediction models is LGBM [23]. Unlike the models above that follow the pre-defined equations, LGBM derives a model automatically using the given data. In other words, it finds the most-fitting model $f()$ that shows a relationship between input variables (i.e. weather conditions) and PV power output as shown in Equation (5). As long as variables are related to PV power output, any variable can be used as an input variable. In this sense, LGBM has much more flexibility than the models above.

$$P_{pred} = f(weather\ conditions). \tag{5}$$

3.2 | Statistical methods

The goal of the statistical methods is to calculate the trend of the PV performance time series (i.e. calculated by the predicted power models) and translate the slope of the trend to the annual degradation rate in %/year (i.e. PLR).

3.2.1 | Linear regression

The most basic statistical method is linear regression (LR). LR fits Equation (6) to the PV performance time series.

$$y' = \alpha t + \beta, \tag{6}$$

where y' represents the fitted values, α is the slope of the trend and β is the y-intercept. LR is simple, but it is very sensitive to outliers and seasonal variations.

3.2.2 | Classical seasonal decomposition

Classical seasonal decomposition (CSD) is more advanced than LR in extracting the underlying trend from the PV performance time series and overcoming the limitations of the LR method. The CSD method separates seasonality and a certain irregular component from a set of measured time-series data, using a centered moving average, to determine the performance trend over time. It assumes that the seasonal component of PV performance is stable year after year. Therefore, the step of the seasonal

period is usually set to 12 for monthly data. The CSD method requires either the additive model (i.e. Equation 7) or the multiplicative model (i.e. Equation 8) depending on the stability of the seasonal component.

$$y' = T_t + S_t + e_t. \tag{7}$$

$$y' = T_t S_t e_t. \tag{8}$$

In Equations (7) and (8), T_t is the trend, S_t is the seasonal, and e_t is the residual component.

3.2.3 | Holt-winters

Another advanced model-based method is Holt-Winters (HW). The HW method applies triple exponential smoothing to the time series. The triple exponential smoothing takes into account seasonal changes, as well as trends, through the minimization of the squared one-step ahead prediction error, in contrast to the CSD method, which bases the calculation of trend, seasonal component, and residuals on a centered moving average. The HW method requires the following equations:

$$\begin{aligned} y'_{n+l|n} &= m_n + b_n + c_{n-S+l}, l = 1, 2, \dots, \\ m_t &= \alpha_0(y_t - c_{t-S}) + (1-\alpha_0)(m_{t-l} + b_{t-l}), \\ b_t &= \alpha_1(m_t - m_{t-l}) + (1-\alpha_1)b_{t-l}, \\ c_t &= \alpha_2 \frac{y_t}{m_t} + (1-\alpha_2)c_{t-S}, \end{aligned} \tag{9}$$

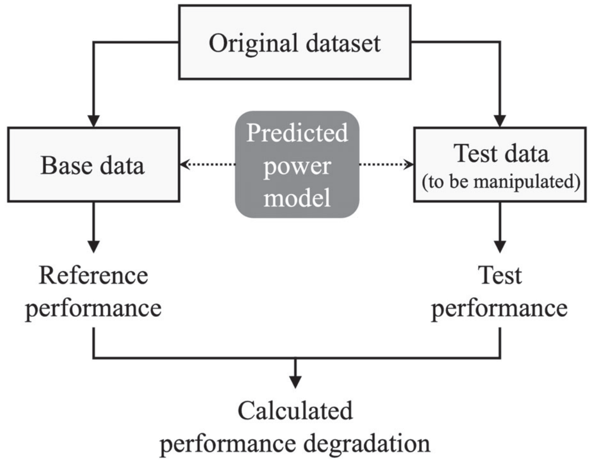


FIGURE 2 Manipulation-based assessment

where m_n is the level component, b_n is the slope component, and c_{n-S+l} is the relevant seasonal component. S is the seasonal period. α_i lies between 0 and 1.

3.2.4 | YoY

The Year-on-Year (YOY) [6] method calculates individual PLR values from points separated by exactly one year, then creates a distribution of individual yearly PLR values. The PLR of the total system is the median of the individual yearly PLRs. This method was developed to be more robust to outliers and seasonality than regression methods, but it does require longer-term data to work effectively.

4 | PROPOSED METHOD FOR COMPARATIVE ANALYSIS

This section proposes a systematic method for assessing the predicted power models in terms of both accuracy and uncertainty.

4.1 | Manipulation-based assessment

Real-world PV-related datasets gathered over several years are available. However, using the raw data in assessing the performance metrics is not preferred because the true performance degradation is unknown. Therefore, this study proposes a manipulation-based assessment in a fully controlled environment.

This study assumes that an original dataset (including weather conditions and corresponding PV power output) over a certain period of time (e.g. the 30-min interval over one year) is given. The first idea of the proposed method is to split the data into two segments: base data and test data (Figure 2). The base data is used to produce reference performance. On the contrary,

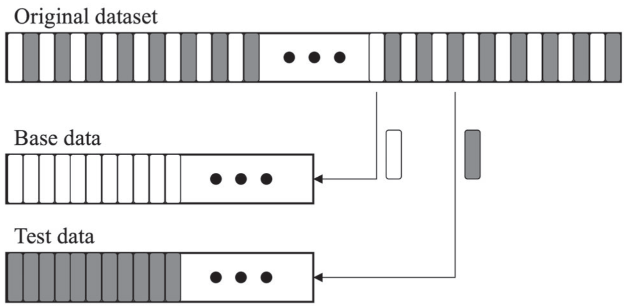


FIGURE 3 Even distribution of data for successful manipulation-based assessment

the test data is manipulated (i.e. decreasing PV power output to simulate the degradation) to produce test performance. The intention of this approach is to calculate PV performance degradation by comparing the reference performance and the test performance and to see how close the calculated PV performance degradation is to the actual manipulated degradation. The performance degradation ratio is calculated as follows:

$$D_{ratio} = \frac{\text{Reference performance - test performance}}{\text{Reference performance}}. \quad (10)$$

The second idea of the proposed method is to split the data evenly (Figure 3). To realize the manipulation-based assessment successfully, the base data and the test data should have similar features. In other words, the base data and the test data have to produce similar performance when there is no manipulation in the test data. To this end, the proposed method splits the original dataset evenly. The odd-numbered data is assigned to the base data and the even-numbered data is assigned to the test data. The rationale behind this approach is that weather conditions and corresponding PV power output do not change abruptly across a few data points. Therefore, the base data and the test data evenly distributed are likely to have similar features. To further examine this point, this study depicts the PV power output of one NREL data described in Section 4. NREL data is 30-min interval data over 1 year. As shown in Figure 4, the base data and the test data show very similar patterns.

4.2 | Model utilization

This subsection describes how the proposed method utilizes the predicted power models to calculate the performance degradation using the base data and the test data.

4.2.1 | Dual model approach

A traditional approach for using predicted power models in calculating PLR is as follows: (1) build a model to predict power output as a function of weather conditions over a period of time and (2) representative weather conditions are applied to the built model. The first proposed approach is to use dual model while adopting the traditional approach (Figure 5a). This approach first builds a base model using the base data. The representative weather conditions are applied to the base model to estimate a reference performance. Then, this approach builds a test model using the manipulated test data. The test model is used to estimate a test performance corresponding to representative weather conditions. The performance degradation is calculated using Equation (10).

This study considers three weather conditions as the representative weather conditions as follows:

- Standard Test Conditions (STC)
  - Irradiance = 1,000 W/m^2
  - Module temperature = 25°C
- Nominal Operating Cell Temperature (NOCT) conditions
  - Irradiance = 800 W/m^2
  - Module temperature = 45+/-3°C
  - Ambient temperature = 20°C
- The mean value of the base data (MEAN) [NREL dataset]
  - Irradiance ≈ 500 W/m^2
  - Module temperature ≈ 24°C
  - Wind speed ≈ 2.3 m/s
  - Dew point ≈ 13°C
  - Relative humidity ≈ 55%

The STC is the condition that is typically used by the PV panel manufacturer. However, it is sometimes difficult to realize the real-world operation of the modules. On the other hand, the NOCT conditions are much more representative of normal operation. In addition to the NOCT conditions and to not cause extrapolations of a model, this study also considers the MEAN conditions because their values always remain in the range of available weather conditions of given data.

4.2.2 | Single model approach

The second proposed approach is to use a single model (Figure 5b). This approach first builds a base model using the base data. Then, this approach applies the weather conditions of the manipulated test data to the base model to estimate a reference performance. PV power output of the manipulated test data is used as test performance. The performance degradation is calculated using Equation (10). The benefit of this single model approach over the dual model approach is that the single model approach does not care about the representative weather conditions as long as the base data and the test data show similar weather conditions.

4.2.3 | PVWatts

PVWatts is a physics-based model. Therefore, unlike other predicted power models, it does not need to be trained. For PVWatts model, this study calculates the reference performance and the test performance as follows:

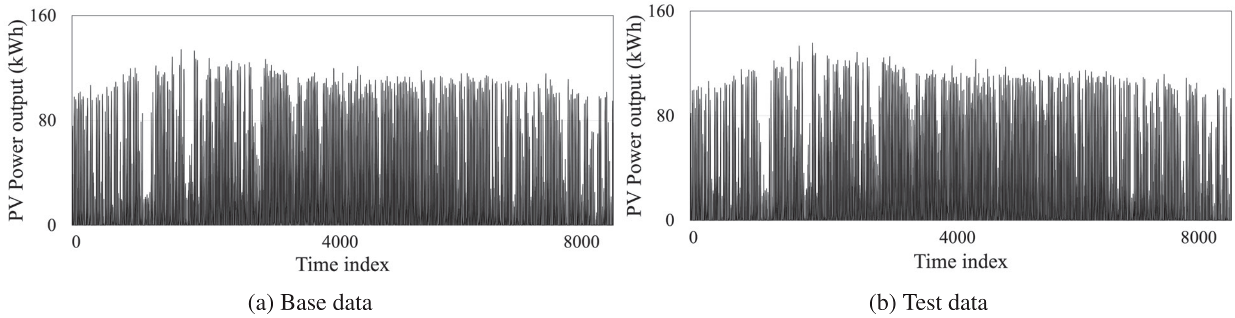


FIGURE 4 The base data and the test data from the same original dataset

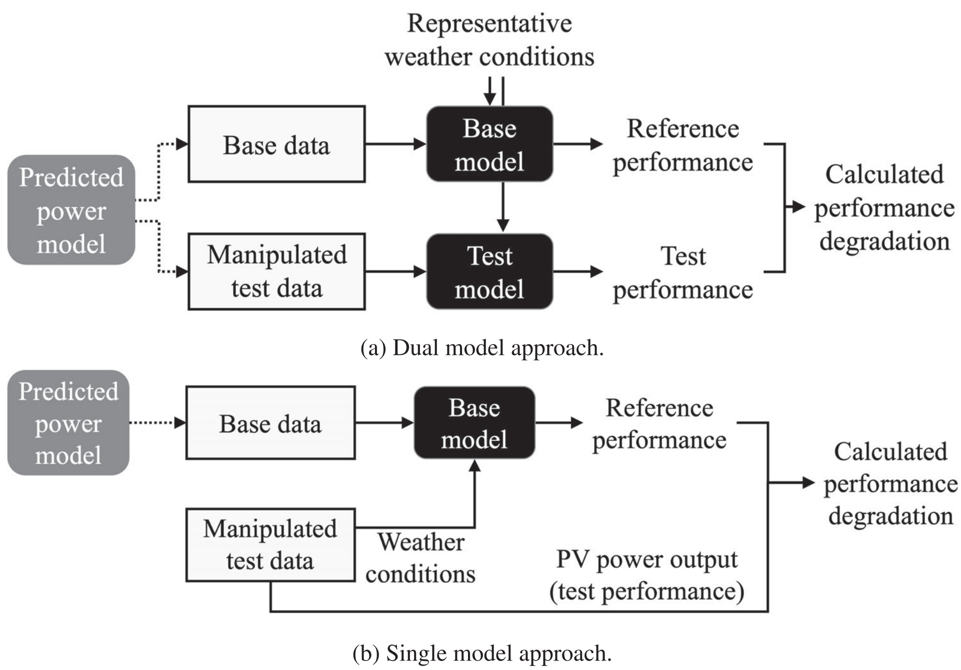


FIGURE 5 Two approaches of utilizing models to generate performance degradation

$$\text{Reference perf.} = \frac{1}{N_b}\sum_i P^i_{pred_b},$$
$$\text{Test perf.} = \frac{1}{N_t}\sum_i P^i_{pred_t}, \tag{11}$$

where $P^i_{pred_b}$ is the predicted power of the ith base data, $P^i_{pred_t}$ is the predicted power of the ith test data, N_b is the number of the base data, and N_t is the number of the test data. The performance degradation is calculated using Equation (10).

4.3 | Accuracy and uncertainty

To calculate reliable performance degradation by the predicted power models, this study applies bootstrap iteration without replacement (Figure 6). The bootstrap iteration randomly samples 65% of the base data and the test data independently. Then, the sampled base data and the sampled test data (after applying manipulation) are used as inputs to the predicted power models. This procedure is repeated 1,000 times. Finally, final performance degradation is calculated by averaging 1,000 calculated performance degradations. The uncertainty of a model is also estimated by calculating a standard deviation of 1,000 calculated performance degradations. A model is considered to be more accurate as its final performance degradation is close to the actual manipulated degradation. Similarly, a model is regarded as more reliable as its uncertainty decreases.

5 | COMPARATIVE ANALYSIS

This section introduces the observations from extensive experiments following the methods described in Section 3.

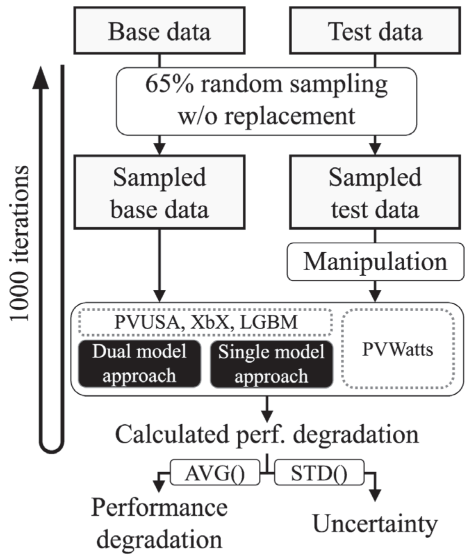


FIGURE 6 Bootstrap iteration-based approach

5.1 | Dataset

For a comparative analysis, this study uses two publicly available datasets: NREL dataset and RTC dataset. For the NREL dataset, historical weather data and PV power output data are downloaded. The weather data acquired from the National Solar Radiation Database (NSRDB) data viewer [18] is 30-min interval data. The weather data includes GHI, dew point, wind speed, relative humidity, and temperature. This study uses PV power output data (acquired from NREL Solar Power Data for Integration Studies [19]) of Texas, USA. More specifically, this study uses NREL data of the seven PV plants whose latitude is 33.45 and longitude is −94.35. The PV power output data is 5-min interval data. Thus, this study transforms the 5-min interval data into the 30-min interval data by cumulating corresponding values. The data from NREL is one-year data (i.e. 2006). The number of data points is 17,516.

The RTC dataset is from the U.S. Department of Energy's Regional Test Center Baseline PV systems [20, 21]. The RTC dataset is freely available from the CWRU-SDLE Research Center Open Science Framework page [22]. The RTC Baseline systems consist of eight PV systems with 1-min interval power output. The eight PV systems are located at four different sites in the U.S.: Florida, New Mexico, Nevada, and Vermont. Each site has a weather station providing 1-min interval weather conditions including GHI, wind speed, and temperature. The eight systems are nearly identical except for their locations, and there are two replicated systems at each location. This study uses the data of four sites (without replicated systems). The number of data points is from 1,281,387 to 1,789,155.

This study applies a low irradiance cut-off at 100 W/m^2 for both datasets. For the purpose of comparative study, this study manipulates the PV power output of the test data by applying a performance degradation ratio from 0 to 0.2 at intervals of 0.005 (i.e. 0.5%).

5.2 | Effects of representative weather conditions

This study applies the three representative weather conditions (i.e. STC, NOCT, and MEAN) to the dual model approach. Therefore, this subsection compares the effects of representative weather conditions on the accuracy of performance degradation calculation. In addition to the three representative weather conditions, this study also considers one additional case (WSUM) where the performance degradation is calculated as the average of the calculated performances of STC, NOCT, and MEAN. In the case of LGBM, this study uses all available weather conditions. Here, this study only considers XbX, PVUSA, and LGBM models. This study does not consider the 6K model, because it requires a nameplate power input, which is unavailable in the public dataset. This study also ignores PVWatts here because it does not follow the representative weather condition-based approach.

Table 1 shows the best- and worst-performing representative weather conditions of the models. To determine the best and the worst weather conditions for models, this study simply counts the number of datasets of seven NREL datasets and four RTC datasets showing the best and worst performances. The best and worst-performing weather conditions are determined to show the best and worst performance on the most data. The numbers in parentheses refer to the number of datasets that a corresponding model performs bests or worst with. XbX shows the best performance with MEAN for both NREL and RTC datasets. LGBM shows the best performance with WSUM for both NREL and RTC datasets. On the contrary, PVUSA shows the best performance with NOCT for the NREL dataset and with STC for the RTC dataset. This result first shows that different models are likely to require different representative weather conditions to show promising results. Another observation is that traditional representative weather conditions such as STC and NOCT may not be a good choice to have accurate performance degradation. In particular, XbX and PVUSA show the worst performance with STC.

Figure 7 shows some exemplary results of XbX, PVUSA, and LBGM. In Figure 7, *absolute error* indicates the absolute difference between the actual manipulated ratio and the estimated performance degradation ratio by models. *Relative error* indicates estimation error in the form of mean absolute percentage error (MAPE) as follows:

$$\frac{|\text{Manipulated degradation - est. degradation}|}{\text{Manipulated degradation}} * 100\%. \quad (12)$$

One common observation is that the absolute error and the relative error decrease as the manipulated degradation ratio increases. In particular, the relative error drops quickly. This first means that it is not easy to estimate accurate performance degradation when actual degradation is marginal. This is due to the fact that the marginal degradation is likely to reside on an error bound of a model built over the base data. The result also means that performance degradation accuracy increases sharply as the actual degradation increases. For example, in the case of

TABLE 1 Performance comparison of models with different representative weather conditions

Data	NREL [7 datasets]			RTC [4 datasets]			ALL [11 datasets]		
Models	XbX	PVUSA	LGBM	XbX	PVUSA	LGBM	XbX	PVUSA	LGBM
Best	MEAN(3)	NOCT(5)	WSUM(4)	MEAN(2)	STC(2)	WSUM(2)	MEAN(5)	NOCT(5)	WSUM(6)
Worst	STC(3)	STC(3)	MEAN(5)	STC(2)	MEAN(2)	NOCT(2)	STC(5)	STC(4)	MEAN(5)

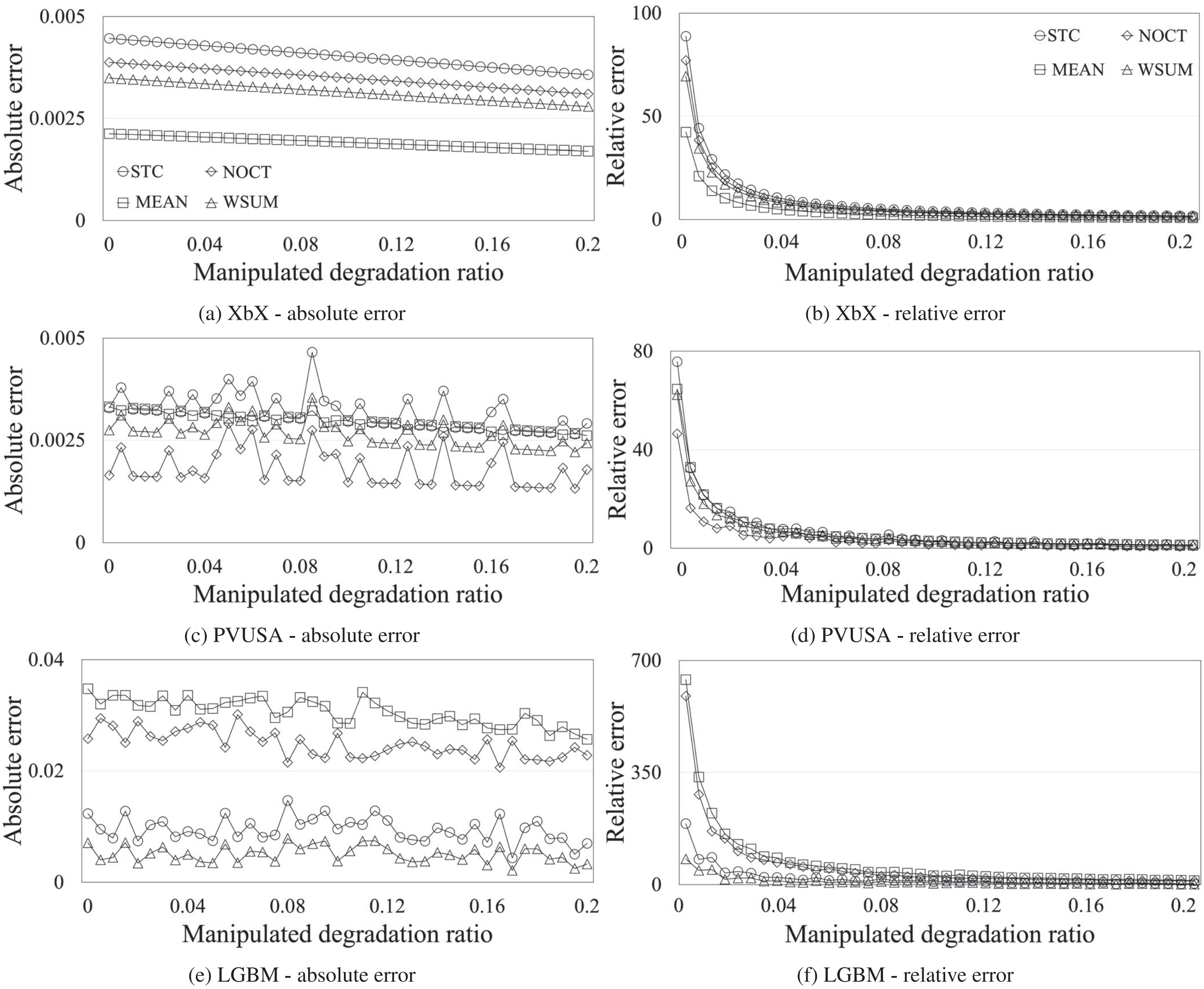


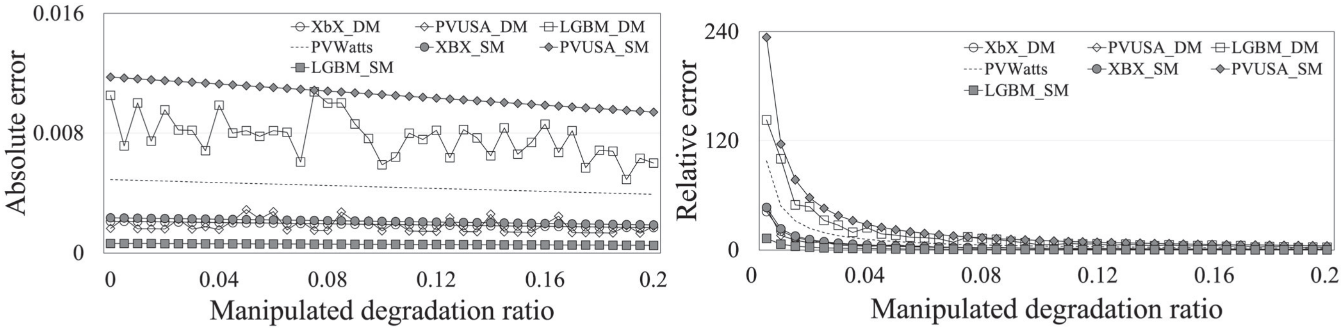
FIGURE 7 Exemplary results of XbX, PVUSA, and LGBM in dual model approach

XbX, the relative error of STC, NOCT, MEAN, and WSUM become less than 10% (5%) when the manipulated degradation ratio is 0.04, 0.035, 0.02, and 0.035 (0.085, 0.075, 0.045, and 0.07), respectively. Another interesting observation is that XbX shows a non-increasing function of the manipulated degradation ratio, which can be considered stable. On the contrary, PVUSA and LGBM show some fluctuations even though they show decreasing patterns.

In the following, this study uses XbX with MEAN, PVUSA with NOCT, and LGBM with WSUM for the dual models approach because those combinations show the best performance.

5.3 | Comparison of predicted power models

This subsection compares all candidate models. For the sake of simplicity, the terms XbX_DM, PVUSA_DM, and LGBM_DM are used to indicate XbX, PVUSA, and LGBM of the dual model approach. The terms XbX_SM, PVUSA_SM, and



(a) Absolute error (b) Relative error

FIGURE 8 Performance comparison of models in terms of accuracy (an exemplary case)

TABLE 2 Performance comparison of models (with all datasets)

	Accuracy	Uncertainty
Best	LGBM_SM(5)	LGBM_SM(11)
	XbX_DM(3)	
	XbX_SM(2)	
	PVUSA_DM(1)	
Worst	PVUSA_SM(9)	LGBM_DM(9)
	XbX_SM(2)	PVUSA_DM(2)

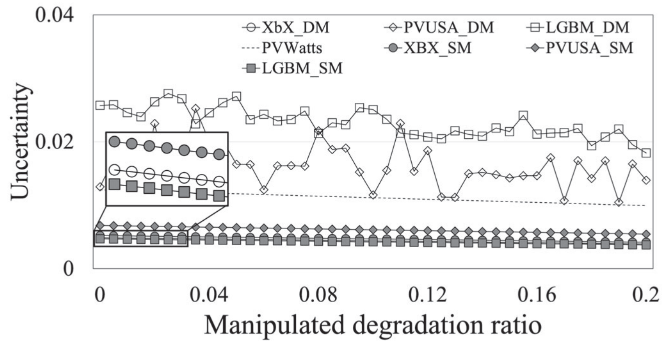


FIGURE 9 Performance comparison of models in terms of uncertainty (an exemplary case)

LGBM_SM are also used to indicate XbX, PVUSA, and LGBM of the single model approach.

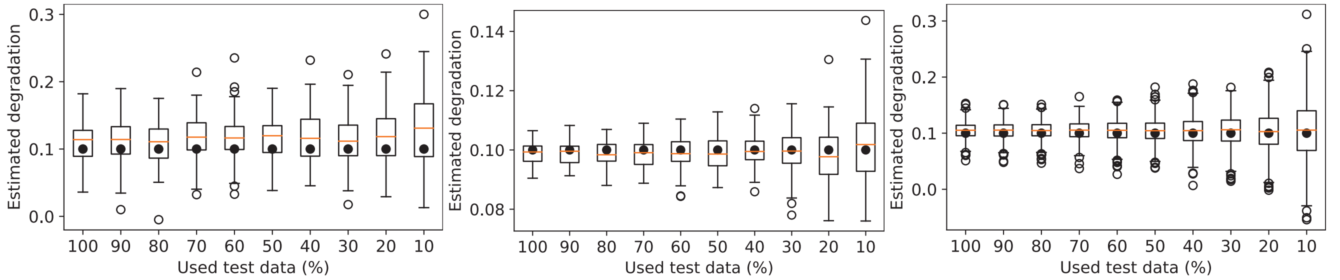
The models are first compared in terms of accuracy. Table 2 shows the results. The numbers in parentheses refer to the number of datasets that a corresponding model performs best or worst with. LGBM_SM shows the best performance in five datasets. It also shows competitive performance even when it does not show the best performance. On the contrary, PVUSA_SM shows the worst performance in most datasets. Figure 8 shows an exemplary comparison of models in terms of accuracy. PVUSA_DM and LGBM_DM show fluctuating results with respect to the manipulated degradation ratio. On the contrary, PVUSA_SM and LGBM_SM show stable performance. XbX and PVWatts also show stable performance. XbX_DM and XbX_SM do not show noticeable differences. On the contrary, PVUSA and LGBM show a noticeable difference in the dual model and the single model approaches. PVUSA and LGBM show better performance in the dual model and the single model approach, respectively. PVWatts does not show good performance. LGBM_SM shows the lowest absolute error. LGBM_SM shows a relative error of less than 5% from the manipulated degradation ratio of 0.015 (i.e. 1.5%).

The models are also compared in terms of uncertainty. Table 2 shows the results. LGBM_SM shows the lowest uncertainty in all datasets. XbX_SM and XbX_DM also show competitive performance. On the contrary, LGBM_DM shows the worst performance in most datasets. Figure 9 shows an exemplary comparison of models in terms of uncertainty. Like the case of accuracy, PVUSA_DM and LGBM_DM show fluctuating results with respect to the manipulated degradation ratio. The other models show stable performance.

In summary, LGBM_SM, XbX_DM, and XbX_SM show good and stable performance (i.e. high accuracy and low uncertainty). Considering that they are data-driven approaches, they are good choices for generating reliable performance time series as long as enough data used to train a model is available.

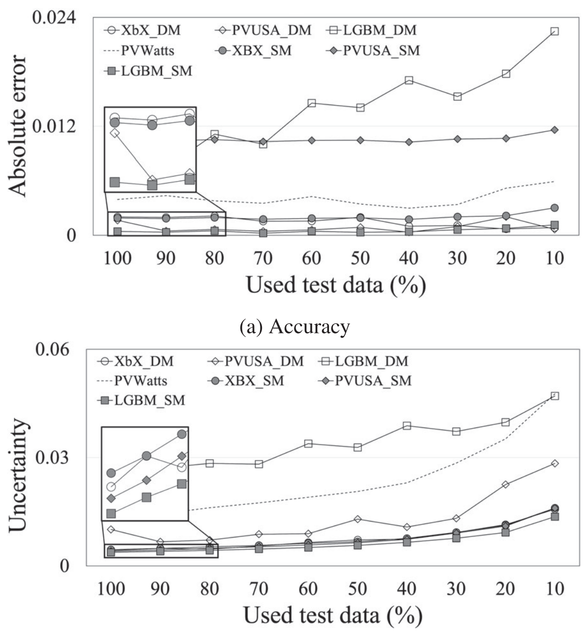
5.4 | Effects of test data size

This subsection examines the effect of the test data size. The intention of this examination is to see how much data is required to estimate reliable performance degradation. For this, the test data is manipulated to have 10% degraded PV power output. The number of test data used for the experiment is changed (i.e. from 100% to 10%). In the case of the NREL dataset, 10% means 875 data points over around 18 days. For a given percentage, the test data is randomly chosen. The bootstrap iteration is also applied for this experiment. Figure 11 shows the effects of the test data size on the accuracy and uncertainty. One common observation is that the accuracy does not change as the amount of test data decreases, except for LGBM_DM. LGBM_SM shows the best performance. On the contrary, the uncertainty increases as the number of test data decrease in all models. Please note that the uncertainty is the standard deviation of 1,000 predictions with the bootstrap iteration. Therefore, con-



(a) LGBM_DM (b) LGBM_SM (c) PVWatts

FIGURE 10 Boxplots of 1,000 predictions through the bootstrap iteration



(a) Accuracy

(b) Uncertainty

FIGURE 11 The effects of the test data size (an exemplary case)

sidering that one dataset to which the bootstrap iteration is applied has the same distribution, it is natural that the more samples are extracted, the less deviation is achieved. To examine this from another perspective, we depict the boxplots of three representative cases in Figure 10. In Figure 10, the dark circles indicate the true degradation. As shown in Figure 10, as the number of test data decreases, the length between the 25th percentile and the 75th percentile and between the lower fence and the upper fence increases while the median value does not change. This result first shows that a small number of data points may be enough to estimate accurate performance degradation. However, the more data is given, the more reliability can be achieved.

6 | CONCLUSION

Toward a reliable PLR, this study proposes a systematic method to assess the predicted power models in terms of both accuracy and uncertainty, which is missing in existing works. Following the proposed method, this study conducts extensive experiments using real-world datasets to assess the well-known predicted power models including XbX, PVUSA, PVWatts, and LGBM. The observations from the experiments may help researchers select appropriate models for their purpose. Considering that the characteristics of data vary depending on the area where PV systems are installed, determining a proper model requires additional analysis. In other words, given the data to which the predicted power model is to be applied, a closer study of the data itself and the variables that can be applied to the model is needed to determine which model will work best. For this, as future work, the proposed systematic method can be used to conduct such a study.

AUTHOR CONTRIBUTIONS

HyunYong Lee: Data curation, formal analysis, investigation, methodology, writing - original draft. Jun-Gi Lee: Resources, software. Nac-Woo Kim: Methodology, writing - review and editing. Byung-Tak Lee: Funding acquisition, project administration, writing - review and editing.

ACKNOWLEDGEMENTS

This work was supported by Electronics and Telecommunications Research Institute (ETRI). Grant was funded by the Korean government. [22ZK1100, Honam region regional industry-based ICT convergence technology advancement support project]

CONFLICT OF INTEREST

The authors have declared no conflict of interest.

DATA AVAILABILITY STATEMENT

Data openly available in a public repository that does not issue DOIs. The data that support the findings of this study are openly available in National Solar Radiation DataBase (NSRDB) Data Viewer at https://maps.nrel.gov/nsrdb-viewer/ , reference number [12], NREL Solar Power Data for Integration Studies at https://www.nrel.gov/grid/solar-power-data.html , reference number [13], and US DOE-RTC-Baseline dataset at https://osf.io/wn35j/ , reference number [16].

ORCID

HyunYong Lee https://orcid.org/0000-0002-0615-4241

REFERENCES

1. Share of electricity production from solar, 2021, https://ourworldindata.org/grapher/share-electricity-solar/.
2. King, D.L., Kratochvil, J.A., Boyson, W.E.: Field experience with a new performance characterization procedure for photovoltaic arrays. In: Proceedings of Second World Conference and Exhibition on Photovoltaic Solar Energy Conversion, pp. 1947–1952. European Commission, Luxembourg (1998)
3. Dobos, A.P.: PVWatts Version 5 Manual. Technical Report (TP-6A20-62641), National Renewable Energy Lab (NREL) (2014)
4. Curran, A.J., Zhang, R., Hu, Y., Haddadian, R., Braid, J.L., Peshek, T.J., French, R.H.: Determining the power rate of change of 353 solar plant inverters using a month-by-month analysis and common data science applications to power time series. In: Proceedings of IEEE Photovoltaic Specialist Conference, pp. 1927–1932. IEEE, Piscataway (2017)
5. Huld, T., Friesen, G., Skoczek, A., Kenny, R.P., Sample, T., Field, M., Dunlop, E.D.: A power-rating model for crystalline silicon pv modules. Sol. Energy Mater. Sol. 95(12), 3359–3369 (2011)
6. Hasselbrink, E., Anderson, M., Defreitas, Z., Mikofski, M., Shen, Y.-C., Caldwell, S., Terao, A., Kavulak, D., Campeau, Z., DeGraaff, D.: Validation of the PVLife model using 3 million module-years of live site data. In: Proceedings of IEEE Photovoltaic Specialists Conference. IEEE, Piscataway (2013)
7. Solis-Aleman, E.M., Casaa, J., Romero-Fiances, I., Silva, J.P., Nofuentes, G.: A study on the degradation rates and the linearity of the performance T decline of various thin film PV technologies. Solar Energy 188, 813–824 (2019)
8. Curran, A.J., Jones, C.B., Lindig, S., Stein, J., Moser, D., French, R.H.: Performance loss rate consistency and uncertainty across multiple methods and filtering criteria. In: Proceedings of IEEE Photovoltaic Specialists Conference (PVSC). IEEE, Piscataway (2020)
9. Kaayaa, I., Vasquezc, J.A., Weiss, K.-A., Topic, M.: Assessment of uncertainties and variations in PV modules degradation rates and lifetime predictions using physical models. Solar Energy 218, 354–367 (2021)
10. Ingenhoven, P., Belluardo, G., Moser, D.: Comparison of statistical and deterministic smoothing methods to reduce the uncertainty of performance loss rate estimates. IEEE J. Photovoltaics 8, 224–232 (2018)
11. Fiances, I.R., Livera, A., Theristis, M., Makrides, G., Stein, J.S., Nofuentes, G., Casa, J., Georghiou, G.E.: Impact of duration and missing data on the long-term photovoltaic degradation rate estimation. Renew. Energy 181, 738–748 (2022)
12. Herteleer, B., Huyck, B., Catthoor, F., Driesen, J., Cappelle, J.: Normalized efficiency of photovoltaic systems: Going beyond the performance ratio. Solar Energy 157, 408–418 (2017)
13. Subramaniyan, A.B., Pan, R., Kuitche, J., TamizhMani, G.: Quantification of environmental effects on pv module degradation: A physics-based data-driven modeling method. IEEE J. Photovoltaics 8, 1289-1296 (2018)
14. Jordan, D.C., Deline, C., Kurtz, S.R., Kimball, G.M., Anderson, M.: Robust pv degradation methodology and application. IEEE J. Photovoltaics 8, 525–531 (2018)
15. Meyers, B., Deceglie, M., Deline, C., Jordan, D.: Signal processing on pv time-series data: Robust degradation analysis without physical models. IEEE J. Photovoltaics 10, 546–553 (2020)
16. Lindig, S., Louwen, A., Moser, D., Topic, M.: New pv performance loss methodology applying a self-regulated multistep algorithm. IEEE J. Photovoltaics 11(4), 1087–1096 (2021)
17. Li, S., Liu, W., Li, J., Sun, S., Wu, Z., Xu, B.: A method for accurately assessing field performance degradation of PV modules in different geographical regions. Sustain. Energy Technol. Assess. 48, 101638 (2021)
18. NSRDB Data Viewer, https://maps.nrel.gov/nsrdb-viewer/
19. NREL Solar Power Data for Integration Studies, https://www.nrel.gov/grid/solar-power-data.html/
20. Stein, J.: US DOE Regional Test Centers Program - 2016 Annual Report. Technical Report SAND-2017- 4853R, Sandia National Lab. (SNL-NM), Albuquerque, NM (2017)
21. Lave, M.S., Stein, J., Burnham, L.: Performance comparison of stion CIGS modules to baseline monocrystalline modules at the new mexico florida and vermont regional test centers: January 2015- December 2016. Technical Report SAND2017-0957, Sandia National Lab. (SNL-CA), Livermore, CA (2017)
22. Curran, A.J., Jones, C.B., Braid, J.L., Stein, J.S., French, R.H.: US DOE RTC Baseline Systems. Case Western Reserve University, Cleveland, OH (2019)
23. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.Y.: LightGBM: A highly efficient gradient boosting decision tree. In: Proceedings of Advances in Neural Information Processing Systems (NIPS). MIT Press, Cambridge (2017)
24. Ogaard, M.B., Riise, H.N., Haug, H., Sartoria, S., Selj, J.H.: Photovoltaic system monitoring for high latitude locations. Solar Energy 207, 1045–1054 (2020)
25. Phinikarides, A., Kindyni, N., Makrides, G., Georghiou, G.E.: Review of photovoltaic degradation rate methodologies. Renew. Sustain. Energy Rev. 40, 143–152 (2014)
26. French, R.H., Bruckman, L.S., Mouser, D., et al.: Assessment of performance loss rate of pv power systems. Technical Report IEA-PVPS T13-22:2021, International Energy Agency (IEA) Photovoltaic Power Systems Programme (PVPS) (2021)
27. Jordan, D., Wohlgemuth, J., Kurtz, S.: Technology and climate trends in pv module degradation. In: Proceedings of the 27th European Photovoltaic Solar Energy Conference and Exhibition. WIP-Renewable Energies, Munchen, Germany (2012)
28. Phinikarides, A., Makrides, G., Georghiou, G.E.: Initial performance degradation of an a-Si/a-Si tandem pv array. In: Proceedings of the 27th European Photovoltaic Solar Energy Conference and Exhibition. WIP-Renewable Energies, Munchen, Germany (2012)
29. Jordan, D., Kurtz, S.: The dark horse of evaluating long-term field performance - Data filtering. IEEE J. Photovoltaics 4, 317–23 (2013)
30. Ransome, S.: The present status of kWh/kWp measurements and modeling. In: Proceedings of the 25th European Photovoltaic Solar Energy Conference and Exhibition. WIP-Renewable Energies, Munchen, Germany (2010)

How to cite this article: Lee, H.Y., Lee, J., Kim, N.-W., Lee, B.-T.: Comparative analysis of photovoltaic performance metrics for reliable performance loss rate. IET Renew. Power Gener. 1–12 (2022). https://doi.org/10.1049/rpg2.12651