Case Western Reserve University
Scholarly Commons @ Case Western Reserve University

Faculty Scholarship

3-25-2021

International Collaboration Framework for the Calculation of Performance Loss Rates: Data Quality, Benchmarks, and Trends (Towards a Uniform Methodology)

Alan J. Curran
Case Western Reserve University

Kunal Rath
Case Western Reserve University

Arash Khalilnejad
Case Western Reserve University

Roger H. French
Case Western Reserve University, roger.french@case.edu

Author(s) ORCID Identifier:

Arash Khalilnejad

Roger H. French

Follow this and additional works at: https://commons.case.edu/facultyworks

Recommended Citation

Curran, Alan J.; Rath, Kunal; Khalilnejad, Arash; and French, Roger H., "International Collaboration Framework for the Calculation of Performance Loss Rates: Data Quality, Benchmarks, and Trends (Towards a Uniform Methodology)" (2021). *Faculty Scholarship*. 64.
https://commons.case.edu/facultyworks/64

This Article is brought to you for free and open access by Scholarly Commons @ Case Western Reserve University. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Scholarly Commons @ Case Western Reserve University. For more information, please contact digitalcommons@case.edu.

Check for updates

DOI: 10.1002/pip.3397

RESEARCH ARTICLE

PROGRESS IN PHOTOVOLTAICS WILEY

International collaboration framework for the calculation of performance loss rates: Data quality, benchmarks, and trends (towards a uniform methodology)

Sascha Lindig[1,2] | David Moser[1] | Alan J. Curran[3] | Kunal Rath[3] | Arash Khalilnejad[3] | Roger H. French[3] | Magnus Herz[4] | Björn Müller[5] | George Makrides[6] | George Georghiou[6] | Andreas Livera[6] | Mauricio Richter[7] | Julián Ascencio-Vásquez[2,7] | Mike van Iseghem[8] | Mohammed Meftah[8] | Dirk Jordan[9] | Chris Deline[9] | Wilfried van Sark[10] | Joshua S. Stein[11] | Marios Theristis[11] | Bennet Meyers[12] | Franz Baumgartner[13] | Wei Luo[14]

[1]Institute for Renewable Energy, EURAC Research, Viale Druso 1, Bolzano, 39100, Italy

[2]Faculty of Engineering, University of Ljubljana, Trzaska cesta 25, Ljubljana, 1000, Slovenia

[3]SDLE Research Center, Materials Science, Case Western Reserve University, White 536, 10900 Euclid Ave., Cleveland, OH, 44106, USA

[4]TÜV Rheinland Energy, TÜV Rheinland, Am Grauen Stein, Cologne, 51105, Germany

[5]PV Power Plants, Fraunhofer ISE, Fraunhofer Institute for Solar Energy Systems, Heidenhofstrasse 2, Freiburg, 79110, Germany

[6]FOSS Research Centre for Sustainable Energy, PV Technology Laboratory, Department of Electrical and Computer Engineering, University of Cyprus, 75 Kallipoleos Avenue, P.O. Box 20537, Nicosia, 1678, Cyprus

[7]iLab, 3E sa, Quai à la Chaux 6, Brussels, 1000, Belgium

[8]EDF R&D, EDF Lab Les Renardières, Moret-Loing-et-Orvanne, France

[9]National Center for Photovoltaics, National Renewable Energy Laboratory, 15013 Denver West Parkway, MS 3411, Golden, CO, 80401, USA

[10]Copernicus Institute of Sustainable Development, Utrecht University, Heidelberglaan 2, Utrecht, 3584 CS, the Netherlands

[11]Photovoltaic Systems Evaluation Laboratory, Sandia National Laboratories, PO Box 5800 MS 0951, Albuquerque, NM, 87185 -0951, USA

[12]Department of Electrical Engineering, Stanford University, 350 Serra Mall, Stanford, CA, 94305, USA

[13]Group Photovoltaic, ZHAW School of Engineering, Technikumstrasse 9, Winterthur, 8400, Switzerland

[14]Solar Energy Research Institute of Singapore, National University of Singapore, 119077, Singapore

Correspondence
Sascha Lindig, Institute for Renewable Energy, EURAC Research, Viale Druso 1, Bolzano 39100, Italy.
Email: sascha.lindig@eurac.edu

Funding information
European Union's Horizon 2020, Grant/Award Number: 721452-H2020-MSCA-ITN-2016; U. S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE), Grant/Award Number: 34366

Abstract

The IEA PVPS Task 13 group, experts who focus on photovoltaic performance, operation, and reliability from several leading R&D centers, universities, and industrial companies, is developing a framework for the calculation of performance loss rates of a large number of commercial and research photovoltaic (PV) power plants and their related weather data coming across various climatic zones. The general steps to calculate the performance loss rate are (i) input data cleaning and grading; (ii) data filtering; (iii) performance metric selection, corrections, and aggregation; and finally, (iv) application of a statistical modeling method to determine the

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2021 The Authors. Progress in Photovoltaics: Research and Applications published by John Wiley & Sons Ltd.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

performance loss rate value. In this study, several high-quality power and irradiance datasets have been shared, and the participants of the study were asked to calculate the performance loss rate of each individual system using their preferred methodologies. The data are used for benchmarking activities and to define capabilities and uncertainties of all the various methods. The combination of data filtering, metrics (performance ratio or power based), and statistical modeling methods are benchmarked in terms of (i) their deviation from the average value and (ii) their uncertainty, standard error, and confidence intervals. It was observed that careful data filtering is an essential foundation for reliable performance loss rate calculations. Furthermore, the selection of the calculation steps filter/metric/statistical method is highly dependent on one another, and the steps should not be assessed individually.

KEYWORDS
degradation rate, performance loss rate, PV system degradation, PV system performance

1 | INTRODUCTION

The calculation of the evolution of a photovoltaic (PV) system's performance is crucial to (i) evaluate if a system is operating within the boundaries of long-term yield assessments and warranties and (ii) provide more accurate values to be used in yield assessments not only in terms of absolute value but also in terms of uncertainty. In order to be able to judge a system's performance, the performance loss rate (*PLR*), which is provided in units of %/a, or %/year, must be calculated in an accurate and well-documented way and with its uncertainty reported. The *PLR* does not just represent the irreversible physical degradation of PV modules; it also measures performance-reducing events, which may be reversible or even preventable through good O&M practices. Data availability, accuracy, and temporal resolution have to be taken into account when choosing and carrying out the necessary steps to calculate *PLR* values. These include (i) input data cleaning; (ii) data filtering; (iii) performance metric selection, corrections and aggregation; and finally, (iv) application of a statistical modeling method to determine the *PLR*. The calculation of *PLR* in PV systems is nontrivial as the "true" value is unknown. Another important issue is that there is no clear agreement of what the *PLR* represents, for example, a partial loss in power output over the complete irradiance range or a verifiable loss at predetermined conditions such as Standard Test Conditions. In the course of this work, the "voted" *PLR* is simply the average value of all calculated values per system and considered to be close to the "true" unknown value.

Several statistical methods have been proposed and compared[1]; however, there is no industry-wide consensus, and thus, no standardized approach to *PLR* calculations. At the IEC level, there was an early effort to introduce a standard or a Technical Reference for the calculation of degradation with the IEC 61724-Part 4; however, the initial activity was discontinued.[2]

The IEA PVPS Task 13 group, which focuses on PV performance, operation, and reliability and consists of experts from several leading R&D centers, universities, and companies, is developing a framework for the calculation of *PLR* of a large number of commercial and research PV power systems located across the globe. The aim of this international collaboration is to elucidate the steps and algorithms for calculating *PLR*. Various data filtering and temporal aggregation approaches coupled with different algorithms and statistical models can be used for *PLR* calculation, each of which can impact the results in terms of absolute values, uncertainties, and confidence intervals. Here, we present *PLR* results on 19 real-world datasets and four datasets from a simulated, or "digital," PV power plant with a predefined degradation rate. The contributors were asked to apply their preferred metrics and statistical methods resulting in the use of 11 data filtering approaches, eight performance metrics, and nine statistical modeling methods. Two additional tools from this comparative study were the development of an A to D grading system based on data quality metrics of percentage of outliers, missing datapoints and data gaps, and an `R` package for *PLR* determination using multiple metrics and methods for PV time series data. The results were then compared to identify the *PLR* that is the most commonly reported *PLR* value for that system, and we use this averaged *PLR* as the most likely approximation for the unknown "true" *PLR* value. Our goal is not to determine a single highly performing filter + metric + correction + model approach, which is based on the results probably not existing. Instead, we want to provide tools to PV researchers and PV plant owners to assess the performance of all their systems, even ones with realistic, and partially severe, data quality issues as well as to enable a cross-comparison among systems. We also analyze uncertainties by comparing the standard deviation of the input datasets, and the standard errors and confidence intervals of the *PLR* results from multiple metrics and methods, to illuminate relative performance differences among these multistep *PLR* determination approaches.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2 | DESCRIPTION OF DATASETS FOR PLR BENCHMARK

Nineteen datasets were made available to the interlab benchmarking participants. An overview together with the most important information can be found in Table A1. The climate zone categorization used in Table A1 is based on the well-known Köppen–Geiger classification,[3-6] and we include the new KGPV classification which includes irradiance intensity (L–K) but has reduced weather distinctions (A–F). In the following, the PV systems are briefly introduced, and Figure 1 shows their geographical distribution. All datasets are publicly available at https://osf.io/vtr2s/.

2.1 | Data characteristics: Time interval, time length, and data types

Data from PV systems are not standardized and can show significant variation based on the source. There are several characteristics related to how the data were collected that are considered before the *PLR* analysis.

Data resolution: The time interval of collected data. Typical values range from 1 to 15 min but can vary between 1 s to one or more hours, depending on the hardware and user settings. Typically, high-resolution data are desired since they provide more detail on the PV performance and also improve the model fitting due to the larger amount of data.

Time length: The total operating time of a system. For understanding the long-term performance of systems, the obvious choice of data would be from systems that have been operating for long periods of time. The absolute minimum of available data is 2 years, as a linear *PLR* describes a yearly change in performance. For reliable PLR evaluation, at least 4 to 5 years of data should be available.[7]

Available variables: Power (*P*), irradiance (*G*), and temperature (*T*) are the foundation of most *PLR* analyses, but they can be measured in a number of different ways. Power can be recorded at the AC or DC side or represented as energy accumulation instead of a power reading. DC power is desired over AC power readings in order to remove inverter influences. Irradiance can be reported as global horizontal irradiance (GHI) or plane-of-array (POA) irradiance. Temperature generally refers to ambient or module temperature, typically measured by a thermocouple. Backsheet temperature measurements of single modules may not be representative of the whole array in given PV system, particularly in larger plants. Additional variables including wind speed and direction, current, voltage, rainfall, and air mass can all be used in certain types of *PLR* analysis, depending on availability. Lastly, metadata are also important, for example, module and inverter characteristics, location, scale, and orientation.

Collection quality: Events such as missing values or gaps, reading errors, or sensor drifting are all commonly observed in PV system data. This is why the proposed PV data quality grading is a useful measure to better understand why one system can be modeled easily, and yet another can have multiple approaches fail. Most *PLR* analyses are robust to a certain amount of such problems or can account

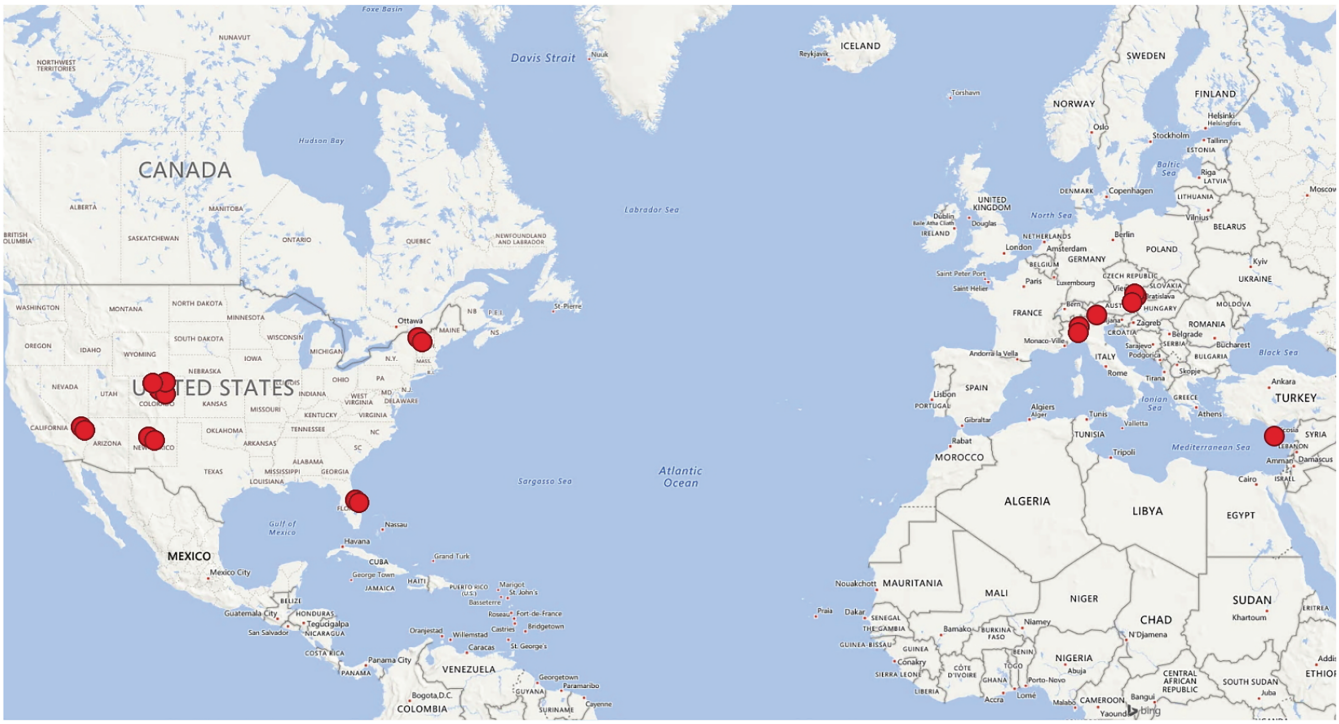


FIGURE 1 The locations of the PV systems used for the benchmarking exercise [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

for some of them. However, *PLR* cannot be calculated reliably in data sets with large proportions of anomalous or missing data or large time gaps in the dataset. Regular maintenance of sensors, site performance, and observation of data collection can reduce the impact of these issues.

2.2 | Datasets

2.2.1 | EURAC PV system

The EURAC PV system was installed at the airport of Bolzano/Italy (ABD) in 2010. The polycrystalline system has a nominal power of 4.2 kWp. The system is ground mounted with a fixed tilt of 30° and an orientation of 8.5° west of south. Additionally, a weather station is installed in close proximity to the test side. Various meteorological parameters are recorded such as POA irradiance, ambient temperature, and wind speed. On the rear side of the system, the module temperature is measured. The sensors are systematically cleaned and periodically calibrated in order to comply with the IEC 61724-1:2017 standard.[8] The weather data are recorded with a time interval of one minute. Since the electrical measurements are taken at time intervals of 15 min, all values are averaged to the same time interval. A period of 8 years is evaluated ranging from February 2011 until January 2019. It is important to mention that the time of observation is not equal to the system age; the system began operating in August 2010, which is roughly 6 months before the observation time starts. The delayed start of observation was set to exclude initial degradation effects.

2.2.2 | FOSS PV system

The FOSS PV system was installed at the outdoor test facility of the University of Cyprus (UCY) in Nicosia, Cyprus, and was commissioned in May 2006. The climate in Nicosia, Cyprus, is characterized as hot semiarid. The PV system dataset used in this investigation is obtained from a ground-mounted monocrystalline silicon (mono-c-Si) system that is rated at 1025 Wp, as depicted from the manufacturer's datasheet. Furthermore, the PV system is installed in an open-field mounting arrangement due South and at the optimum inclination angle of 27.5°.

The monitoring of this system started in June 2006, and both climatic data and operational measurements were acquired and stored. More specifically, the electrical performance of the system along with the prevailing irradiance and environmental conditions were recorded according to the requirements set by the IEC 61724-1 standard standard[8] and stored with the use of a robust measurement monitoring system. The monitoring system records POA irradiance (secondary standard pyranometer), wind, and temperature measurements. Periodic calibrations and inspections of the sensors were performed, in order to ensure high-quality data and reveal any deviations from the real measurements.

The PV system time series constructed for the purpose of this evaluation covers a period of 10 years starting from June 2006 until June 2016.

2.2.3 | RSE PV systems

RSE PV systems are based in the experimental area of Milan (north of Italy), where various PV technologies are analyzed. The data analyzed in this document refer to two PV power plants, respectively, with c-Si (polycrystalline silicon) and CdTE (cadmium telluride) technology, which started to operate in June 2009. The c-Si PV plant is a ground-mounted PV plant, south orientation, and tilt of 30°. The PV plant has a nominal power of 1.61 kW, constituted by a string of eight PV modules of 210 W. The CdTe PV plant is a ground-mounted PV plant, south orientation, and tilt of 30°. The PV plant has a nominal power of 1.16 kW and consists of four strings with four PV modules of 72.50 W each. A weather station is installed close to the test site and allows the acquisition of irradiance measurements (POA) and air temperature. Sensors are periodically cleaned and calibrated according to the IEC 61724-1:2017 standard.[8] Operational data are acquired every 10 s and sent to the remote unit which stores them as mean or integral values (1- and 15-min intervals).

2.2.4 | Pfaffstaetten PV systems

The Pfaffstaetten PV system is a 5-kWp rooftop system, running from January 2013 until April 2019. The system has three strings, two with second-hand polycrystalline modules (110 and 120 Wp, dating back to the end of the 1990s) and the third with CIGS modules (150 Wp). The inverter has three separate maximum power point (MPP) trackers, and these are connected to

- Pfaffstaetten A: 18 × pc-Si modules (initially measured power 1.812 kWp);
- Pfaffstaetten B: 18 × pc-Si modules (initially measured power 1.669 kWp); and
- Pfaffstaetten C: 15 × CIGS modules (rated at 2.250 kWp).

The monitoring data are inverter based with a 10-min timestep, but during morning start-up and closing down, additional measurement data with arbitrary timesteps are produced. If the inverter is idle during the night, no data are recorded.

Irradiance values are based on a c-Si reference cell, and the ambient temperature sensor is integrated in the case underneath. Therefore, the ambient temperature readings do not correspond to the air temperature (i.e., in the shade). As such, these temperature readings approximate the module temperatures obtained from the rear side of one of the Kyocera modules. There are no additional temperature measurements for the Miasole CIGS glass/glass modules, which may operate at slightly higher temperatures.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2.5 | US DOE RTC baseline systems

The US Department of Energy (DOE) Regional Test Center (RTC) Project has five sites.[9] At each site, there are two systems that are used as the "baseline systems."

This dataset is currently version 0.2 and consists of a series of eight PV systems in four locations/climate zones, 3.24-kW strings with the same module and inverter at each site. The data are 1-min time series inverter data with ground and satellite weather data.

These systems have proved to be useful for research purposes given that they are nearly identical at each climate zone, including the same number and brand of modules and inverters.

2.2.6 | NREL systems

Four PV systems from the US National Renewable Energy Laboratory (NREL) were included in the benchmark study, all located at the NREL main campus in Golden, Colorado. Short time interval 1-min data are collected for three systems, with 15-min data collected for the fourth and largest system. System #1 is similar to the RTC baseline systems described above using the same PV module type in one string of 10, total system size 2.7 kWdc, 3 kWac, with data available from April 2016 until July 2019.

System #2 is also a small research system using a string of five silicon heterojunction modules, 1 kWdc, 1.8 kWac, with data availability from August 2007 until December 2016. Both of these small research systems have colocated calibrated broadband pyranometer G_{POA} irradiance measurements, wind speed, and back-of-module temperature measurement and are mounted on free-standing open-rack structures.

Systems #3 and #4 are larger building-mounted systems with lower quality silicon photodiode G_{POA} irradiance measurement. A nearby weather station provides calibrated GHI, wind speed, and ambient temperature. System #3 is a 94-kWdc building-mounted system at 10° tilt and using multicrystalline-Si modules and a single 75-kVA central inverter. System #4 is a 524-kWdc carport using high-efficiency back-contact modules connected to two 250-kVA central inverters, also at ∼9° tilt angle.

2.2.7 | Digital power plant

The simulated "digital" power plant has been created with EDF R&D's tool "PV NOV." It consists of a string of 10 PV modules in series, with the following characteristics:

- From datasheet: P_{mpp} = 180 W (+/−3%), I_{sc} = 5.29 A, V_{oc} = 44.8 V
- From Flash simulated: P_{mpp} = 182 W, I_{sc} = 5.44 A, V_{oc} = 44.8 V

The string is connected to a 2kW inverter. The behavior of the plant is simulated with a Dymola/BuildSyspro Software, developed by EDF.[10,11] The model for the PV modules is a two-diode model.

Solar data were derived from HelioClim with a temporal resolution of 15 min. All in all, four PV systems have been created with the following settings:

- 1-year weather data repeated for 5 years with and without known degradation;
- 5-year weather dataset (real weather conditions given by Helioclim for a period of 4 years and an added fifth year which is the minimum value of each previous years) with and without known degradation.

The degradation of PV panels is simulated with a linear variation of parameters: short-circuit current (I_{sc}, initial value: 5.44 A, variation of −4%/a), series resistance per cell ($R_{s_{cell}}$, initial value: 0.00012 Ω.m^2, variation of +7%/a), shunt resistance per cell ($R_{sh_{cell}}$, initial value: 0.14745 Ω.m^2, variation of −6%/a), and the temperature coefficient of the I_{sc} ($\alpha_{I_{sc}}$, initial value: 0.037%/K, variation of −1.5%/a). Those values were chosen to induce an overall high degradation of the output power in a scenario that belongs to a bigger experiment including a wide range of degradation cases. The theoretical degradation for the simulated dataset with induced power loss was calculated from two viewpoints:

- @STC the P_{mpp} degradation is −4.41% per year.
- Absolute energy degradation is −4.89% per year, which has been quantified for the systems with repeating weather data.

The PV system files consist of timestamp, POA irradiance, ambient temperature, wind speed, AC power, and DC power.

3 | METHODS

3.1 | General approach

Each organization applied their preferred filters, metrics, and statistical methods for the calculation of the *PLR*. The necessary steps to calculate performance losses depend on the available data. A general approach for high-quality data treatment is presented below.

In this work, *PLR* have been calculated based on DC power readings if available; otherwise, AC power has been used. An overview of available measurement data for the individual systems can be found in Table A1.

Figure 2 presents the necessary calculation steps for retrieving a *PLR*. The steps include gathering, understanding, and grading of the input data; the application of certain filter; the selection and aggregation of a performance metric including possible corrections; and the application of models to calculate *PLR*.

First, we have to understand which data are available and the format conditions of our raw data. A quality assessment of measured data is always recommended and will ensure a smooth application of the steps to follow. To characterize the quality of time series datasets, we use a grading scheme that assess a dataset for outliers, missing

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

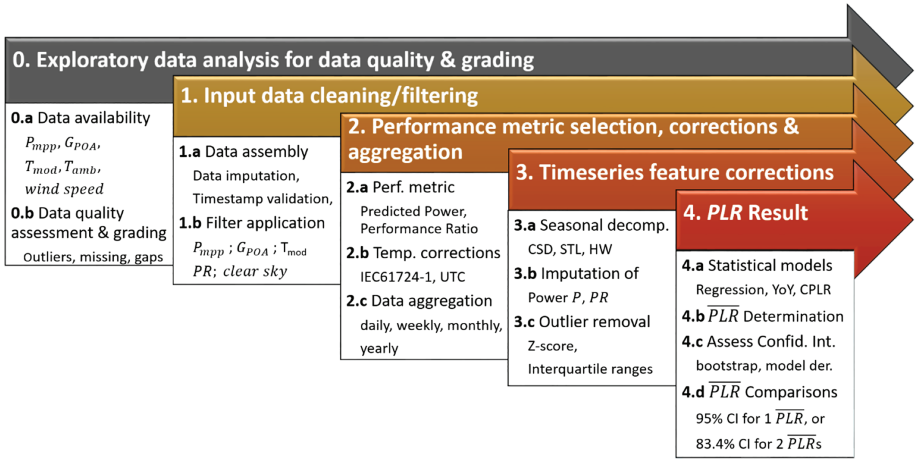


FIGURE 2 The four *PLR* analysis steps for high-quality time series power and weather data [Colour figure can be viewed at wileyonlinelibrary.com]

datapoints, and larger gaps in the data. Next, we apply filters to extract the essence of our data. This step is performed to get rid of outliers, measurement errors and nonrepresentative data. Usually irradiance, power, temperature, and performance ratio (*PR*) are considered. In cases where local weather data (irradiance and temperature) are not available, it is possible to use interpolated (weighted regression) data of different peered weather stations in relatively close proximity to the test side, satellite, or reanalysis-based datasets or clear-sky modeled data. Afterwards, a performance metric is selected. These metrics are usually not only power or *PR* but also empirically defined metrics. The next step involves data correction for temperature. It is not required but in most cases suggested. The correction attenuates seasonal variations of the chosen metric. Additionally, the data will be aggregated to a desired time interval, which is usually days, weeks, months, or years. In the next step, the performance metric is prepared for the final *PLR* determination through the application of seasonal decomposition, data imputation, and a final outlier removal. The last step involves the application of a statistical modeling method to determine the system's final *PLR* followed by an uncertainty assessment. Currently, there are numerous statistical *PLR* determination methods in the literature to choose from. A comparative study of methods found in the literature has been performed by Lindig et al.[1]

Two different definitions for the *PLR* are found in the literature. The relative *PLR* is calculated by

$$PLR[\%/a] = \left(\beta_1 \frac{t}{\beta_0}\right) 100, \tag{1}$$

and absolute *PLR* by

$$PLR[\%/a] = (\beta_1 t) 100. \tag{2}$$

β_1 is the slope and β_0 is the *y*-intercept of the corresponding linear model of the linear trend line for the *PLR* calculation. *t* is a scaling parameter that converts the time scale at which power or *PR* is observed to a yearly scale, as *PLR* is per year (12 for monthly, 52 for weekly, etc.). The absolute *PLR* (Equation 2) is independent of the initial starting value of the chosen metric. The absolute *PLR* gives an indication of the absolute loss rate, but it is important that the fitting parameter β_0 is also given.[12] The relative *PLR* (Equation 1) makes it easier to generalize the findings to the energy yield of the array using the initial yield of the plant. In the course of this work, the calculated *PLR* refers to the relative performance loss rate.

3.2 | Data imputation, filtering, and correction approaches

In case of missing data different strategies can be implemented. If only a small fraction of data is missing, imputation is not necessary and usually data aggregation solves the issue. If instead a larger share of data is missing, data imputation is the recommended approach, although many different imputation techniques are existing. A recent study by Livera et al.[13] proposes a unified methodology for data processing, quality verification, and reconstruction. It was shown that *PLR* studies are sensitive to invalid or missing data rate. If less than 10% of data are missing, the study recommends to use the list-wise deletion method, where simply data with invalid measurements are omitted. If more than 10% of the data are missing, data imputation techniques should be applied. In this study, the Sandia PV Array Performance Model[14,15] is recommended for missing power measurements, multiple imputation by predictive mean matching for missing irradiance measurements,[16] and the Sandia module temperature model[14] for temperature measurements. In another study by Lindig et al.[17] data imputation techniques for a considerable amount of missing POA irradiance measurements were compared, where other on-site measured climate data were available. Here, the histogram-based gradient boosting regressor performed with highest accuracy among several tested classical irradiance transposition and machine learning-based models.

Filtering serves to identify and remove data within the time series that are influenced by factors that cannot be modeled. The basic relationships between the output of a solar panel, incident irradiance, and the temperature are well understood; however, real-world applications cannot be well controlled and the performance of the plant may have external dependencies. Natural occurrences such as night,

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

shading/soiling/snow coverage or inconsistent irradiance, operational features such as inverter saturation and outages, or extreme conditions including high temperature and irradiance, can all influence the instantaneous power production of a system. These features are typically difficult to control, model, or quantify, and may not necessarily relate to the temporal performance of the system, so it is prudent to remove these data in any given analysis.

It is a common approach to remove such data; however, the extent of filtering is often an arbitrary process that varies by individual analysts, or must be tailored to individual systems in many cases.

3.2.1 | Irradiance threshold filtering

The irradiance threshold is one of the most standard filters applied to PV time series. Data with irradiance values that fall below or above given values are removed. Low cutoff values (filtering out irradiance data below a given value) are intended to remove nighttime and low irradiance periods. High irradiance thresholds remove outliers and potential errors in measurement. High cutoff values are typically set at 1200 W/m^2 based on typical maximum terrestrial irradiance readings, this generally concerns a small portion of the total data. The low irradiance cutoff, however, applies to a much larger portion of data. Low irradiance threshold values have varied significantly between research groups. Previously, data were subset to a high irradiance level, typically 800 W/m^2 and above to maintain conditions similar to STC. This has become less popular recently given the massive amount of data removal and current low irradiance cutoffs are generally around 100–200 W/m^2.

3.2.2 | Power threshold filtering

Power thresholding and irradiance thresholding have strong overlap with each other, given their fundamental link for PV systems. Removing low irradiance values will also remove low power values and vice versa. System outages are a common occurrence in commercial systems which can be easily removed with a low power filter, as power values will be low during these periods even when irradiance is high. High power cutoffs target outliers in the time series; power values that are unreasonably high. Power presents a unique problem since it is not uniform across systems due to the different technologies installed at different locations that are exposed under different environmental conditions. Power outputs of different systems can vary by many orders of magnitude, so threshold values have to be tailored to individual systems. A common method is to remove data based on a percentage of maximum power.

3.2.3 | Inverter saturation and curtailment

Inverter saturation occurs in a PV system when the power output produced by the modules is higher than the allowed AC power output of the inverter. At this point, the inverter will be "saturated" and the power output will be maintained at this maximum value and will not be able to increase, even if the module DC power increases. Curtailment is commonly used to stabilize the power output of PV plants and increase the capacity factor, making the systems easier to integrate into existing grids, but proactive curtailment can lead to reduced availability. As such, inverter saturation is most commonly observed in larger scale commercial PV systems. Saturation poses a unique problem in PV data analysis as it occurs at higher irradiances, when systems are assumed to perform under ideal operating conditions. Power values exceeding saturation limits are no longer a function of weather conditions and should not be used in modeling.

Saturated data can be removed quickly if the saturation limit is known by filtering out power above 99% of the limit. Ninety-nine percent is commonly used, but other values can be applied if needed for different data sets. Unknown saturation limits can be identified by observing maximum power trends in the data, appearing as flat plateaus at the peaks of daily power trends.

3.2.4 | Clear-sky filters

Clear-sky filters attempt to subset data to periods of time with little to no cloud cover during operation. There are several different reasons why someone might want to perform this filter step. Clear-sky filters may be used to reduce the influence of inconsistent shading on a system. Large systems in particular may experience variation in irradiance between different strings or the pyranometer under periods of partial or intermittent cloud cover, leading to a discrepancy between power produced and irradiance measured. Additionally, clear-sky filters are often used to merge in modeled irradiance values for a system, which do not perform well in cloudy periods. Comparing sensor and modeled irradiance during clear-sky periods is a common method for detecting sensor drift.

There are two well-used methods of identifying clear-sky periods in a system, the 5-factor moving average by Reno and Hansen[18] available in PVLib[19] and a clear-sky index (CSI), used by NREL in RdTools.[20] The first uses a comparison between modeled and sensor irradiance with a moving average evaluating which periods show strong similarity. Periods where sensor and modeled irradiance show strong overlap are noted as clear-sky periods. The CSI is a less strict method and simpler to apply. It is also used as a comparison between sensor and clear-sky modeled irradiance but identifies clear periods using a ratio between the two, defaulting to 85%. Any period where the sensor irradiance is within 15% of the modeled irradiance is flagged as clear sky or near clear sky. The five-criteria method is stricter causing the removal of larger amounts of data, which ends up removing large amounts of data, and is generally not used in a direct *PLR* analysis. The CSI method keeps more data and is incorporated into the standard RdTools *PLR* analysis pipeline.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.5 | Shading, soiling, and snow corrections

Shading, soiling, and snow coverage may refer to events that inhibit light reaching the surface of the modules, while not being represented in the local irradiance if the sensors are cleaned periodically or freed from snow. This effect is observed as a drop in power without a corresponding drop in irradiance. Identifying these periods can be tricky as their influence on the power output can vary greatly from minor affects to large-scale loss, making them difficult to detect. System logs can identify snow events or dust build up, however these may not be available or accurate for all systems. Automated soiling removal is usually done with outlier detection. When converting power measurements to PRs, soiling events will produce lower PR values than regular operating periods and can be filtered out.[21] Other methods can also be applied which use power and irradiance trends and clustering to detect and remove data influenced by soiling.[22] Shading, soiling, and snow can vary greatly between systems, and it is recommended that PV analysts should view power corrected time series (PR, weather regression, etc.) of their systems to identify any potential areas of concern.

3.2.6 | Performance metric IQR filters

Power *P* and performance ratio *PR* are the most common performance metrics used. *PR* is a unit-less parameter, which describes the relationship between incoming irradiation and produced power by a PV system. Since power and irradiance follow a nearly linear trend over a wide range of irradiance, this relationship can be used to detect and remove nonrealistic power–irradiance pairs created through sensor shadowing, alignment, or other issues. Usually, statistical thresholds based on interquartile ranges (IQRs) around the median or mode[23] of the performance metric values are used to filter irradiance and power data.

TABLE 1 Chosen statistical model, metric, filter and aggregation steps

			Filter				
Nr	Models	Metric	Irradiance (W/m^2)	Module temperature (°C)	Power	Performance ratio	Aggregation
1	STL1, YoY1	PR_{Tcorr}	500–1200	−40 to 100	(0.01–1.2) · P_{nom}	±2sd around monthly *PR* mode	Monthly
2	STL3, STL4, STL5, STL6 LS-LR4, LS-LR5, LS-LR6, LS-LR7, STLYoY1, YoY4	6K PVUSA Xbx Xbx(UTC)	>100		(0.01–1.2) · P_{nom}	1.5× interquartile range	Monthly
3	VAR1	Power	350–850			±2sd around instantaneous *PR*	Daily then yearly
4	R-LR1, LS-LR1	Power	800–1000	5k bin containing largest share of datapoints		±5% from yearly median *PR*	None
5	CSD1, LS-LR2	*PR*				0% < PR < 100%	Monthly
6	STL2, YoY3 CSD2, LS-LR3	PR_{Tcorr}	200–1200	−50 to 100	0.01-(98th percentile of 0.99 P_{ac})	*PR* > 0	Monthly
7	YbY1	Power	780–820	18–22			Yearly
	or						
	YbY2		980–1020	23–27			
8	LS-LR8, CSD3 STL7,[a] STL8[a] HW1, FBP1	*PR*	50–1300		(0.1–1.3) · P_{nom}	±3sd around monthly *PR* mode	Monthly
9	YoY2	PVWatts	200–1200	−50 to 110	*P* > 0		Daily
10	YoY5	*PR*	100–1000			1.5× interquartile range	Daily
11	SCSF1	Power	Strict clear-sky filter				Daily

[a]STL7 and STL8 follow the exact same approach including filtering and metric. The only difference is that STL7 uses the STL function of the `R` stats package[24] implemented with the rstl package[25] in `Python` and STL8 the function directly implemented in the statsmodel package[26] of `Python`.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2.7 | Data filters summary

In Table 1, the applied filters and the chosen aggregation steps are summarized together with the model names and the performance metric used.

3.3 | Performance metrics

A metric is a certain measure that provides information about the performance of a PV system in one way or another. In the following, the most commonly used metrics in PV are described.

3.3.1 | Power (*P*) metrics

This metric refers to the measured system power, filtered, and adapted depending on the selected statistical method for *PLR* determination. For instance, the power metric was subject to very strict irradiance filters and temperature binning for the R-LR1 and LS-LR1 models.

3.3.2 | *PR* models

The PR at the DC side is calculated by[8]

$$PR_{DC} = \frac{Y_a}{Y_{ref}} = \frac{E_{DC}/P_{nom}}{H_{POA}/G_{STC}}, \tag{3}$$

where Y_a is the array yield, and Y_r is the reference yield; E_{DC} is the DC energy produced over a certain time t; P_{nom} is the nominal power at STC; H_{POA} is the POA irradiation over a certain time t; and G_{STC} is the irradiance of 1000 W/m^2. We have decided to use DC values to eliminate losses due to DC/AC conversion.

The *PR* can be corrected for temperature[8] using temperature coefficients as provided by the manufacturers or obtained from the time series data to better reflect the actual outdoor performance of the module (PR + temperature correction [PR_{Tcorr}]). The advantage of correcting temperature based on power data over the PV power plant's lifetime is the large range of available temperature, increasing the certainty of the power versus temperature trend. The correction should be performed according to standard IEC 61724-1:2017.[8] Seasonal fluctuations are still evident even when temperature-corrected PR is used; this is due to other effects such as angle of incidence and spectrum. Furthermore, if the temperature coefficients are biased, a seasonality due to changing temperature ranges will be introduced.[27]

3.3.3 | Power predictive models

Generally, a power prediction model is built, for a specific time period such as a day, week, or month, to predict power as a function of weather, and then, standard or representative weather variable parameters are applied to each aggregated temporal period model. This produces a predicted power value, for the given weather conditions and time period, which is therefore independent of the actual measured weather values. Four power predictive models are used in this study to compare the effects of the subsequent time series they produce on *PLR*. The models are described in detail by Curran et al.[28]

XbX: The XbX model, is a data-driven, multiple regression predictive model[29] with a temperature and irradiance term. The flexibility of this model enables nonlinear, change-point *PLR* calculations and allows for either POA or GHI to be used in the irradiance term and air or module temperature in the temperature term (*T*).

$$P_{pred} = \beta_0 + \beta_1 G + \beta_2 T + \epsilon. \tag{4}$$

The X in the name refers to the given time aggregation period that the power prediction model is built to predict for; a model built on a day of data would be day-by-day (DbD), while in week-by-week (WbW) or month-by-month (MbM) modeling, data would be subset by weeks or months. The time aggregation period is chosen based on the condition of the data being modeled and what modeling will be performed on the overall data set.

XbX + UTC: When modeling on small time scales such as individual days, it can be difficult to properly model temperature given the low variation that typically occurs in that time. Days staggered by season (i.e., summer vs. winter) have very different ranges of temperature, so modeling temperature between them can lead to extrapolation. By introducing a universal temperature correction (UTC), one can produce a single temperature coefficient that can be used to convert to the desired representative temperature value.

$$P_{cor} = \frac{P_{obs}}{1 + \gamma_T (T_{obs} - T_{rep}) \left(\frac{G_{obs}}{G_{rep}}\right)}, \; P_{cor} = \beta_0 + \beta_1 G + \epsilon. \tag{5}$$

Here, data are subject to a high irradiance G_{rep} of 900 W/m^2, and the slope of the irradiance over temperature becomes γ_T. *Obs* represents observed or measured values, and T_{rep} is a representative temperature. This method is most similar to a temperature-corrected performance ratio method used in other *PLR* packages such as RdTools[20,21,30] but structured as a power predictive model for better comparison with other models.

PVUSA: The well-known PVUSA model[31] is physics based and described by

$$P = G_{POA}(\beta_0 + \beta_1 G_{POA} + \beta_2 T_{amb} + \beta_3 WS). \tag{6}$$

Here, T_{amb} is the ambient temperature (°C), and *WS* is the wind speed (m/s). The model assumes that current is a function of the in-plane irradiance G_{POA} and voltage is a function of both G_{POA} and the module temperature T_{mod}. T_{mod} is predicted using T_{amb} and *WS*.

6K: The 6K model[14] is the most complicated power predictive model used in this study. The name "6K" refers to the coefficients fit by the model.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\begin{aligned} G' &= G_{POA}/G_{STC}, \\ T' &= T_{mod} - T_{STC}, \\ P &= G'(P_{nom} + k_1 ln(G') + k_2 ln(G')^2 + k_3 T' \\ &\quad + k_4 T' ln(G') + k_5 T' ln(G')^2 + k_6 T'^2). \end{aligned} \tag{7}$$

This model uses in-plane irradiance (G_{POA}) and module temperature (T_{mod}) but models them as a fraction of standard irradiance (G_{STC}) and difference from standard temperature (T_{STC}). Additionally, this model requires the nameplate power P_{nom} as an input and will always predict P_{nom} at STC conditions.

PVWatts: This simple power predictive model follows the irradiance and temperature scaling approach of PVWatts[32] as implemented in the PVLib Python software package:[19,33]

$$P = \frac{G_{POA}}{1000 \cdot P_{nom}}(1 + \gamma_T(T_{mod} - 25°C)). \tag{8}$$

3.4 | Statistical methods

Finally, a statistical method is applied to compute the *PLR*, given in percentage per year (%/a). The methods applied in this paper are

Linear regression (LR): *PLR* is commonly assumed as linear, where a single *PLR* value is representative of the entire lifetime of a system. Alternately, nonlinear *PLR* methods[27,34] can be used to determine change in the trend of performance between different periods during the lifetime of the system.

Assumed linear *PLR* is determined by regression of the predicted metric versus time or through year-on-year (YoY) modeling. For regression determined *PLR*, the slope and intercept of the trend directly relate to the change in system performance. Both least squares linear regression (LS-LR) and robust regression (R-LR) have been used in this study. Least squares regression can be simple if only one dependent variable predictor, or it can be ordinary least squares (OLS) regression if there are multiple predictors, and the errors are homoscedastic and uncorrelated. If the errors are normally distributed; then, OLS regression provides maximum likelihood estimation, and the coefficients are the most probable.[35] Robust regression is another form of regression that is less sensitive to assumptions about the data-generating process and can be less affected by outliers, compared to OLS regression, while being more computationally demanding.[36]

Classical seasonal decomposition (CSD): CSD 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.[37] The step of the seasonal period depends on the data resolution and is usually set to 12 for monthly data. In this case, 6 months at the beginning and six at the end of the observation period are not included in the averaged time series. By removing the trend from the measured data and averaging months of consecutive years, the remainder corresponds to the residuals.[38]

Seasonal and trend decomposition using loess (STL): The idea behind STL is to decompose the *PR* time series into a seasonal component, a remainder or residual component and a trend using locally weighted, nonparametric regression.[37] The trend is a nonlinear curve,[39] and STL functions are available in `R` in both the base `R` stats package and the STL-Plus package.[24,40] Afterwards, a linear fit of the trend is performed to get a regression representation of the performance evolution of the PV system, of which the gradient is multiplied by a factor to present yearly values (12 for months, 365 for days, etc.) of the final *PLR*. This statistical method is suitable for time series with a seasonal behavior and where the data are of high quality.[1]

YoY: The YoY approach for *PLR* determination was first applied by Hasselbrink et al.[41] and is now available in the RdTools package[20] in `Python` and the PVplr package in `R`.[22] In YoY, the differences between one datapoint in a calendar year with the datapoint at the same position in the subsequent year are accumulated over a 1-year period. The median value of these multiple yearly *PLR* represents the overall system *PLR*. The *PLR* of the YoY method is normalized to the first year's median, though one can choose not to normalize. The confidence interval is calculated using a Monte Carlo or bootstrap resampling of the distribution.[20,21,35]

VAR method: The VAR method gives degradation rates from one year to the next. By averaging the annual degradation rates, we get the PL of a system. Regression models of power variations with respect to environmental variations (irradiance and ambient temperature) are fitted. The basic idea of the VAR method is to build a model of correlation between yearly variations of output power with respect to yearly variations of environment, hence the name: the VAR method.[42] After processing and filtering the data, it fits a regression $\Delta Power = f(\Delta Irr, \Delta T_{amb}) + d$, meaning that if f is accurate enough, d is the variation of power not due to environmental changes, but only due to the system condition itself, and then interpreted as a performance degradation. It gives degradation rates from 1 year to the next and by averaging the annual degradation rates we get the *PLR* of a system.

Year-by-year (YbY): A yearly aggregation of strictly filtered data is the basis for this method. Consequently, the first year of measurements is set as a base value to 0% and the yearly difference in produced power within the filtered frame is evaluated in the following years. The average of differences between yearly values in respect to Year 1 is the final *PLR*.

Statistical clear-sky fitting (SCSF): The SCSF method fits a constrained, nonparametric clear-sky model to the data.[43] This model is adaptive and can model sites with complex shade patters, as well as unobstructed fixed-tilt and tracking systems. The model is very robust to missing data and poor data quality and can be used for data imputation, clear-sky condition detection, and clear-sky adaptive forecasting. The algorithm compares data on subdaily, daily, seasonal, and yearly time scales to estimate daily and seasonal patterns. One of the constraints on the problem is a consistent year-over-year percent change in daily energy, which becomes the estimate of system degradation.[44] This approach is unique in that no other information or data is required besides measured power—no irradiance data, no

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

temperature data, no meteorological data, no system configuration information, and no metadata. Therefore, this method is suitable for the analysis of distributed rooftop PV systems and the more highly instrumented and well-modeled centralized PV power plants. In addition, irradiance sensors can themselves be treated as a PV power signal source, allowing the automated analysis of sensor drift.

Holt-Winters (HW): The HW seasonal model can be used to forecast and smooth performance time series of PV systems. It consists of three smoothing parameters, a level, slope, and seasonal component. Although the HW model can be used in an additive or multiplicative manner, the additive method should be used for PV time series because seasonal variations are expected to be fairly constant throughout the time of observation. A weighted average is used to compute the slope of the level, and the smoothing parameter determines how fast the exponential weights decline over the past observations.[37,45]

Facebook Prophet (FBP): This statistical method is a modular regression model with an additive structure. It consists of four parameter, namely trend, seasonality, holiday, and error. The holiday term is used in business applications and omitted for this study. Seasonality is considered for daily, weekly, and yearly recurring patterns. Since PV power time series are expected to show monthly seasonality, the built-in yearly seasonality option of the model is set to TRUE which takes into account monthly patterns. Time is used as a regressor, and the trend is fit using a piecewise linear and a saturating growth model.[46] FBP has the advantage of incorporating change-point analysis which is useful for computing nonlinear PLR. However, in order to calibrate this model to provide meaningful results for PV degradation behavior, the flexibility of the extracted trend, number of potential change points, and range had to be adjusted according to the process and settings reported by Theristis et al.[27,47]

4 | RESULTS AND DISCUSSION

This section is divided into six different parts. First, the steps of calculating a *PLR* are shown as a detailed example using the EURAC dataset. In Section 4.2, the statistical characteristics of the PV power time series datasets are characterized and presented. A discussion on the contributions to the uncertainty of *PLR* results follows, including measurement uncertainty and the standard deviation of the power time series. Afterwards, the impact of different data filters utilized in the benchmarking exercise are compared by applying them on two PV system datasets while holding the other calculation steps the same. In Section 4.5, the calculated *PLR* of the digital plants are investigated. Finally, the results based on the real PV power time series datasets are examined in Section 4.6 including a discussion on the impact of data quality with respect to the magnitude and uncertainty of *PLR*.

4.1 | Example calculation of a PV system PLR

Here, the necessary steps for calculating the *PLR* of a PV system are shown graphically based on the monitoring data of the EURAC system. The specifications of the corresponding system can be found in Section 2.2 and in Table A1.

In Figure 3, the graphical evolution of possible *PLR* calculation steps is shown. In this example, STL has been applied to the monthly

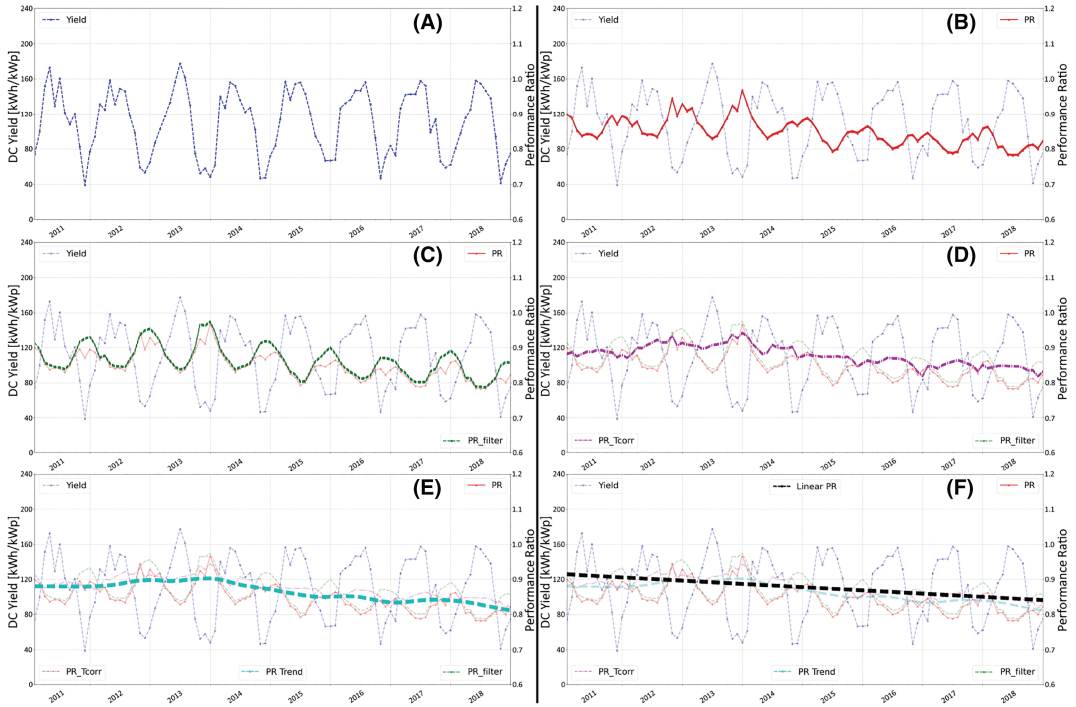


FIGURE 3 Example calculation steps for retrieving *PLR* value: (A) yield, (B) *PR*, (C) filtered *PR*, (D) temperature-corrected *PR*; (E) performance trend of temperature-corrected *PR*; and (F) assumed linear *PLR* [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

temperature-corrected *PR*. The performance trend, extracted with STL, is fit with a simple linear model to determine the asumed linear *PLR*.

First, the power time series or the yield (the power divided by the installed capacity, in kWh/kWp), is shown in Figure 3A as the monthly aggregated yield of the power plant throughout the time of observation. The system shows peak yield in the summer time and the lowest yield in the winter months. In Figure 3B, the selected metric (i.e., PR) is added to the plot. It is shown that due to the strong temperature dependence of PV modules, especially in crystalline, the PR exhibits high seasonality with low values during the warmer months and higher values during the colder periods. The application of a strict irradiance filter combined with a *PR* filter, to exclude values out of the range of two times the standard deviation of the monthly *PR* mode, yields the filtered *PR*, visible in Figure 3C. The applied filters correspond to Filter #1 of Table 1. It can be seen that the *PR* time series exhibit a sinusoidal shape, which is an indicator of the exclusion of nonrepresentative measurements or measurement conditions through filtering. Figure 3D shows the effect of temperature correction according to the standard IEC 617214-1:2017.[8] The correction was performed using measured module temperature values, and one can see a reduction in the apparent seasonality of the time series. The chosen statistical method for receiving the final linear *PLR* was a combination of STL[39] and LR. STL is a locally weighted regression, which extracts a nonlinear trend from a dataset by excluding the remaining seasonality and the residuals. This nonlinear trend can then be fitted with a simple linear model using regression to determine the best fit linear trend line, visible in Figure 3F. The yearly aggregated gradient of the linear function, divided by the intercept, is the final, assumed linear, relative PLR of our system (see Equation 1). The intercept of the function represents the *PR* value at the starting time of the time series. For this system, a relative *PLR* of −0.90%/a was calculated using the approach explained above. For the associated uncertainty of the *PLR* value, the residuals are added back into the STL trend component. The uncertainty between the STL combined component and the linearized trend is ±0.09%.

4.2 | EDA of the power time series datasets

To assess the meaning, accuracy, and robustness of the calculated results for a particular dataset, it is useful to determine during initial exploratory data analysis (EDA), the appropriate statistical measures of PV dataset quality. This provides insights into which datasets are robust to analysis, and which datasets may fail at particular steps or for certain types of analysis, such as the case where different aspects of data missingness makes analysis impossible.[48,49] In addition these measures can guide the user on the expected uncertainties in comparing multiple PV systems, which may have quite different equipment and operational histories, and therefore dataset quality. From time series analysis of building electricity time series data to perform virtual building energy audits, a time series dataset grading schema has been developed which has proven useful to alert users to expect high, or low, quality of results of data analysis of the systems.[50] We have adapted this approach for application to PV system time series datasets such as the power and irradiance time series and have implemented PV dataset quality grading in the PVplr package. We statistically characterize the power time series and then grade each dataset in three areas, outliers, missing datapoints, and data gaps. We have developed a grading schema, summarized in Table 2, and the measures and grades are summarized in Table 3.

TABLE 2 Data quality grading criteria

Letter grade	Outliers (%)	Missing percentage (%)	Longest gap (days)
A	Below 10	Below 10	Below 15
B	10 to 20	10 to 25	15 to 30
C	20 to 30	25 to 40	30 to 90
D	Above 30	Above 40	Above 90

Note: Outliers include the impact of clouds and anomalous datapoints; missing data is 5 or fewer sequential datapoints; and the longest gap is of all the data gaps in the dataset. The dataset length needs to be >2 years for a P grade; otherwise, it is graded F.

Outliers are typically defined as points that are greater than ±1.5 times the IQR and may be anomalous datapoints.[48] In time series, such as PV time series, outliers can arise from causes that fall beyond the expectations of a model, so, for example, for the power generated by a PV system, if there were no clouds, then a linear second-order model can fit the daily and seasonal changes in power production.[51] But clouds, being statistically random, are not easily modeled, and the power drops due to cloud shading would be outlier datapoints, while still physically meaningful. Anomalies are a subset of outliers, and correspond to datapoints that are not physically reasonable, but arise due to a mistaken measurement or malfunction of a piece of equipment.[52] For outlier detection we use the tsoutliers R package, which identifies time series outliers arising from clouds and anomalous datapoints such as arise from measurement errors.[53,54] Various of the filtering and correction methods discussed here are examples of approaches to address dataset outliers. Missing datapoints in a dataset is another typical data error that can impact analysis. A set of up to 5 sequential missing datapoints can be imputed rather easily using simple interpolation, so we consider the % missing datapoints the second important dataset characteristic to identify, and if desired, to impute. Longer data gaps in a time series dataset can arise from system or communications outages, can be quite problematic for different analysis methods, and are hard to correct or impute in an attempt to mitigate their impact, so we consider this the third important characteristic. These three categories of statistical properties give us a quantitative sense of the "missingness" of the dataset, and learning which filters, and methods are robust in the face of outliers, missing datapoints and data gaps, is important to advance the field. There is much active work on data imputation to address both outliers, missing datapoints and data gaps, but we have not implemented these here.[55]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Statistical characteristics of PV datasets used in the *PLR* benchmarking exercise

Dataset ID	Grade (outliers, missing, gaps, P/F)	Power variable (kW)	Length (years)	Outliers (%)	Missing (%)	Data gaps: # of gaps/longest gap (days)	
EURAC	BAAP	P_{DC}	7.95	11.5	2.1	2847	7.0
FOSS[a]	BCBP	P_{DC}	10.9	13.8	32.9	134	26
RSE CdTe	AABP	P_{AC}	9.59	10.0	0.3	2	25.3
RSE pc-Si	BABP	P_{AC}	9.59	11.0	0.3	2	25.3
Pfaffstaetten A[a]	ADAP	P_{DC}	6.33	2.0	41.2	2082	0.9
Pfaffstaetten B[a]	ADAP	P_{DC}	6.33	2.2	40.1	2014	0.9
Pfaffstaetten C[a]	ACAP	P_{DC}	6.33	2.2	39.1	2061	0.9
US DOE c10hov6	BAAP	P_{DC}	3.16	14.6	1.2	69	13.1
US DOE kobdpi8	BAAP	P_{DC}	3.44	13.2	0.4	15	5.2
US DOE luemkoy	AAAP	P_{DC}	2.45	10.0	0.5	16	3.7
US DOE lwcb907	BACP	P_{DC}	3.47	13.8	3.7	33	49
US DOE t3pg1sv	BACP	P_{DC}	3.47	12.2	3.7	33	49
US DOE wca0c5m	BAAP	P_{DC}	3.16	12.8	1.2	69	13.1
US DOE wxzsjaf	AAAP	P_{DC}	2.45	9.9	0.5	16	3.7
US DOE z0aygry	BAAP	P_{DC}	3.44	14.8	0.4	15	5.2
NREL1[a]	BACP	P_{DC}	3.31	13.0	6.3	727	76.7
NREL2[a]	BABP	P_{DC}	6.06	15.2	4.3	1733	22.2
NREL3[a, b]	AADP	P_{DC}	7.88	8.7	10.0	669	146.1
NREL4[a]	ABBP	P_{DC}	6.82	1.7	18.5	1999	27.9
Digital power plant 1	AAAP	P_{DC}	5.0	8.2	0	0	0
Digital power plant 2	AAAP	P_{DC}	5.0	7.9	0	0	0
Digital power plant 3	AAAP	P_{DC}	5.0	8.6	0	0	0
Digital power plant 4	AAAP	P_{DC}	5.0	8.6	0	0	0

[a]Incomplete cases omitted.
[b]Negative and high values (>60) for P_{DC} and a high power time series standard deviation.

4.2.1 | Dataset quality issues

Various dataset quality issues are present in the selection of PV system power time series, as can be seen in Table 3, and are discussed here.

- *EURAC system*: No major data quality issues have been detected. This system has more than 10% outliers, probably due to cloudiness.
- *FOSS system*: No major data quality issues have been detected. This system had a higher amount of missing datapoints.
- *RSE systems*: No major data quality issues have been detected. These systems had a 25-day-long gap in the datasets.
- *Pfaffstaetten systems*: A relatively low amount of measured data has been reported. This can be seen in the missing % and the # of datagaps.
- *US DOE RTC baseline systems*: Several data quality problems have been detected, which resulted in some filter-metric-methods to be unable to calculate sensible *PLR* results for some methodologies. System c10hov6 experienced a 4-month-long initial inverter clipping followed by a period of 4 months without data. Afterwards, normal data acquisition without major issues is reported. The systems luemkoy and lwcb907 are also, at least partially, subject to inverter clipping, and negative power values are recorded. It is likely that the polarity has been switched for the time period of recorded negative values. For the luemkoy system, positive *PLR* values have been calculated which can be traced back to an initial power limitation due to inverter clipping followed by a period in which the power was not capped. System t3pg1sv is subject to significant inverter clipping. In the data of the systems wca0c5m and z0aygry, a data shift in the power output measurements has been detected additionally to inverter clipping.
- *NREL systems*: The power versus measured irradiance data for the PV systems NREL1 and NREL2 show a substantial number of outliers. Therefore, a large share of the raw data has to be filtered to ensure reliable data. For the PV systems NREL3 and NREL4, the measured irradiance sensor data were faulty and should be replaced by modeled clear-sky values (provided in raw data). The irradiance sensor used for NREL3 is installed in a distance of a few hundred meters away from the PV system, has a different tilt, which was translated to the POA, and shows decreasing irradiance values over time, possibly a result of a degrading reference cell.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4.3 | Uncertainty contributions to reported PLR results

Uncertainties in the final reported *PLR* result arise from multiple contributions, including measurement and sampling uncertainties. The power, temperature, and irradiance time series datasets which are used as the input to *PLR* determination, are a sample of a real-world PV system. Consider two different samples of this system, a 1-min interval and a 5-min interval time series, acquired over the same time period, but with different instruments. These two time series, measured by different instruments, would exhibit different standard deviations (σ_{meas}) due to the instrumentation's characteristics. If instead we use two identical measurement instruments, each measuring the system over the same time period, for example, by lagging on measurement by 30 s, these should have the same measurement uncertainty σ_{meas}, but two independent samples, their sampling uncertainties will vary when one considers the population mean μ of the *PLR* of the PV system, and will have sampling standard deviations (σ_{samp}) that are also different. Of course, each dataset is a sample of the system and contains contributions to their standard deviation arising from measurement and sampling uncertainty, intrinsically. The easiest way to determine the dataset standard deviation s of the time series is by seasonal decomposition, into its seasonal, trend, and residual components, and the sampling standard deviation is easily calculated from the time series residuals. The PVplr package reports the dataset standard deviation, as a % of initial power, as one of its measures of PV data quality.[22]

In many fields of research, it has long been reported that researchers do not distinguish the descriptive statistics of the standard deviation of a set of measurements and the standard error of the mean.[56] As a descriptive statistic the standard deviation is a measure of the variability of a set of measurements, arising from instrument and sampling effects. The standard error of the mean provides an estimate of our uncertainty in the "voted" value of a population mean, which in this study is supposed to be the true *PLR* value of the real-world PV system.[57] In this study we are benchmarking which filter-metric-method approach can determine the "true" voted value for a system. Researchers using common approaches to *PLR* determination, such as in the RdTools or PVplr packages, or using their own methods, are producing results which should replicate the "true" *PLR* of the system. We are benchmarking methods applied to diverse datasets, to identify the most robust approaches to determine the replication means for these 19 systems.[58] Typically replication studies attempt exactly the same method, and the standard error of the mean measures the variance among the attempts, to determine the population mean value of *PLR*. Here, since we have no a priori basis to know which of these many methods is "correct," we can expand to using differing methods, and by comparing standard errors, or better yet overlapping confidence intervals we can determine statistically, what the true, or as we refer to it, the "voted" *PLR* of that real-world PV system is. We are interested in the standard error of the population mean, and if we desire the standard 5% type I error rate[35] that is related to a p value of 0.05, we should compare *PLR* determination results (and filter-metric-method approaches) using 95% confidence intervals, determined from the standard error of the mean *PLR* of these results.

Consider that we want to know the true *PLR* of one PV system, and we use one or many *PLR* determination methods and calculate *PLR* 100 times, the *PLR* will vary around the mean, or "true" *PLR* and the important descriptive statistic is the standard error of the mean, as we calculate more values, our confidence in the mean improves. This is the basis of the "voting" method we apply here so that we can determine the most likely mean value of *PLR* for the 19 real-world PV systems, but utilizing many data filtering and statistical modeling methods and for each given dataset, the mean *PLR* value is probably the most likely.[59] And the standard error of the mean is a measure of the variance of these methods in determining the mean *PLR*. For comparing multiple methods of *PLR* determination for a single system, we can compare the different results and their 95% confidence interval (CI). With this, we will find the true population *PLR* mean, with in range of the 95% CI 19 out of 20 times. To determine if different methods show statistically similar, or different, estimates of the PV system's true *PLR*, we can check that the 95 % CIs are overlapping.[60] This is the approach we use here, for example, as shown for the EURAC system and multiple methods in Figure 9. If in this figure 95% CIs for each result were used, then we could define which methods give similar estimates of the PV systems mean *PLR* and which methods provide distinctly different estimates.

Regression based *PLR* uncertainty can be evaluated from the variance of the linear model coefficients of the corrected performance metrics with time (section Linear Regression), returning a standard deviation or 95% CI of the final *PLR* result. This process is described in Lindig et al.[1] When using this method with time series decomposed into components using either CSD or STL to determine the *PLR*, it is recommended to add back the residuals component into the trend component so that the final time series has the same signal to noise characteristics, as a *PLR* determined without using decomposition. In this way the uncertainties for regression *PLR* and regression on decomposed time series *PLR* are comparable. Otherwise, STL would have an apparent advantage due to the separation of the residuals' standard deviation from the regressed model line on the trend component. This cannot be used for YoY *PLR* as the individual *PLR* values used in YoY have no error given they are between two points only. YoY instead uses the probability distribution of the individual *PLR* results to represent the uncertainty in the reported *PLR*, and we will need to determine the correct comparative measure of uncertainty; should this be standard deviations, standard errors, or 95% or 83.4% confidence intervals.

Since the coefficient uncertainties from different methods are important to be able to compare, we wish to address this question so as to address this divide between regression and YoY *PLR* uncertainties. *PLR* uncertainty can also be evaluated using bootstrap resampling of the time series with replacement.[28,35,61,62] For regression determined *PLR*,[1] 65% of the days are randomly chosen from the total time series and the *PLR* is recalculated, and this process is repeated for 1000 iterations. The standard error of the distribution

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the *PLR*s obtained from each iteration relate to the uncertainty of the overall *PLR* value. A more stable *PLR* is expected to have less variance through resampling. This process can also be used with YoY as well; however, instead of resampling individual days, which would bias YoY, the final *PLR* distribution is resampled. Bootstrap resampling requires large computational capabilities, and the final uncertainty can be made artificially lower by increasing the number of iterations; however, it does offer a direct comparison between regression and YoY *PLR*.

The graphical display of 95% CIs and whether the ranges overlap is an effective way to show the relative uncertainties of the estimates of the mean *PLR*, and this graphical approach enables multiple comparisons, where as the traditional student *t* test is only a pairwise comparison.[63] For a 5% type I error rate, a 0.05 *p* value, we can compare methods for one system by plotting the 95% CIs. But the criteria will be different when we wish to determine if the mean *PLR* for two different PV systems are statistically similar (the null hypothesis), or different. In this case, the comparison of the estimates of two means, to achieve a 5% type I error, we should graphically compare the estimate of the mean *PLR* with 83.4% CIs, so as to attain a 5% capture rate.[64,65] This suggests that researchers who wish to compare on a common basis, different *PLR* determination methods should use 95% CIs, while a PV system fleet owner, who is wanting to determine which PV systems in the fleet are exhibiting similar or distinctly different performance loss, will compare 83.4% CIs.

4.4 | Data filtering PLR

It is useful to demonstrate the strong impact that data filtering has on all of these different *PLR* determination approaches, which is a concept that has not been emphasized in data analyses done to date. Instead, many researchers just stated the filtering they felt was reasonable, without documenting filtering's role and impact on reported *PLR* results. Here, we benchmark the complex role of filters on otherwise identical PLR data analyses. The applied filter can be found together with the corresponding number in Table 1. As a metric, PR_{Tcorr} has been used in monthly aggregation together with STL as the statistical method. In the case of missing data in the monthly PR_{Tcorr} time series, data imputation is accomplished using linear interpolation to address this aspect of missingness. The *PLR* of the digital system with degradation and real weather data was investigated next to the EURAC system. Figure 4 shows an example power versus irradiance plot for Filter #3 applied to the EURAC system. The power along the *y*-axis has been normalized to the nominal power of the system and a POA irradiance interval from 0 to 1250 W/m^2 is depicted along the *x*-axis. The blue dots represent the raw data, the green ones a first threshold filter, and the red points the final filtered dataset which is used for the subsequent *PLR* calculation steps. Some filters only include a threshold filter. In this case, the in-between filter step has been omitted, and just the final filtered data are shown.

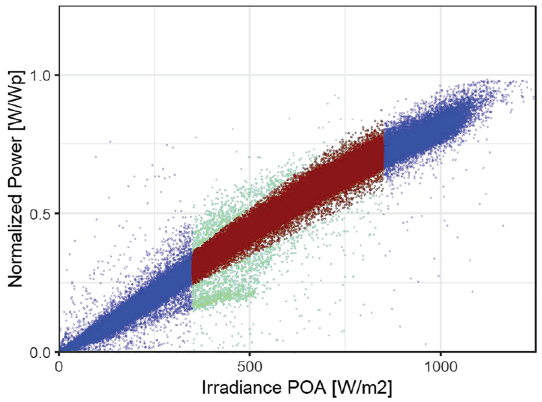


FIGURE 4 Normalized power versus irradiance plot for EURAC dataset with applied Filter #3; blue, raw data; green, threshold filter; red, final filter [Colour figure can be viewed at wileyonlinelibrary.com]

Figure 5 shows the calculated *PLR* values of the digital plant in dependence on the used filter together with power versus POA irradiance plots in order to get an idea of the impact of the individual filter. The same is shown for the EURAC system in Figure 6.

4.4.1 | Digital plant with degradation and real weather data

Looking at Figure 5, it is clear that the choice of filter does affect the *PLR* to a large extent. The calculated *PLR* values range from −4.48%/a to −5.47%/a. It seems that filters with similar irradiance cutoff thresholds are clustered together when using the same metric and calculation method. This correlation is especially pronounced for the datasets of the digital plants which are, due to their nature, not subject to outliers, and the outlier grade or filtering (mainly *PR* related) does not have any effect on the outcome. Stricter irradiance thresholds yield lower *PLR*s, an observation that could be exploited by an analyst to arrive at "desirable" *PLR* results. The application of Filter #3, #1, #4, and #7 results in *PLR* with the lowest calculated values, particularly close to the STC *PLR*, while the irradiance thresholds stretch from 350 to 800 W/m^2. Filter #6 and #9 apply a 200-W/m^2 cutoff resulting in *PLR* of −4.82%/a each, being quite close to the energy related *PLR*. Filter #2 and #8, both applying a very low irradiance threshold of 50 to 100 W/m^2, yield the results closest to the energy *PLR*. Filter #10 also applies a similar lower bound threshold, but additionally as well a high threshold at 1000 W/m^2, which appears to further increase the calculated *PLR* above the indicated value. Filter #5, which does not apply any threshold filter, but only a very loose *PR* filter, yields the highest deviations from both indicated degradation values. Considering PV plant data free of any data outliers, that is, Grade A in outliers, such as the digital once presented here, and the approach of applying STL together with PR_{Tcorr}, it seems that low irradiance cutoffs between 50 and 200 W/m^2 yield the most accurate results based on the energy related *PLR* described in Section 2.2 for the digital plants.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

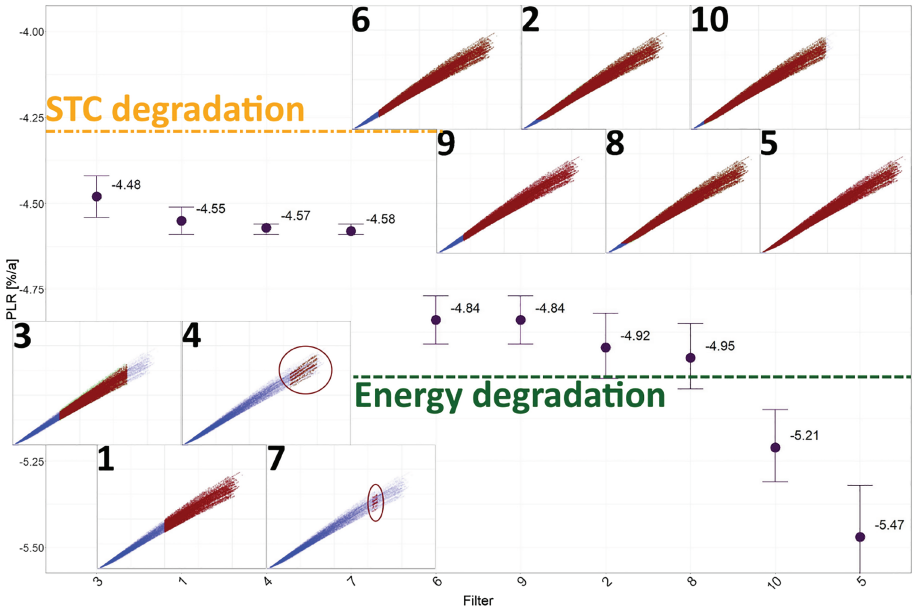


FIGURE 5 Calculated *PLR* using all proposed filter with PR_{Tcorr} as metric and STL as calculation statistical method for digital plant with degradation and real weather data [Colour figure can be viewed at wileyonlinelibrary.com]

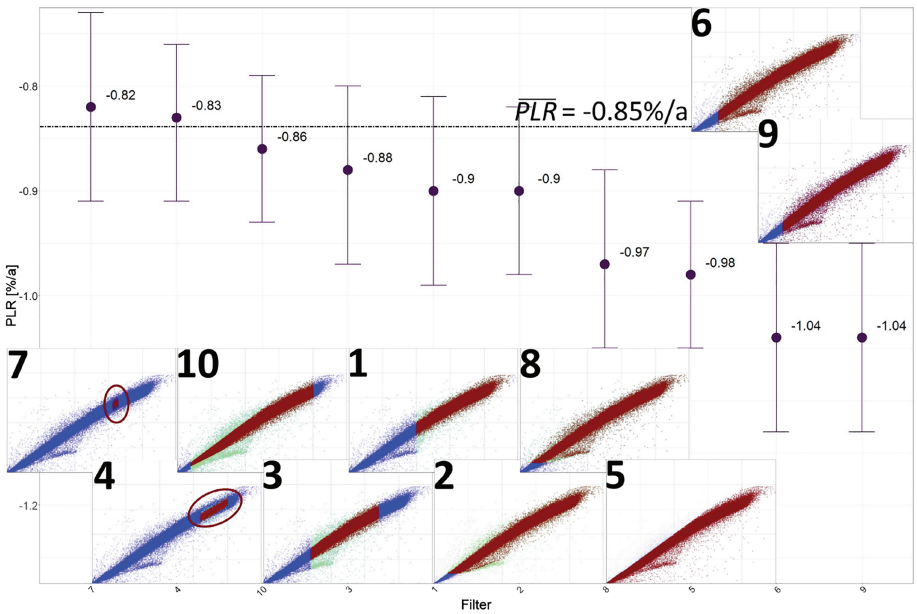


FIGURE 6 Calculated *PLR* using all proposed filter with PR_{Tcorr} as metric and STL as calculation statistical method for EURAC system [Colour figure can be viewed at wileyonlinelibrary.com]

4.4.2 | EURAC system

Looking at Figure 6, one can see that similar observations can be made to a certain extent for the EURAC plant. Again, higher irradiance thresholds tend to yield lower *PLR*. Additionally, low irradiance thresholds (e.g., Filter #10 and #2) give, in certain circumstances, accurate *PLR* results. As we deal here with real performance data, outlier accountability seems to play an important role as well.

Filter #7, #4, #10, #3, #1, and #2 yield *PLR* very close to the mean reference. Four of these six filters are in relatively narrow intervals, excluding power–irradiance pairs which are not representing the nearly linear relationship between both variables. If a metric is directly irradiance related, such as the *PR*, accounting for outlier in power–irradiance pairs is crucial to provide clean and representable data. It is visible that the usage of Filter #4 and #7, both subject to very strict irradiance filtering approaches, provides results close to the mean reference, at least for high-quality data. A problem of both approaches is the amount of filtered data. Below, the amount of data used for the final *PLR* calculation after filtering is shown in respect to raw data excluding nights for four different filter:

- Filter #1: 33.7%
- Filter #2: 61.1%
- Filter #4: 2.7%
- Filter #7: 0.4%

Filter #1 already applies a strict irradiance threshold at 500 W/m^2, but the data within the considered irradiance interval, although

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

just being 33.7% of the total amount of data, still account for roughly 80% of the produced power by the EURAC system and can therefor be considered as being representative. Instead, a vast amount of data is excluded in the *PLR* calculation using Filter #4 and #7. Although the methodologies perform well on the example dataset above and on some of the high-quality datasets in Section 4.6.2, it is believed that such a small amount of remaining data (2.7% and 0.4%) does possibly not represent the overall performance evolution well. Furthermore, depending on the location, such strict irradiance thresholds might reduce the amount of available data even further. Instead, given the used metric and statistical method, a narrow power–irradiance interval seems to be the filter of choice for real datasets including outlier. It has to be stressed that this does not hold for all metric-method combinations. For instance, Filter #9 used with PVWatts as metric and the YoY approach as statistical method yields quite accurate results for the EURAC system (see model YoY2 in Figure 9) whereas Filter #9 applied to PR_{Tcorr} combined with STL shows in direct comparison the largest deviation from the mean reference *PLR*. This circumstance underlines the strong dependency between all calculation steps from filtering up until the choice of a statistical method.

4.5 | Benchmarking PLR digital datasets

The digital datasets have been introduced along with the real PV plants in Section 2. Table 4 shows the most important characteristics of the plants.

Four PV systems have been modeled, two plants with 5 years of repeating weather data and two PV systems with 4 years of satellite data (location: Rennes in the west of France) followed by a fifth colder year. The benchmarking results of the digital plants serve as a reference since the real *PLR* values are known for these systems, which is not the case for the real datasets. Two definitions of *PLR* are indicated, the loss in P_{mpp} at STC and the constant absolute loss in energy from year to year. Both approaches are presented in the results. From a practical point of view, the parameter of interest is the degradation in energy.

TABLE 4 Main characteristics of digital PV plants

Characteristics	Repeating weather data	Real weather data with 5th induced cold year
No degradation	Same Meteo_0Deg	Real Meteo_0Deg
Induced degradation	Same Meteo_xDeg	Real Meteo_xDeg

The individual methods used in this comparison have been introduced in Table 1. Figure 7 shows the calculation results for the two digital plants without degradation. The colors of the dots indicate the chosen metrics and the symbols the applied statistical models. Two horizontal lines are visible in the figures, the orange one representing the degradation value for P_{mpp} at STC, and the green one the energy degradation. For this systems with no induced degradation, these values are very similar to one another.

Looking at the results, it is obvious that the calculation accuracy varies depending on the source of weather data. While all used approaches yield the correct value within a ±0.05% interval for the system with repeating weather data, greater deviations can be detected looking at the figure where real weather data with an induced colder year were used. Especially, the last induced colder year in the weather dataset seems to bias the results towards negative values. This colder year, subject to artificially induced lower radiation, yields just 63% of the initial power output for this system, without being subject to degradation. Methods using high irradiance threshold filter (VAR1, R-LR1, LS-LR1, and STL1) or predictive power approaches as metrics (i.e., XbX(UTC)) seem to be able to remove this induced effect and provide reliable results.

Figure 8 shows the results for the digital plants with simulated degradation. Here, the indicated degradation values (orange and green lines) vary distinctively from one another. Among other things, this has to do with the irradiance distribution for this particular simulated site. In Figure A1, the yearly POA irradiation per irradiance interval is shown for the dataset measured at the test site in Rennes, France. It is visible that more than 60% of the irradiation energy is provided from instantaneous irradiance values below 600 W/m^2. Thereby, a large amount of PV energy is produced under irradiance conditions far

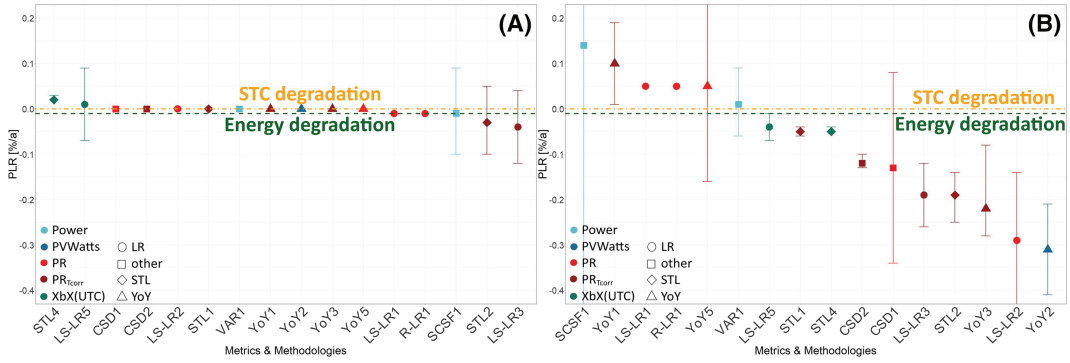


FIGURE 7 (A) Calculated *PLR* of digital plant with no degradation and repeating weather data. (B) Calculated *PLR* of digital plant with no degradation and real weather data; orange line: *PLR* at STC; green line: *PLR* in absolute energy [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

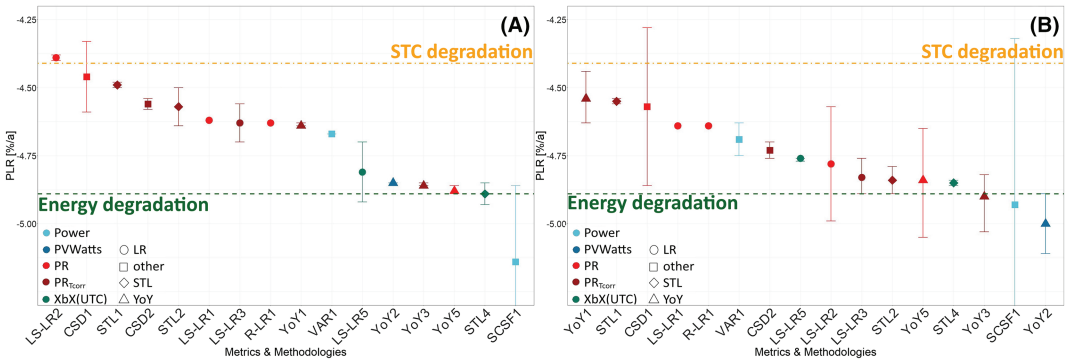


FIGURE 8 (A) Calculated *PLR* of digital plant with degradation and repeating weather data. (B) Calculated *PLR* of digital plant with degradation and real weather data; orange line: *PLR* at STC; green line: *PLR* in absolute energy [Colour figure can be viewed at wileyonlinelibrary.com]

away from STC, and therefore, both degradation values differ. It is expected that such degradation estimations are closer together in areas with a high amount of irradiation in higher irradiance level regimes.

For the digital systems with induced degradation, the deviations among the calculated *PLR* values are smaller between the systems, and most calculated values are higher than the STC *PLR* value and lower compared to the energy degradation value. As mentioned before, the degradation in energy is the indicator of interest. All approaches are within a 10% interval of the true energy *PLR*. It is visible that several YoY approaches yield values very close to the energy *PLR*. Furthermore, it seems that methods based on *PR* and PR_{Tcorr} metrics, except for YoY3 and YoY5, yield results with lower *PLR* values, laying between the lower and upper bound. It seems that the fifth colder year shifts most approaches to larger average *PLR* values.

For the two systems without degradation, the methodologies VAR1, LS-LR5, STL4, R-LR1, LS-LR1, YoY5, and STL1 yield the most accurate results with average deviations below −0.025%/a. For the systems with induced degradation, the application of YoY3, STL4, YoY5, and YoY2 returns results with deviations lower than −0.05%/a compared to the energy related *PLR*. It is believed that a threshold filter for irradiance and power affect the final *PLR* calculation results in a way such that higher thresholds yield lower *PLR* values. All four systems have very high-quality data where the effect of filtering irradiance–power pairs seems secondary in the sense that the main purpose of these filters, namely, outlier removal, is unnecessary as the raw datasets do not have any outlying values.

4.6 | Benchmarking PLR real datasets

4.6.1 | Evaluation methodology

An evaluation of the overall results for real datasets is fairly complicated since the true value of the PLR for each respective system is unknown. To rate the methodologies among each other, we used a replication study approach, where multiple filter-metric-method approaches to *PLR* were used to determine the *PLR* values and comparing this sample of *PLR* values as a sample of the true population mean of the PV system. Since we don't know the "correct" value, we use the sample mean from the calculations and using a voting procedure to identify the mean *PLR* as the mostly reasonable value for each PV system. First, the mean *PLR* for an individual PV system is calculated using all calculated *PLR* values (Figure 9A). Next, the relative difference of all methodologies from the reference (mean) *PLR*, which is set at 0%, is calculated for this particular system (Figure 9B). The closer a result is to 0%, the more accurate the calculated *PLR* is. We see, for example, that the highest deviations for the EURAC system are observed for statistical method LS-LR7 followed by SCSF1. LS-LR7 uses the 6K method as metric, which might be the root cause for the deviation. Statistical method SCSF1 has a different approach from the other methods, as it does not account for irradiance in the *PLR* calculation. SCSF1 practitioner's analysis suggested a positive sensor drift for the EURAC POA pyranometer accounts for the difference of their reported result from the mean result (a difference of +0.47%/a). Combining the suggested sensor drift with the estimation of degradation for the system, the deviance of the results using SCSF1 is reduced compared to the mean and yields a value similar to the majority of other methods.

The difference from the normalized mean *PLR* was subsequently calculated for all systems, and the values were averaged to see which methodologies seemingly yield the highest accuracy by cross-comparison. Here, two key performance indicators (KPIs) were then identified to benchmark the proposed combinations for the calculation of *PLR*.

- Absolute average deviation from the mean value considering all datasets

$$= \frac{\left|\sum_{i=1}^{n}\left(\frac{(\overline{PLR_i} - PLR_i)}{\overline{PLR_i}}\right)\right|}{n}. \quad (9)$$

- Standard error of the average deviation

$$= \sqrt{\frac{\sum_{i=1}^{n}\left(\frac{(\overline{PLR_i} - PLR_i)}{\overline{PLR_i}} - \mu\right)^2}{n-1}} / \sqrt{n}. \quad (10)$$

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

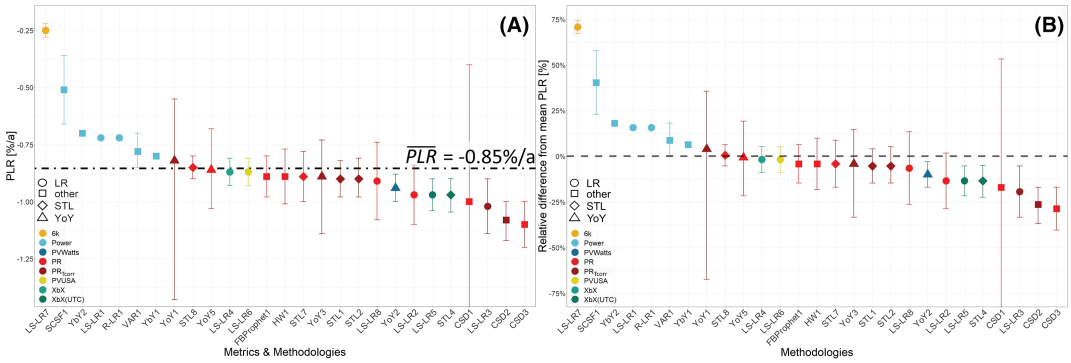


FIGURE 9 (A) Calculated *PLR* of EURAC system. (B) Relative calculated *PLR* values of EURAC system [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 5 Mean *PLR*, across all filter/methods used, for all systems included in the benchmarking study

System	*PLR* (%/a)	System	*PLR* (%/a)	System	*PLR* (%/a)
EURAC	−0.85	NREL1	−0.33	US DOE luemkoy[b]	0.95
FOSS	−0.71	NREL2	−0.54	US DOE lwcb907	−0.03
RSE CdTe	−1.75	NREL3[a]	0.06	US DOE t3pg1sv	−0.75
RSE pc-Si	−0.96	NREL4	−0.25	US DOE wca0c5m[c]	−1.00
Pfaffstaetten A	−3.57	US DOE c10hov6	−0.50	US DOE wxysjaf	−0.97
Pfaffstaetten B	−3.96	US DOE kobdpi8	−0.73	US DOE z0aygry[c]	−2.32
Pfaffstaetten C	−1.29				

[a]The provided modeled irradiance dataset should have been used, which was not done by all participants. The reported *PLR* corresponds to the average *PLR* of SCSF1, YoY2, and STL1 (see Figure A3).
[b]The system power was for approximately the first half of its recorded lifetime limited by inverter clipping. Afterwards, the output power was not capped anymore. A calculation of *PLR* using this power data series, which has been done by all participants, does not correspond to the mean *PLR*.
[c]The power datasets were subject to data shifts at the beginning of operation. These shifts were detected only by the participants applying STL1 and YoY2. The average of the results of these methodologies is reported as *PLR* (see Figure A4).

The first KPI provides an indication of how a particular filter-metric-method performs overall in terms of estimating the average value over all considered datasets. Thereby, the absolute average of the differences between the mean *PLR* ($\overline{PLR}_i$) and the *PLR* for each statistical method is calculated, where *i* refers to a specific PV system of the *n* PV systems being analyzed. The second KPI provides an indication on how the average value deviates from dataset to dataset. Here, μ is the mean of the numerator of Equation (9) over all systems for one filter-metric-method. Finally, all results are averaged in a target plot to see which methodologies perform the best across all systems. The results are discussed in Section 4.6.2. Figure 9 shows, in addition to the absolute *PLR* values, the uncertainties reported by the analysts for each applied methodology, which unfortunately were a variety of standard deviations, standard errors, and CIs, so are not actually comparable. Since there is no consensus on how to report *PLR* uncertainty values in the PV community, the analyst reported uncertainties were omitted in the final evaluation. In Table 5, the mean *PLR* for all systems are depicted to get an overall impression of the degradation of the systems under evaluation. The calculated *PLR* for all systems considering all methodologies can be found in Appendix C1. Unfortunately, some of the datasets under investigation had certain dataset quality issues, as discussed in Section 4.2 and Table 3, and individual methodologies failed to yield *PLR* results for these datasets. In Table 5, several systems belonging to the US DOE and NREL datasets are marked because of PLR calculation issues related to data quality.

4.6.2 | PLR analysis

Based on the discussion above the *PLR* evaluation is subdivided into different groups considering a varying number of analyzed systems. That was done to also study the variability of the results. For example, the *PLR* of the systems NREL3 and NREL4 was only calculated using five different methods of which a few of these results were strong outliers. Since the *PLR* evaluation is based on the mean of the calculated *PLR* values, a small predictor dataset may yield biased results. The idea is that an increasing number of calculated values decreases the average *PLR* variability and therefore increases the accuracy of the estimated reference *PLR*. Thereby, the trustworthiness of the *PLR*

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

as being close to the "real" *PLR* is higher if more *PLR* values are included.

With the voted results of our interlab/intermethod comparison shown in Table 5, we may actually have an ensemble learning approach to determine the $P\bar{L}R_i$ accurate and reproducible. Ensemble models in machine learning defines the usage of different modeling approaches, and the final result is a voted result across all models.[66] An example of this is random forest machine learning, where the result of a "forest" of decision tree models are averaged together, and this ensemble averaging allows the different approaches to counterbalance their uncertainties. In the case of PLR determination, with tools such as RdTools and PVplr, it is easy to perform an ensemble of PLR results on a system and then calculate the PV systems $P\bar{L}R_i$.

Thresholds for a minimum amount of calculated *PLR* values were set per evaluation group for a given filter-metric-method to be included in the benchmark comparison:

- G7-3-24: evaluate seven systems with a minimum of three calculated *PLR* values per filter-metric-method. This includes 24 filter-metric-method combinations.
- G13-7-17: evaluate 13 systems with a minimum of seven calculated *PLR* values per filter-metric-method of which at least one has not been included in first evaluation group. This includes 17 filter-metric-method combinations.

The threshold for an evaluation over all systems was set to 15 calculated *PLR* values. This reduces the number of considered filter-metric-method combinations to two, namely, STL1 and YoY2. Since these results are strongly biased, an evaluation over all systems has not been carried out.

G7-3-24: Evaluation of seven systems with 24 approaches

The first evaluation is based on the results calculated for seven datasets excluding all NREL and US DOE PV systems based on the dataset issues discussed before. For a *PLR* calculation method to be included in this analysis, at least three *PLR* values have to be calculated. The number of calculated *PLR* per system over these seven systems ranges from 20 to 27.

Figure 10A shows the *PLR* statistical method ratings when considering seven different PV systems. The deviance, a goodness-of-fit statistic for a statistical model, shown along the *x*-axis, describes the absolute overall difference from the reference mean *PLR*, and along the *y*-axis the standard error (se) of the average differences from the reference *PLR* across the systems under consideration is shown. Uncertainties are omitted because no homogenized way has been used throughout the application of different methodologies and therefore, the usage of the indicated uncertainties would be misleading. The difference in colors describes the usage of different metrics and the difference in symbols the usage of different statistical methods. The isobands, at 10% increase, are a guide for the eye to categorize the results into different groups of accuracy. In Figure 10A, a minimum of three calculated *PLR* was set for a statistical method to be included in the benchmark evaluating seven different PV systems. These datasets are considered as being of high quality without serious data quality issues apart from the minor ones reported in Section 4.2. It can be seen that the majority of applied filter-metric-method approaches have results in the first and second isoband with a relative average difference from the mean *PLR* of up to 17% and a corresponding standard error (se) of 1% to 6%. According to the results evaluating seven different PV systems, YbY1, STL4, STL8, LS-LR3, and HW1 provide the most accurate results (all in the first isoband). YbY1 has been applied to three PV systems, and the remaining three methods have been applied to all seven PV systems considered in the study. Two out of five of this methods use temperature-corrected metrics (PR_{Tcorr} and XbX(UTC)). YbY1 is one of the methodologies not applying any temperature correction, whereas this methodology applies the strictest overall filter by only including data within an irradiance interval of 40 W/m^2 around NOCT conditions.

From 24 tested filter-metric-method approaches, seven are not in the first and second isoband with deviance values from the mean *PLR* greater than 20% and perform thereby in direct comparison with lower accuracy. YbY2 is at the edge of the cluster with methodologies performing with higher accuracy. It seems that the usage of power as metric combined with LR and CSD as statistical methods results in higher uncertainty results. Especially, LR where the metric was not

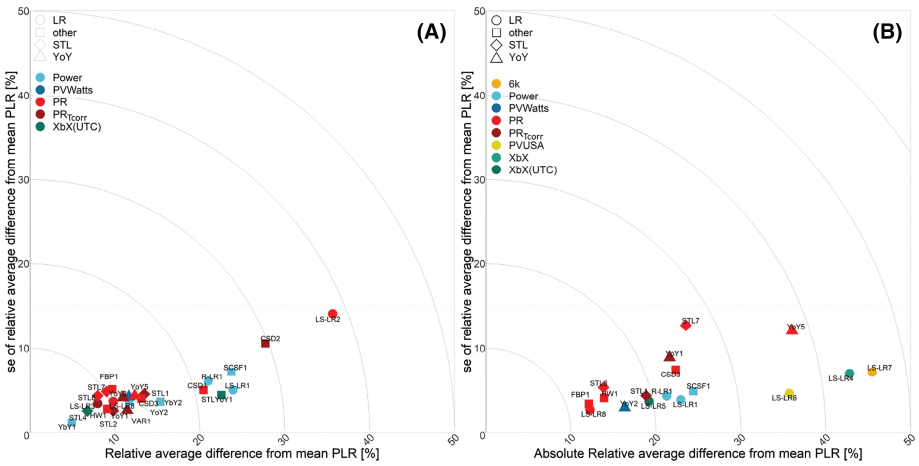


FIGURE 10 Target plot with absolute average deviations from mean *PLR* value and standard error considering (A) seven PV systems (excluding all NREL and US DOE datasets) and (B) 13 PV systems (excluding US DOE luemkoy, lwcb907, wca0c5m, and z0aygry and NREL3 and NREL4) [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

subject to temperature correction is subject to high variations. The statistical method with the lowest accuracy in direct comparison is LS-LR2, an approach using the *PR* as metric and LR as calculation method. The filter (see Table 1) used for LS-LR2 only applies a *PR* threshold of 0% to 100% and thereby does not exclude outlier sufficiently, which are ultimately affecting the final result since LR is strongly affected by nonvalid datapoints.

Methods using the common approaches of YoY and STL are performing with relatively lower uncertainty throughout, but also alternative models such as HW, FBP, or the VAR method yield satisfactory results.

It is interesting to observe the deviance in the results looking at YbY1 and YbY2. For both approaches, only three *PLR* values have been calculated. The only difference between these approaches are the applied filters. While YbY1 filters a narrow irradiance band around NOCT conditions (780–820 W/m^2), YbY2 filters around STC conditions (980–1020 W/m^2). Just considering this small sample of datasets, it seems that NOCT conditions represent the calculated mean *PLR* better in direct comparison. Furthermore, both methodologies, together with LS-LR1 and R-LR2, use narrow irradiance band and temperature filter and exclude thereby the vast majority of datapoints. Three out of four (except the mentioned YbY1) of this heavy filtering approaches yield results with higher uncertainties, possible because of filtering out large amounts of data. These filtering approaches are further discussed in Section 4.4.

A direct comparison of statistical method SCSF1 is more complicated as it does not include irradiance values for the *PLR* calculation. It evaluates power and irradiance time series independently. For instance, the application of SCSF1 on the EURAC systems irradiance data suggests a positive drift (see Section 4.6.1). If one combines that with the *PLR* calculated for the EURAC system using SCSF1, the difference of the *PLR* using statistical method SCSF1 compared to the result decreases. A similar observation was made for the FOSS system. The calculated *PLR* (using SCSF1) is with −0.35%/a clearly below the mean *PLR* of −0.7%/a. An evaluation of the irradiance sensor data is still ongoing to verify a possible drift, but it has been ensured that both sensors are calibrated according to existing guidelines and standards.

Instead, the results for the Pfaffstaetten systems using SCSF1 are quite close to the mean *PLR*, although a strong irradiance sensor drift of +0.67%/a is suggested using this statistical method. Irradiance drifts have to be considered with care, since interannual variations might contribute to this effect and are not excluded while estimating a sensor drift using the SCSF approach. Additionally, solar brightening effects are taking place since the early 1980s to this date, describing an increase of solar irradiation on the earth's surface in certain parts of the world.[67] For example, Kiefer et al.[68] saw an average increase in irradiance of +1.1% per year while studying the performance of several PV plants in Germany. These effects might as well influence such measurements. It seems that a direct comparison of the SCSF approach to others is quite complex for real datasets. A comparison based on a greater number of digital plants may present a better foundation for further evaluation.

Furthermore, CSD did not perform well. CSD is based on a centered moving average where the first and last datapoints are removed by creating a statistical smoothing function. This may lead to the exclusion of important performance data, especially in shorter time series. Comparing the individual CSD approaches with one another, it is visible that CSD3 performed the best. This is the only CSD approach having a statistical *PR* filter which excludes outlier of power–irradiance pairs.

Overall, the majority of the test approaches calculate *PLR* with relatively low uncertainties considering this datasets. This study serves thereby as a first indicator of *PLR* estimation accuracy for high-quality datasets without major measurement and operation issues.

G13-7-17: Evaluation of 13 systems with 17 approaches

The second evaluation considers 13 PV systems excluding NREL3, NREL4, US DOE luemkoy, US DOE lwcb907, US DOE wca0c5m, and US DOE z0aygry. The latter two systems were excluded because of the detected data shift in the power output. As explained in Appendix C1, only two participants, using the approaches STL1 and YoY2, took the data shift into account. The inclusion of the *PLR* evaluation results for these two PV systems would alter the results based on a thorough data quality check instead of the actual statistical method application.

A minimum of seven calculated *PLR* was set for a statistical method to be included in this second benchmark category. At the same time at least one calculated *PLR* had to be for a NREL or US DOE dataset in order to avoid having redundant results compared to the first evaluation. The threshold reduced the number of considered methodologies from 24 to 17. The results can be seen in Figure 10B.

The inclusion of methodologies in itself is already a first quality characteristic by being applicable to a wider set of PV system performance data, which are partially subject to certain data related issues. Comparing both benchmark evaluations, it is visible that the spread of methodologies in the target plot in Figure 10B increased substantially. While the majority of methodologies in Figure 10A yields results corresponding to values in the first and second isoband, seven of now 17 methodologies are remaining in this area and the rest stretches over the other isobands. The approaches with the highest accuracy are LS-LR8 and FBP1. Both methodologies use the *PR* as metric and apply the same Filter #8 from Table 1. It has to be noted that, although very accurate results have been achieved using this two methodologies, results where serious data issues have been detected were omitted. Four more statistical methods have been tested using the same metric and filter. It is interesting to note that two of these use STL as a statistical method, the only difference is that one function was taken from `Python` (STL8) while the other one was taken from `R` (STL7). The latter exhibits the highest se in Figure 10B. A cross-comparison of STL7 and STL8 shows that both approaches yield very similar results, except for system NREL1. Here, STL7 overestimates the mean *PLR* substantially, whereas STL8 results in a *PLR* lower than the mean value, but with a lower deviation. This overestimation explains the highest standard error across all tested methodologies.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Apart from LS-LR8, FBP1, and STL8, four more methodologies are in the second isoband, namely, HW1, YoY2, STL1, and LS-LR5. It should be noted that YoY2 and STL1 are next to YoY5, the only approaches for which all 13 *PLR* have been calculated and provided. Furthermore, LS-LR5 stands in direct comparison to LS-LR6, LS-LR4, and LS-LR7, three methodologies with high deviations from the mean *PLR*. For all four methods, Filter #2 from Table 1 and LS-LR as statistical method are used together with different power predictive models. Thereby, 6K, PVUSA, and XbX do not seem to yield reliable results with the 6K metric performing the poorest. Due to its nature the 6K metric tends to predict values close to the nameplate power at STC and underestimates thereby the *PLR*. Instead, XbX(UTC) provides already satisfactory results by just applying LR. It is expected that the usage of a more sophisticated statistical method such as STL or YoY would return results with lower uncertainties. Method STL4, which uses XbX(UTC) as metric, performed with high accuracy in the first benchmark considering seven systems but was not applied to a sufficient number of systems to be included in this second benchmark.

Overall, a "perfect" combination of filter, metric, and *PLR* calculation method probably does not exist, since there are complex interactions of filters, metrics, and methods with the characteristics of the datasets. Instead, based on the results discussed in this section, case-to-case-dependent arrangements of dataset-dependent adaptive filter (possibly automated based on quantitative data quality measures), temperature-corrected metrics, and suited methods are recommended. Although LS-LR did yield some good results, it is not recommended for more complicated datasets as it gives too much weight to outliers. The commonly used statistical methods STL and YoY performed well if suited filter and metrics have been applied.

5 | CONCLUSIONS

In this work, several experts in the field of PV reliability have applied their preferred methods for the calculation of *PLR* including various combinations of filters, metrics, and statistical methods. The aim of the exercise was to determine the capabilities and uncertainties of all the various approaches tested and to determine the "best" voted *PLR* estimates for the 23 systems under investigation. Furthermore, we wanted to quantify the variation and sensitivity of the individual *PLR* calculation steps based on the uncertainty, reproducibility, and bias of the results. This work is an important step forward to support the community to have tools to assess data quality of research and commercial PV systems, to utilize various filters, metrics, corrections and methods, and to have these tools be accessible and broadly usable. Examples are the software packages RdTools, PVlib and PVplr.

Based on the presented results, one uniform way of calculating reliable *PLR*, including fixed filtering approaches, the same metric, and statistical method, does not seem to exist at that point. Instead, it was shown that a thorough data quality check together with careful filtering approaches are absolutely crucial steps in calculating *PLR*, especially if the PV system dataset is subject to monitoring data related issues. In terms of calculation approaches, not only the most popular ones, STL and YoY, but also newly developed or less common ones, such as the VAR method, HW or FBP, demonstrated reliable results. While a standardized way of calculating *PLR* would be the desirable outcome of this study in order to reliably intercompare results across PV systems and operators, it was highlighted that a sensitive combination of filtering/metrics and statistical methods is crucial for achieving reliable results.

Even if we currently cannot define a single way to calculate the PLR of a PV system, this study does suggest that the voting, or preference aggregation, approach used here, may itself represent an accurate ensemble approach for PLR determination. It seems that a calculating PLR using many filters, performance metrics, corrections, and statistical modeling approaches does appear to provide consistent and robust estimates of $\bar{PLR}_i$ for system *i*. This multiple method approach may serve as an ensemble model in which inaccuracies of all the different approaches are minimized in the voted result of the ensemble calculation of $\bar{PLR}_i$.

Often, the filtering step is either performed insufficiently or not reported clear enough in corresponding literature. Therefore, when calculating *PLR*, an exhaustive report on filter selection and data flagging is vital to better comprehend, or even reproduce, the calculation steps and to enable the reader to compare the results. This circumstance gave way to the implementation of a direct filter comparison, where 10 different filtering approaches have been applied to one digital plant and one real dataset with an otherwise identical *PLR* calculation approach. Filtering can be divided into two sets, namely, threshold filter and statistical filter used to remove outliers of power–irradiance pairs.

High irradiance thresholds tend to lower the overall *PLR*, especially when modules/systems are experiencing degradation in low light performance due to a decreasing shunt resistance. Additionally, filtering approaches including statistical filter, which remove a large share of power–irradiance data pair outliers, together with low to medium irradiance thresholds seem to provide the most reliable datasets for further treatment and consequently result with the highest accuracy.

ACKNOWLEDGEMENTS

The authors kindly acknowledge the data provided by Karl Berger (AIT) for the Pfaffstaetten system and Giosué Maugeri (RSE) for the RSE datasets. The research has received funding from the European Union's Horizon 2020 program under GA. No. 721452-H2020-MSCA-ITN-2016. This material is based upon work supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under the Solar Energy Technologies Office Award Number 34366. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of Energy or the United States Government. The authors thank the Department of Innovation, Research and University of the Autonomous Province of Bozen/Bolzano for covering the Open Access publication costs.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Sascha Lindig https://orcid.org/0000-0001-5421-8265
David Moser https://orcid.org/0000-0002-4895-8862
Arash Khalilnejad https://orcid.org/0000-0002-7138-8095
Dirk Jordan https://orcid.org/0000-0002-2183-7489
Chris Deline https://orcid.org/0000-0002-9867-8930
Wilfried van Sark https://orcid.org/0000-0002-4738-1088
Wei Luo https://orcid.org/0000-0002-7075-3521

REFERENCES

1. Lindig S, Kaaya I, Weiß K, Moser D, Topic M. Review of statistical and analytical degradation models for photovoltaic modules and systems as well as related improvements. *IEEE J Photovoltaics*. 2018;8(6):1773-1786.
2. Jordan DC, Deceglie MG, Kurtz SR. PV degradation methodology comparison—a basis for a standard. In: 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC); 2016:0273-0278.
3. Bryant C, Wheeler NR, Rubel F, French RH. KGC: Koeppen-Geiger Climatic Zones. https://CRAN.R-project.org/package=kgc, R package version 1.0.0.2; 2017.
4. Kottek M, Grieser J, Beck C, Rudolf B, Rubel F. World map of the Köppen-Geiger climate classification updated. *Meteorol Z*. 2006;15:259-263.
5. Rubel F, Brugger K, Haslinger K, Auer I. The climate of the European Alps: shift of very high resolution Köppen-Geiger climate zones 1800–2100. *Meteorol Z*. 2016. http://www.schweizerbart.de/papers/metz/detail/prepub/87237/The_climate_of_the_European_Alps_Shift_of_very_hig?af=crossref
6. Rubel F, Kottek M. Observed and projected climate shifts 1901-2100 depicted by world maps of the Koppen-Geiger climate classification. *Meteorol Z*. 2010;19(2):135-141.
7. Jordan DC, Marion B, Deline C, Barnes T, Bolinger M. PV field reliability statusâĂŤanalysis of 100,000 solar systems. *Prog Photovoltaics: Res Appl*. 2020;28(8):739-754. https://doi.org/10.1002/pip.3262
8. IEC61724-1:2017. Photovoltaic system performance, part 1: monitoring. *Standard*, Geneva, CH, International Electrotechnical Comission; 2017.
9. Sandia National Laboratories and United States. Department of Energy. Office of Scientific and Technical Information. The US DOE Regional Test Center Program: driving innovation quality and reliability. https://books.google.it/books?id=-x0YzQEACAAJ; 2015.
10. Braisaz B, Dupeyrat P, Lindsay A, Radouane K, Helm P, Jager-Waldau A, Mine A. An advanced model of PV power plants based on Modelica. In: 28th European Photovoltaic Solar Energy Conference and Exhibition. WIP; 2013:3644-3648. http://www.eupvsec-proceedings.com/proceedings?paper=21666
11. Plessis G, Kaemmerlen A, Lindsay A. BuildSysPro: a Modelica library for modelling buildings and energy systems. In: Proceedings of the 10th International Modelica Conference; 2014:1161-1169.
12. 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*. 2018;8(1):224-232.
13. Livera A, Theristis M, Koumpli E, Theocharides S, Makrides G, Sutterlueti J, Stein JS, Georghiou GE. Data processing and quality verification for improved photovoltaic performance and reliability analytics. *Prog Photovoltaics: Res Appl*. 2020;29:143-158. https://doi.org/10.1002/pip.3349
14. Huld T, Friesen G, Skoczek A, Kenny RP, Sample T, Field M, Dunlop ED. A power-rating model for crystalline silicon PV modules. *Sol Energy Mater Sol Cells*. 2011;95(12):3359-3369.
15. Kratochvil JA, Boyson WE, King DL. *Photovoltaic array performance model*: Sandia Technical Report; 2004.
16. Buuren S, Groothuis-Oudshoorn C. mice: multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1-67.
17. Lindig S, Louwen A, Moser D, Topic M. Outdoor PV system monitoring input data quality, data imputation and filtering approaches. *Energies*. 2020;13(19):5099. http://doi.org/10.3390/en13195099
18. Reno MJ, Hansen CW. Identification of periods of clear sky irradiance in time series of GHI measurements. *Renew Energy*. 2016;90:520-531.
19. Holmgren W, Lorenzo T, Krien U, Mikofski M, Hansen C, Driesse A, Boeman L, Miller E, Anoma MA, Beutner V, Hendricks T, Dollinger J, Leroy C, Stark C, Anderson K, Gaffiot J, Oos J, Peronato G, Schachler B, Mathew A, Kafkes A. Pvlib/pvlib-python: V0.7.2. Zenodo, https://zenodo.org/record/2850192; 2020.
20. Deceglie MG, Jordan DC, Nag A, Deline CA, Shinn A. RdTools: an open source Python library for PV degradation analysis. In: PV Systems Symposium; 2018; Albuquerque, United States.
21. Jordan DC, Deline C, Kurtz SR, Kimball GM, Anderson M. Robust PV degradation methodology and application. *IEEE J Photovoltaics*. 2018;8(2):525-531.
22. Curran AJ, Burleyson TJ, Lindig S, Moser D, RH French. PVplr: SDLE performance loss rate analysis pipeline. https://CRAN.R-project.org/package=PVplr; 2020.
23. Diez DM, Çetinkaya-Rundel M, Barr CD. *OpenIntro Statistics*. 4th edition: OpenIntro, Inc.; 2019. https://www.openintro.org/book/stat/
24. R Core Team. R: the R Stats Package. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html; 2020.
25. Rauch E. rstl. https://pypi.org/project/rstl/; 2018.
26. Seabold S, Perktold J. statsmodels: econometric and statistical modeling with Python. In: 9th Python in Science Conference; 2010:92-96.
27. Theristis M, Livera A, Micheli L, Jones CB, Makrides G, Georghiou GE, Stein JS. Modeling nonlinear photovoltaic degradation rates. In: 2020 47th IEEE Photovoltaic Specialists Conference (PVSC); 2020:0208-0212.
28. Curran AJ, Jones CB, Lindig S, Stein J, Moser D, French RH. Performance loss consistency and uncertainty across multiple methods and filtering criteria. In: IEEE 46th Photovoltaic Specialists Conference (PVSC); 2019; Chicago:1328-1334.
29. Curran AJ, Zhang R, Hu Y, Haddadian R, Braid JL, Peshek TJ, French RH. 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: IEEE 44th Photovoltaic Specialists Conference (PVSC); 2017; Washington:1927-1932.
30. Jordan DC, Deceglie MG, Kurtz SR. PV degradation methodology comparison—a basis for a standard. In: 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC); 2016; Portland, United States:273-278.
31. King DL, Kratochvil JA, Boyson WE. Field experience with a new performance characterization procedure for photovoltaic arrays. Sandia Technical Report, United States; 1998.
32. National Renewable Energy Laboratory. PVWatts version 5 manual. https://pvwatts.nrel.gov/downloads/pvwattsv5.pdf; 2014.
33. Holmgren W, Hansen C, Mikofski M. pvlib python: a python package for modeling solar energy systems. *J Open Source Softw*. 2018;3:884. PVLIB version 0.5.2 https://doi.org/10.5281/zenodo.1246152

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

34. Lindig S, Moser D, Müller B, Kiefer K, Topic M. Application of dynamic multi-step performance loss algorithm. In: 2020 47th IEEE Photovoltaic Specialists Conference (PVSC); 2020:0443-0448.
35. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning: With Applications in R*. 1st ed. 2013, Corr. 5th printing 2015 edition, Springer Texts in Statistics: Springer; 2013. http://www-bcf.usc.edu/gareth/ISL/index.html
36. Faraway JJ. *Linear Models with R, 2nd Ed*. 2nd edition: Chapman and Hall/CRC; 2014(English).
37. Hyndman RJ, Athanasopoulos G. *Forecasting: Principles and Practice*. 2nd edition: OTexts; 2018. https://otexts.com/fpp2/
38. Jordan DC, Kurtz SR. Analytical improvements in PV degradation rate determination. In: 35th IEEE Photovoltaic Specialists Conference; 2010:688-693.
39. Cleveland RB, Cleveland WS, McRae JE, Terpenning I. STL: a seasonal-trend decomposition procedure based on loess. *J Off Stat*. 1990;6(1):3-33.
40. Hafen R. stlplus: enhanced seasonal decomposition of time series by loess. https://CRAN.R-project.org/package=stlplus, R package version 0.5.1; 2016.
41. 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: 2013 IEEE 39th Photovoltaic Specialists Conference (PVSC); 2013: 0007-0012.
42. Meftah M, Lajoie-Manzenc E, van Iseghem M, Perrin R, Boubil D, Radouane K. A less environment-sensitive and data-based approach to evaluate the performance loss rate of PV power plants. In: 36th EU PVSEC Proceedings; 2019; Marseille, France:1554-1559.
43. Meyers B, Tabone M, Kara EC. Statistical clear sky fitting algorithm. In: 2018 45th IEEE Photovoltaic Specialists Conference (PVSC); 2018; Hawai.
44. Meyers B, Deceglie M, Deline C, Jordan D. Signal processing on PV time-series data: robust degradation analysis without physical models. *IEEE J Photovoltaics*. 2019;10:546-553.
45. Holt CC. Forecasting seasonals and trends by exponentially weighted moving averages. *Int J Forecast*. 2004;20(1):5-10.
46. Taylor SJ, Letham B. Forecasting at scale. *Am Stat*. 2018;72(1): 37-45.
47. Theristis M, Livera A, Jones CB, Makrides G, Georghiou GE, Stein JS. Nonlinear photovoltaic degradation rates: modeling and comparison against conventional methods. *IEEE J Photovoltaics*. 2020;10(4):1112-1118.
48. Kwak SK, Kim JH. Statistical data preparation: management of missing values and outliers. *Kor J Anesthesiol*. 2017;70(4):407-411. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5548942/
49. Tierney NJ, Cook DH. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. arXiv:180902264 [stat], http://arxiv.org/abs/1809.02264; 2020.
50. Khalilnejad A, Karimi AM, Kamath S, Haddadian R, French RH, Abramson AR. Automated pipeline framework for distributed processing of large-scale building energy time series data. *PLoS ONE*. 2020;18:1-22.
51. Aggarwal CC. *Outlier Analysis*. 2nd ed. 2017 edition: Springer; 2016 (English).
52. Hodge V, Austin J. A survey of outlier detection methodologies. *Artif Intell Rev*. 2004;22(2):85-126.
53. Chen C, Liu L-M. Joint estimation of model parameters and outlier effects in time series. *J Am Stat Assoc*. 1993;88(421):284-297. http://www.jstor.org/stable/2290724
54. López-de-Lacalle J. TsOutliers: detection of outliers in time series. https://CRAN.R-project.org/package=tsoutliers; 2019.
55. Gelman A, Hill J. *Chapter 25. Missing Data Imputation, from Data Analysis Using Regression and Multilevel/Hierarchical Models*. 1st edition: Cambridge University Press; 2006.
56. Horan BF. Standard deviation, or standard error of the mean? Anaesthesia and Intensive Care; 1982.
57. Curran-Everett D. Explorations in statistics: standard deviations and standard errors. *Advances in Physiology Education*. 2008;32(3): 203-208. https://journals.physiology.org/doi/full/10.1152/advan.90123.2008
58. Cumming G, Williams J, Fidler F. Replication and researchers' understanding of confidence intervals and standard error bars. *Underst Stat*. 2004;3(4):299-311. https://doi.org/10.1207/s15328031us0304_5
59. Muravyov SV, Marinushkina IA. Processing data from interlaboratory comparisons by the method of preference aggregation. *Measurement Techniques*. 2016;58(12):1285-1291. https://doi.org/10.1007/s11018-016-0886-4
60. Goldstein H, Healy MJR. The graphical presentation of a collection of means. *J R Stat Soc. Ser A (Stat Soc)*. 1995;158(1):175. https://www.jstor.org/stable/10.2307/2983411?origin=crossref
61. Efron B, Tibshirani RJ. An introduction to the bootstrap. *Monographs on Statistics and Applied Probability*, Vol. 57: Chapman & Hall/CRC; 1993.
62. Efron B, Tibshirani R. Improvements on cross-validation: the 632+ bootstrap method. *J Am Stat Assoc*. 1997;92(438):548-560. https://doi.org/10.1080/01621459.1997.10474007
63. Payton ME, Greenstone MH, Schenker N. Overlapping confidence intervals or standard error intervals: what do they mean in terms of statistical significance? *J Insect Sci*. 2003;3:34-40. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC524673/
64. Cumming G, Finch S. Inference by eye: confidence intervals and how to read pictures of data. *Am Psychol*. 2005;60(2):170-180.
65. Kamath SM. Energy use intensities across building use types and climate zones using the CBECS dataset. *Ph.D. Thesis*: Case Western Reserve University; 2020. https://etd.ohiolink.edu/pg_10?::NO:10:P10_ETD_SUBID:184439
66. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*. 2nd edition: Springer; 2009.
67. Wild M, Gilgen H, Rsch A, Ohmura A, Long CN, Dutton EG, Forgan B, Kallis A, Russak V, Tsvetkov A. From dimming to brightening: decadal changes in solar radiation at earth's surface. *Science*. 2005;308(5723): 847-850.
68. Kiefer K, Farnung B, Müller B, Reinartz K, Rauschen I, Külnter C. Degradation in PV power plants: theory and practice. In: 36th EU PVSEC Proceedings; 2019; Marseille, France:9-13.
69. Ascencio-Vasquez J, Brecl K, Topic M. Methodology of Köppen-Geiger-photovoltaic climate classification and implications to worldwide mapping of PV system performance. *Sol Energy*. 2019;191: 672-685.

How to cite this article: Lindig S, Moser D, Curran AJ, et al. International collaboration framework for the calculation of performance loss rates: Data quality, benchmarks, and trends (towards a uniform methodology). *Prog Photovolt Res Appl*. 2021;29:573–602. https://doi.org/10.1002/pip.3397

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX A

TABLE A1 Description of datasets used in the *PLR* benchmarking exercise

Dataset	Technology	Country	P_{nom} (kWp)	Time period	Azimuth	Tilt	Measurement data available	Köppen–Geiger climate classification[3a, b, c]	PV sensitive climate classification[69a, b, d]
EURAC	pc-Si	Italy	4.20	02/11-01/19	188.5°	30°	G_{POA}, T_{amb}, WS,	Dfb to ET	EM to DM
							T_{mod}, P_{DC}, P_{AC}		
FOSS	mc-Si	Cyprus	1.03	06/06-05/16	180°	27.5°	G_{POA}, T_{amb}, WS,	Csa to BSk	DH to CH
							T_{mod}, P_{DC}		
RSE CdTe	CdTe	Italy	1.16	06/09-12/18	180°	30°	G_{POA}, T_{amb}, P_{AC}	Cfa to Cfb	DM
RSE pc-Si	pc-Si	Italy	1.68	06/09-12/18	180°	30°	G_{POA}, T_{amb}, P_{AC}	Cfa to Cfb	DM
Pfaffstaetten A[a]	pc-Si	Austria	2.11	01/13-04/19	220°	22°	G_{POA}, T_{amb},	Cfb to Dfb	DM
							T_{mod}, P_{DC}, P_{AC}		
Pfaffstaetten B[a]	pc-Si	Austria	2.06	01/13-04/19	220°	22°	G_{POA}, T_{amb},	Cfb to Dfb	DM
							T_{mod}, P_{DC}, P_{AC}		
Pfaffstaetten C[a]	CIGS	Austria	2.25	01/13-04/19	220°	22°	G_{POA}, T_{amb},	Cfb to Dfb	DM
							T_{mod}, P_{DC}, P_{AC}		
US DOE c10hov6	mc-Si	USA	3.24	03/15-05/18	180°	35°	G_{POA}, T_{amb},	BSk to Cfb	CK to DH
							T_{mod}, P_{DC}		
US DOE kobdpi8	mc-Si	USA	3.24	11/15-05/18	180°	35°	G_{POA}, T_{amb},	BWk to BWh	BK to CK
							T_{mod}, P_{DC}		
US DOE luemkoy	mc-Si	USA	3.24	11/14-05/18	180°	35°	G_{POA}, T_{amb},	Dfb	EM
							T_{mod}, P_{DC}		
US DOE lwcb907	mc-Si	USA	3.24	11/14-05/18	180°	35°	G_{POA}, T_{amb},	Dfb	EM
							T_{mod}, P_{DC}		
US DOE t3pg1sv	mc-Si	USA	3.24	03/15-05/18	180°	35°	G_{POA}, T_{amb},	BSk to Cfb	CK to DH
							T_{mod}, P_{DC}		
US DOE wca0c5m	mc-Si	USA	3.24	12/14-05/18	180°	30°	G_{POA}, T_{amb},	Cfa	DH
							T_{mod}, P_{DC}		
US DOE wxzsjaf	mc-Si	USA	3.24	11/15-05/18	180°	35°	G_{POA}, T_{amb},	BWk to BWh	BK to CK
							T_{mod}, P_{DC}		

(Continues)

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A1 (Continued)

Dataset	Technology	Country	P_{nom} (kWp)	Time period	Azimuth	Tilt	Measurement data available	Köppen–Geiger climate classification[3a, b, c]	PV sensitive climate classification[69a, b, d]
US DOE z0aygry	mc-Si	USA	3.24	12/14-05/18	180°	30°	G_{POA}, T_{amb},	Cfa	DH
							T_{mod}, P_{DC}		
NREL1	mc-Si	USA	2.70	05/16-07/19	180°	30°	G_{POA}, T_{amb}, WS,	Dfb to BSk	DH to CH
							T_{mod}, P_{DC}, P_{AC}		
NREL2	HIT	USA	1.00	08/07-12/16	180°	40°	G_{POA}, T_{amb},	Dfb to BSk	DH to CH
							T_{mod}, P_{DC}, P_{AC}		
NREL3	mc-Si	USA	94.00	09/09-01/18	175°	10°	G_{POA}, $G_{POA_{mod}}$, T_{amb},	Dfb to BSk	DH to CH
							WS, T_{mod}, P_{AC}		
NREL4	mc-Si	USA	524.00	07/11-05/18	165°	9.1°	G_{POA}, $G_{POA_{mod}}$, T_{amb},	Dfb to BSk	DH to CH
							WS, T_{mod}, P_{AC}		
4 digital	c-Si	France	4 × 1.82		180°	21°	G_{POA}, T_{amb}, WS,	Cfb	DM
power plants[b]							P_{DC}, P_{AC}		

[a]Second-hand modules.
[b]Irradiance data from HelioClim for Rennes, France.
[c]A, tropical; B, arid, C, temperate; D, continental; E, polar climates; f, no dry season; m, monsoon, s, dry summer; w, dry winter; S, steppe; W, desert; a, hot summer; b, warm summer; c, cold summer; d, very cold summer; h, hot; k, cold.
[d]A, tropical; B, desert; C, steppe; D, temperate; E, cold; F, polar climates; L, low; M, medium; H, high; K, very high irradiation zones. The first climate zone corresponds to the nearest location labeled, while the second one corresponds to the most adjacent climate zone.

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

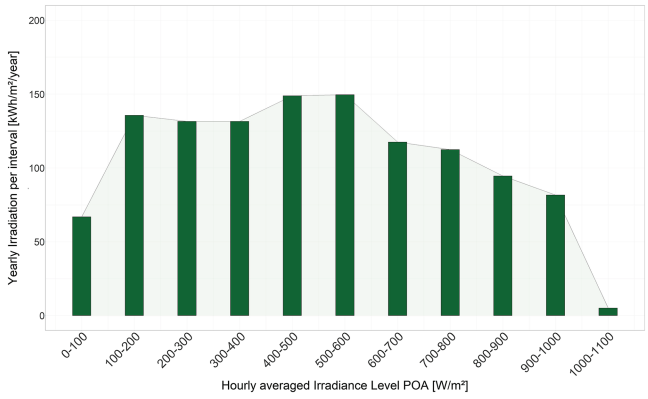
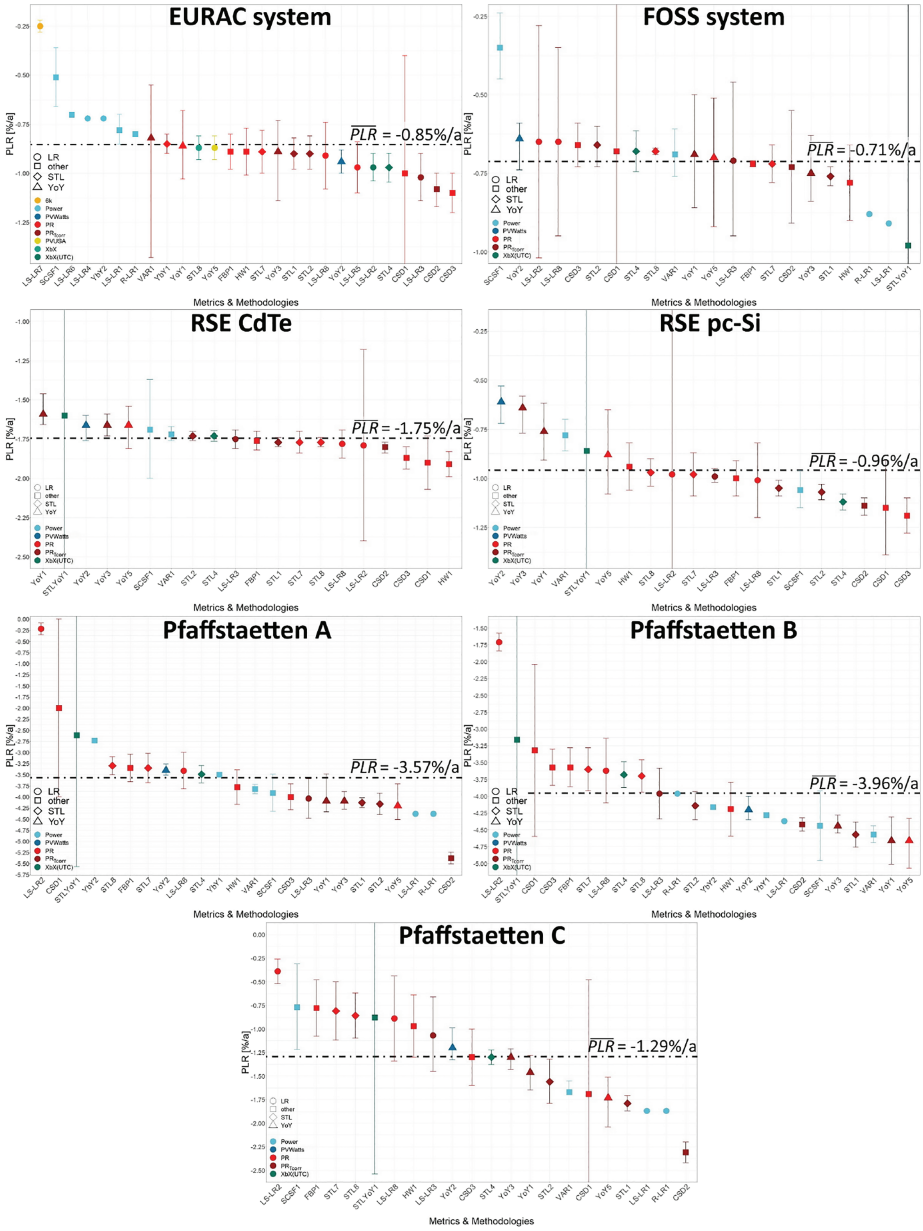


FIGURE A1 Irradiance distribution for digital power plant (location: Rennes, France). Plane-of-array irradiance distribution per 100-W/m^2 interval [Colour figure can be viewed at wileyonlinelibrary.com]

Here, the calculated *PLR* of all systems are listed. The number of calculated values per system varies due to monitoring data issues participants were facing while working on the data. They are discussed in greater detail in Section 4.2.1. The methodologies are depicted on the *x*-axis and the *PLR* on the *y*-axis. The colors indicate the chosen metrics and the symbols the applied statistical models. The horizontal lines indicate the mean *PLR* for the respective system. An exception are the US DOE systems wca0c5m and z0aygry as well as NREL3. For the US DOE systems, the mean *PLR* corresponds to the average *PLR* of the methodologies STL1 and YoY2. That is because both system datasets were subject to data shifts at the beginning of operation (see Section 4.2.1). This shift should have been detected and excluded for the *PLR* calculation. This was done only with statistical method STL1 and YoY2. For NREL3, the mean *PLR* corresponds to the average *PLR* of the methodologies

FIGURE A2 Calculated *PLR* of EURAC system, FOSS system, RSE CdTe & pc-Si systems, Pfaffstaetten A, B & C systems [Colour figure can be viewed at wileyonlinelibrary.com]



1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

SCSF1, YoY2, and STL1. YoY2 and STL1 used the provided modeled clear-sky irradiance data series as an input, and the methodology SCSF1 is not based on any irradiance data series. The remaining approaches used the faulty measured irradiance dataset as input the corresponding results deviate thereby substantially from the "true" *PLR*.

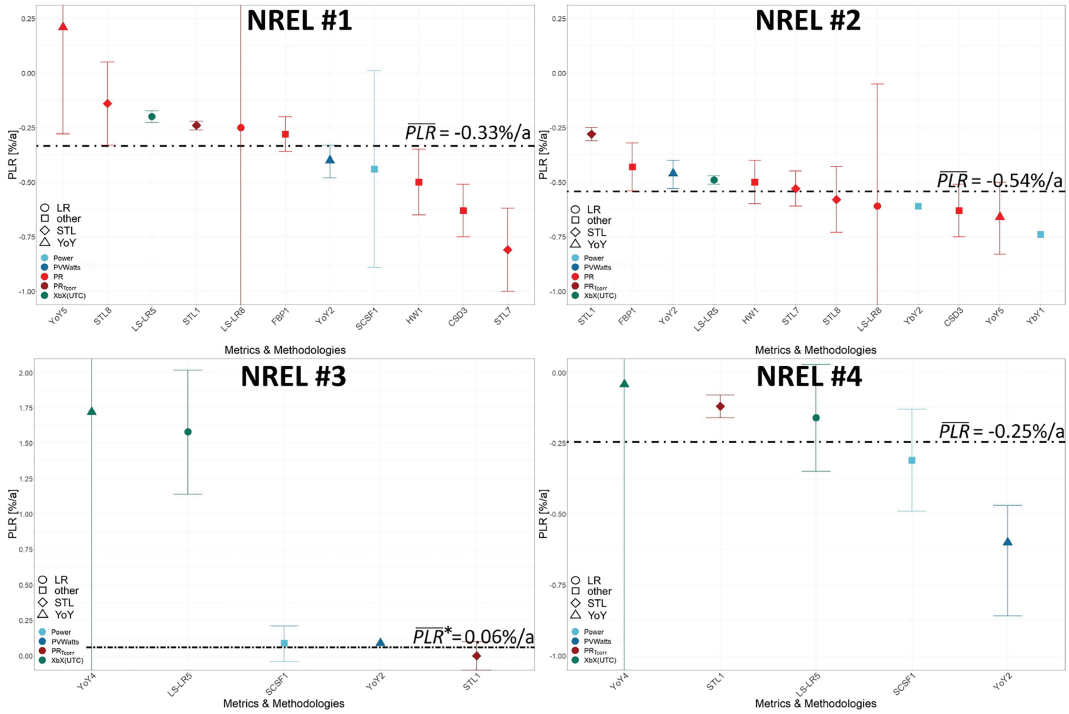


FIGURE A3 Calculated *PLR* of NREL systems; mean *PLR* * for NREL3 corresponds to average *PLR* of SCSF1, STL1 and YoY2 [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

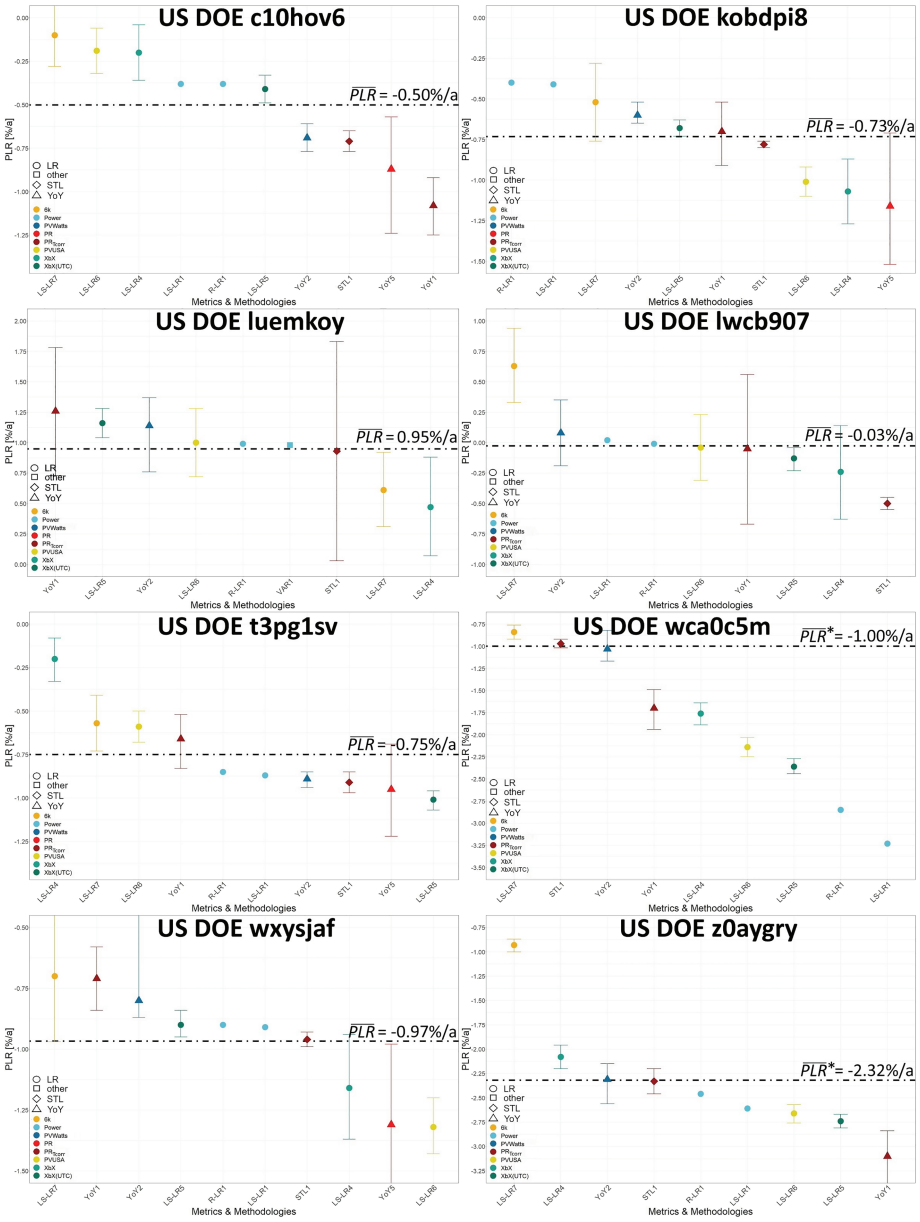


FIGURE A4 Calculated *PLR* of US DOE systems; mean *PLR** for US DOE wca0c5m and z0aygry corresponds to average *PLR* of STL1 and YoY2 [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

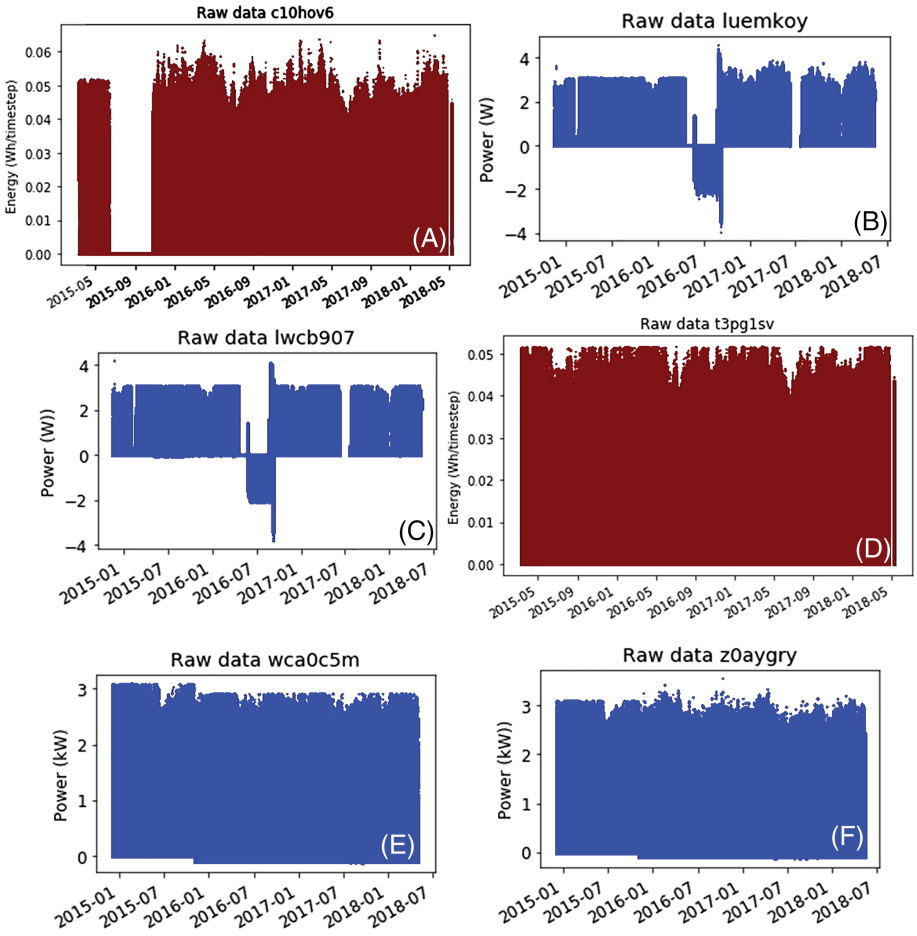


FIGURE A5 Data quality issues US DOE systems: Energy (red)/power (blue) over time for US DOE datasets: (A) US DOE c10hov6—initial inverter clipping and missing data; (B) US DOE luemkoy—negative power values and initial inverter clipping; (C) US DOE lwcb907—negative power values and inverter clipping; (D) US DOE t3pg1sv—inverter clipping; (E) US DOE wca0c5m—data shift after 1 year and inverter clipping; and (F) US DOE z0aygry—data shift after 1 year and initial inverter clipping [Colour figure can be viewed at wileyonlinelibrary.com]

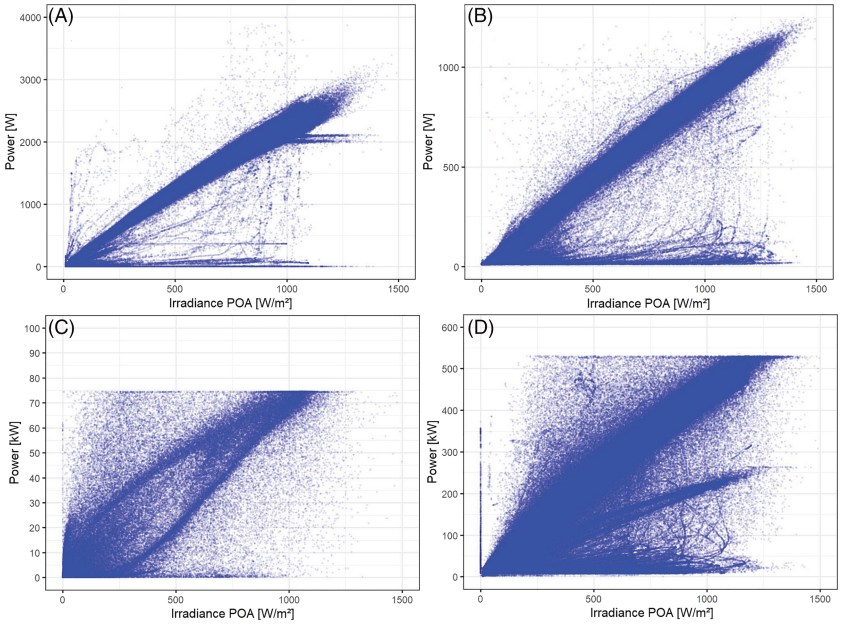


FIGURE A6 Data quality issues NREL systems: Power over plane-of-array irradiance for NREL datasets: (A) measured irradiance—NREL1—numerous outlier detected; (B) measured irradiance—NREL2—numerous outlier detected; (C) modeled irradiance—NREL3—extreme outlier and inverter clipping detected; and (D) modeled irradiance—NREL4—extreme outlier and inverter clipping detected [Colour figure can be viewed at wileyonlinelibrary.com]

1099159x, 2021, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/pip.3397 by CASE WESTERN RESERVE UNIVERSITY, Wiley Online Library on [23/01/2023]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License