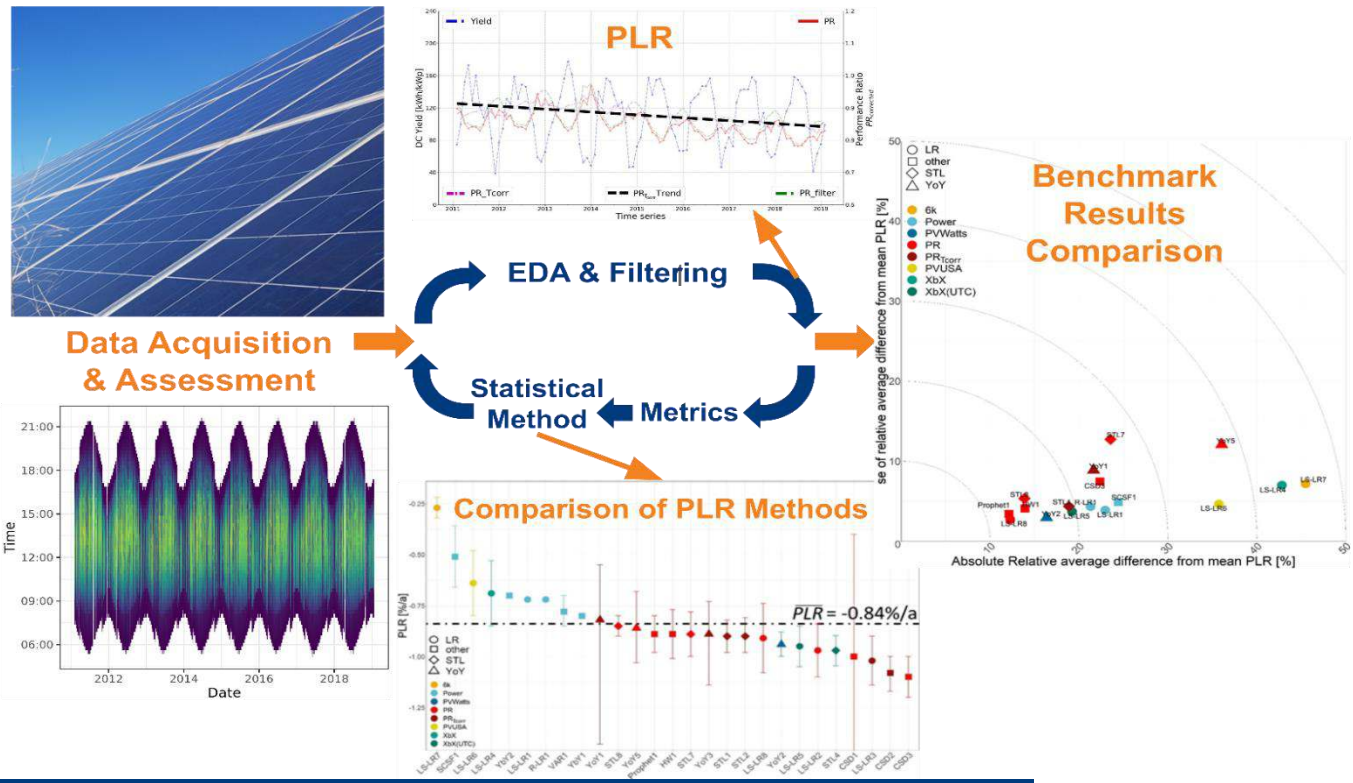
International Energy Agency
Photovoltaic Power Systems Programme



Task 13 Performance, Operation and Reliability of Photovoltaic Systems

PVPS

Assessment of Performance Loss Rate of PV Power Systems

2021

Report IEA-PVPS T13-22:2021

What is IEA PVPS TCP?

The International Energy Agency (IEA), founded in 1974, is an autonomous body within the framework of the Organization for Economic Cooperation and Development (OECD). The Technology Collaboration Programme (TCP) was created with a belief that the future of energy security and sustainability starts with global collaboration. The programme is made up of 6000 experts across government, academia, and industry dedicated to advancing common research and the application of specific energy technologies.

The IEA Photovoltaic Power Systems Programme (IEA PVPS) is one of the TCP's within the IEA and was established in 1993. The mission of the programme is to "enhance the international collaborative efforts which facilitate the role of photovoltaic solar energy as a cornerstone in the transition to sustainable energy systems." In order to achieve this, the Programme's participants have undertaken a variety of joint research projects in PV power systems applications. The overall programme is headed by an Executive Committee, comprised of one delegate from each country or organization member, which designates distinct 'Tasks,' that may be research projects or activity areas.

The IEA PVPS participating countries are Australia, Austria, Belgium, Canada, Chile, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Malaysia, Mexico, Morocco, the Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Thailand, Turkey, and the United States of America. The European Commission, Solar Power Europe, the Smart Electric Power Alliance (SEPA), the Solar Energy Industries Association and the Cop- per Alliance are also members.

Visit us at: www.iea-pvps.org

What is IEA PVPS Task 13?

Within the framework of IEA PVPS, Task 13 aims to provide support to market actors working to improve the operation, the reliability and the quality of PV components and systems. Operational data from PV systems in different climate zones compiled within the project will help provide the basis for estimates of the current situation regarding PV reliability and performance.

The general setting of Task 13 provides a common platform to summarize and report on technical aspects affecting the quality, performance, reliability and lifetime of PV systems in a wide variety of environments and applications. By working together across national boundaries we can all take advantage of research and experience from each member country and combine and integrate this knowledge into valuable summaries of best practices and methods for ensuring PV systems perform at their optimum and continue to provide competitive return on investment.

Task 13 has so far managed to create the right framework for the calculations of various parameters that can give an indication of the quality of PV components and systems. The framework is now there and can be used by the industry who has expressed appreciation towards the results included in the high-quality reports.

The IEA PVPS countries participating in Task 13 are Australia, Austria, Belgium, Canada, Chile, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, Thailand, and the United States of America.

COVER PICTURE

The process of *PLR* determination, after initial exploratory data analysis and data quality grading, consists of the four steps are 1) input data cleaning and filtering, 2) performance metric selection, corrections and aggregation, 3) time series feature corrections and finally 4) application of a statistical modeling method to determine the Performance Loss Rate value.

ISBN 978-3-907281-10-9

INTERNATIONAL ENERGY AGENCY

PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

IEA PVPS Task 13

Performance, Operation and Reliability of Photovoltaic Systems

Assessment of Performance Loss Rate of PV Power Systems

Report IEA-PVPS T13-22:2021

April 2021

ISBN 978-3-907281-10-9

AUTHORS

Main Authors

Roger H. French, Case Western Reserve University, USA
Laura S. Bruckman, Case Western Reserve University, USA
David Moser, EURAC Research, Italy
Sascha Lindig, EURAC Research, Italy
Mike van Iseghem, EDF, France
Björn Müller, Fraunhofer-ISE, Germany
Joshua S. Stein, Sandia, USA
Mauricio Richter, 3E, Belgium
Magnus Herz, TÜV Rheinland, Germany
Wilfried Van Sark, Utrecht University, The Netherlands
Franz Baumgartner, Zürcher Hochschule für Angewandte Wissenschaften, Switzerland

Contributing Authors

Julián Ascencio-Vásquez, 3E, Belgium
Dario Bertani, RSE, Italy
Giosué Maugeri, RSE, Italy
Alan J. Curran, Case Western Reserve University, USA
Kunal Rath, Case Western Reserve University, USA
JiQi Liu, Case Western Reserve University, USA
Arash Khalilnejad, Case Western Reserve University, USA
Mohammed Meftah, EDF, France
Dirk Jordan, NREL, USA
Chris Deline, NREL, USA
Georgios Makrides, FOSS Research Centre for Sustainable Energy, Cyprus
George Georghiou, FOSS Research Centre for Sustainable Energy, Cyprus
Andreas Livera, FOSS Research Centre for Sustainable Energy, Cyprus
Bennet Meyers, Stanford University, USA
Gilles Plessis, EDF, France
Marios Theristis, Sandia, USA
Wei Luo, SERIS, Singapore

Editors

Roger H. French, Case Western Reserve University, USA
Boris Farnung, VDE Renewables GmbH, Germany

TABLE OF CONTENTS

Acknowledgements 6
List of abbreviations 7
Executive summary 9
1 Introduction 11
1.1 How is performance loss rate calculated? 11
1.2 Data imputation, filtering and correction approaches 13
1.3 Metrics 16
1.4 Statistical methods 18
1.5 Combinations of performance metrics & *PLR* calculation models 21
2 Description of *PLR* benchmarking datasets 22
2.1 Data characteristics: time interval, time length, data types 22
2.2 Systems with monthly power only 22
2.3 Systems with high quality time series power & weather data 25
2.4 Systems with higher-order time series data types 32
3 Calculation of *PLR* by multiple methodologies 33
3.1 *PLR* pipeline workflow 33
3.2 Example calculation PV system *PLR* 33
3.3 Low quality data *PLR* results 35
3.4 High quality data *PLR* results 37
3.5 New opportunities from analysis of time-series *I-V*, P_{mpp} datasets 60
3.6 The role of *PLR* in PV system long term yield assessments 61
3.7 Critical factors in *PLR* determination 61
4 Conclusions 63
References 65
Appendices 72
A. Irradiance distribution for digital power plant (location: Rennes/France) 72
B. Data quality issues of PV system datasets 73
C. *PLR* results for all PV systems 75

ACKNOWLEDGEMENTS

This report received valuable contributions from several IEA-PVPS Task 13 members and other international experts. Many thanks to:

The contribution of data by Karl Berger, supported by the Austrian government, by means of the Austrian Federal Ministry for Climate Action, Environment, Energy, Mobility, Innovation and Technology (bmk.gv.at), represented by the Austrian Research Promotion Agency (FFG), under contract No. 876763.

Case Western Reserve University's work on this report was supported by the U.S. Department of Energy's Office of Energy Efficiency and Renewable Energy (EERE) under Solar Energy Technologies Office (SETO) Agreement Number DE-EE-0008172. We acknowledge Erdmut Schnabel and Michael Köhl of Fraunhofer ISE for the I-V, P_{mpp} time series dataset in Section 3.5. Leonie Kemper of Case Western Reserve University prepared the cover figure.

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 multi-mission 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 of Energy or the United States Government.

The research has received funding from the European Union's Horizon 2020 programme under GA. No. 721452 – H2020-MSCA-ITN-2016.

This report is supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) under contract no. 0324304A and 0324304B.

LIST OF ABBREVIATIONS

6k	The 6k model for PV performance
ABD	Airport code for Bolzano/Italy
AC	Alternating Current
BSh	Arid climate, steppe climate, hot desert Köppen-Geiger climate zone
BSk	Arid climate, steppe climate, cold desert Köppen-Geiger climate zone
BWh	Arid climate, desert climate, hot desert Köppen-Geiger climate zone
CI	Confidence Interval
Cfa	Warm temperate climate, fully humid, with hot summer Köppen-Geiger climate zone
Dfb	Snow climate, fully humid, with warm summer Köppen-Geiger climate zone
ET	Polar climate, Frost Köppen-Geiger climate zone
CdTe	Cadmium Telluride, a PV absorber
CIGS	Copper Indium Gallium Selenide, a PV absorber
CPLR	Change Point Linear Regression
CRAN	Comprehensive R Archive Network
CSD	Classical Seasonal Decomposition
CSI	Clear Sky Index
DAQ	Data Acquisition (Device)
DbD	Day by Day
DC	Direct Current
DOE	Department of Energy of the United States
EDA	Exploratory Data Analysis
EDF	Électricité de France
ESL	Electronic Solar Load
EURAC	Accademia Europea Bolzano/Europäische Akademie Bozen
FOSS	The Research Centre for Sustainable Energy, at the University of Cyprus
IEA	International Energy Agency
IEC	International Electrotechnical Commission
ISE	Fraunhofer Institute for Solar Energy Systems
I-V	Current – Voltage
Prophet	The Prophet R Package
GHI	Global Horizontal Irradiance
HW	Holt-Winters
KG	Köppen-Geiger, name of the climate zone system
KPI	Key Performance Indicator
kWp	Kilowatts Peak
kWh	Power in Kilowatt Hours
kVA	kilo volt amps
LID	Light Induced Degradation
LOESS	Locally estimated scatterplot smoothing, a non-parametric regression method

LR	Linear Regression
LS	Least Squares
LS-LR1	Least Squares, Linear Regression method version 1
LTYP	Long Term Yield Prediction
MbM	Month by Month
MPP	Maximum Power Point
NM	New Mexico
NOCT	Nominal Operating Cell Temperature
NREL	National Renewable Energy Laboratory
P	Power
PLR	Performance Loss Rate
POA	Plane of Array
PP	Predicted Power
PR	Performance Ratio
PV	Photovoltaic
PVlib	A Python3 Package for Modeling Solar Energy Systems
PVplr	An Open Source R Package for PV Performance Loss Rate Determination
PVPS	Photovoltaic Power Systems Programme
PVSC	Photovoltaics Specialist Conference
PVUSA	The Photovoltaics for Utility Scale Applications project
PVWatts	A Simple-to-use Photovoltaic System Energy Model
R	Robust Regression
R-LR1	Robust Regression, Linear Regression method version 1
RdTools	An Open Source Python Library for PV Degradation Analysis
RSE	Ricerca sul Sistema Energetico
RTC	Regional Test Center
SCSF	Statistical Clear Sky Fitting
Si	Silicon, a PV absorber
SNL	Sandia National Laboratory
STC	Standard Test Conditions
STL	Seasonal and trend decomposition using Loess
STL7	STL version 7 method
STL8	STL version 8 method
UCY	University of Cyprus
US	United States
UTC	Coordinated Universal Time
VAR	Yearly variations of output power with respect to yearly variations of environment
WbW	Week by Week
XbX	X by X
YbyY	Year by Year
Yield	The power divided by the installed capacity, in kWh/kWp
YoY	Year on Year

EXECUTIVE SUMMARY

This IEA PVPS Task 13, Subtask 2.5 reports on a benchmarking study of the various approaches for calculating the Performance Loss Rates (*PLR*) of commercial and research photovoltaic (PV) power plants in diverse climatic zones. *PLR*s are calculated with data from the PV systems' power and weather data. The *PLR* is used by power plant owners, operators, and investors to determine the expected power output of a PV system over its installed life. Therefore, discrepancies in various calculation methods can greatly impact the financial around a PV installation. This benchmarking study is necessary due to the inconsistency in reported *PLR* results based on the many different approaches currently used to calculate *PLR* of PV systems. This study is focused on identifying which of the various approaches produce similar results and what causes inconsistencies between these different methods.

The findings of the study lead to a *PLR* framework which defines the basic four steps common to *PLR* determination. After initial exploratory data analysis and data quality grading, the four steps are 1) input data cleaning and filtering, 2) performance metric selection, corrections, and aggregation, 3) time series feature corrections, and 4) application of a statistical modeling method to determine the *PLR* value. The *PLR* of 19 high quality research PV systems and four simulated (aka "digital") PV systems using the various available PLR methodologies. These 23 datasets are now open access datasets for the PV community. This reports shows the impact of data quality and missing data on PLR calculations. Additionally, the "true value" of *PLR* (i.e., mean *PLR* ($\overline{PLR_i}$)) of each of the *i* systems studied is reported.

The *PLR* results were compared between the different calculation methods using statistical, data-driven, and deterministic analytical methods. These results help define which analysis methods produce results that cluster around the mean *PLR* of the individual PV systems. The results of the *PLR* framework for each PLR calculation method are benchmarked in terms of a) their deviation from the $\overline{PLR}$ value, and b) their uncertainty, standard error and confidence intervals. Of the 19 systems studied, nine systems had $\overline{PLR_i}$ values between -0.4%/annum to -1%/a, 3 systems showed lower $\overline{PLR_i}$ values, and six had larger $\overline{PLR_i}$ values in the range from -1%/a to -4%/a.

Various statistical modeling methods can be applied for the calculation of the *PLR* of PV systems. Furthermore, the selections made at each calculation step are highly interdependent such that the individual steps cannot be assessed individually. In addition, the different methods used are impacted by the quality and missingness of the specific dataset in a complex manner such that one cannot identify particular methods as more relatively more robust.

The key findings of this report are:

- Data quality of the research and commerical PV systems impact the calculated *PLR* results. Exploratory data analys is important to assess, quantify, and grade the input datasets in order to understand the reliabilty or bias of reported results and to make choices on the appropriate methodology. If more than 10% of the daytime data is missing, then data imputation techniques are recommended.
- The degree of data filtering can impact the stability of the *PLR* results. Heavy data filtering can introduce strong bias in the *PLR* results, enabling a user to raise or lower the reported *PLR* of a PV system When calculating and reporting *PLR*, an exhaustive report on filter selection and data cleaning is vital to better comprehend the steps in the *PLR*

calculation. Reported *PLR* values need to be reproducible by others and have clearly reported confidence intervals, so that results among systems are comparable at a 5 % significance level.

- The choice of Performance Ratio *(PR)* or Power (*P)* does not strongly influence the *PLR* results and give comparable results ; therefore, neither metric is preferred over the other.
- The uncertainity of the *PLR* is determined by the quality of data (power and weather). When there is high quality of data to compare between different types of *PLR* calculations on a single PV system, the results should be standardized on the 95 % confidence intervals. When comparing *PLR* results between multiple systems, the results should be standardized at the 83.4 % confidence intervals. In both of these cases, this standardization corresponds to a p-value, capture ratio and significance level of 0.05 and is suggested be best practice. If a time series decomposition is used in the statistical modeling, then the residuals should be retained with the trend, to report comparable confidence intervals.
- In cases where local weather data is not available, it is possible to use satellite-based weather data.
- Higher order time series data such as *I-V, P_{mpp}* (max power point) datastreams, by virtue of containing more information, represent an important opportunity for advanced analytics of PV system performance and degradation.

Careful data filtering is an essential foundation for reliable *PLR* analysis. Filtering can be divided into two categories: threshold filters and statistical filters used to remove outliers in power-irradiance pairs. High irradiance threshold filters tend to lower the reported *PLR* which is not necessarily representative of real system performance. Statistical filtering (to remove the anomalous power-irradiance data pairs) in combination with low to medium irradiance thresholds (to retain a larger amount of the system's data) provides the most reliable datasets for the next steps in *PLR* determination and produces the most accurate results.

These results will inform standards development for *PLR* determination, which was previously attempted with an initial proposal for a new IEC 61724-4 standard. However, the results reported here suggest that proposing a specific standardized method is still premature.

Even if we have not yet defined a single way to calculate the *PLR* of a PV system, this study suggests that the preference aggregation approach may itself represent an accurate ensemble approach for *PLR* determination. By calculating *PLR* using many filters, performance metrics corrections, data aggregations, time series corrections, and statistical modeling approaches we can provide consistent and robust estimates of $\overline{PLR_i}$ for PV system *i*. This ensemble, multiple method, approach may serve as the best model for minimizing the inaccuracies found in the different approaches for determining $\overline{PLR_i}$.

1 INTRODUCTION

The Performance Loss Rate (*PLR*) of a photovoltaic (PV) system is a parameter, which indicates the decline of the power output over time and is provided in units of % per annum (%/a, or %/year). 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 operations and maintenance (O&M) practices. The goal of this Task 13 Subtask is to define a framework of analytical steps that are required for *PLR* determination, and assess the reliability and reproducibility of the many different approaches and methods used by the PV community to determine and report the *PLR* of a PV system. Another important aspect is to establish an approach to assess the quality of PV system power and weather datasets from research PV systems and commercial PV power plants, and to identify the applicability of these analytical approaches, and their steps, to different types of PV system datasets that can be of varying quality[1].

1.1 How is performance loss rate calculated?

In this work, *PLR* has been calculated based on DC power readings if they are available, otherwise AC power has been used. An overview of the available measurement data for the individual systems can be found in Section 1.2.8, Table 1. Figure 1 presents the necessary steps for calculating the *PLR*. The steps include gathering and understanding of the input data, the application of certain filters, the selection and aggregation of a performance metric including possible corrections and the application of models to calculate *PLR*. Typically *PLR* has been reported as a linear rate, which is the simplest, first order model of the temporal change in PV system power production, which we refer to as the "assumed linear *PLR*". This linear *PLR* is simplest manner of quantifying the temporal drop in power output over the system's lifetime. However, field experience has shown non-linearities in a system's *PLR*, so we have moved to a second *PLR* model fit based on change-point segmented regression, which reports the time of the change point and the linear *PLR* for both segments in the dataset. This non-linear model can capture and quantify the more complex observed behavior of real systems, and can also be extended if the data is of sufficient quality.

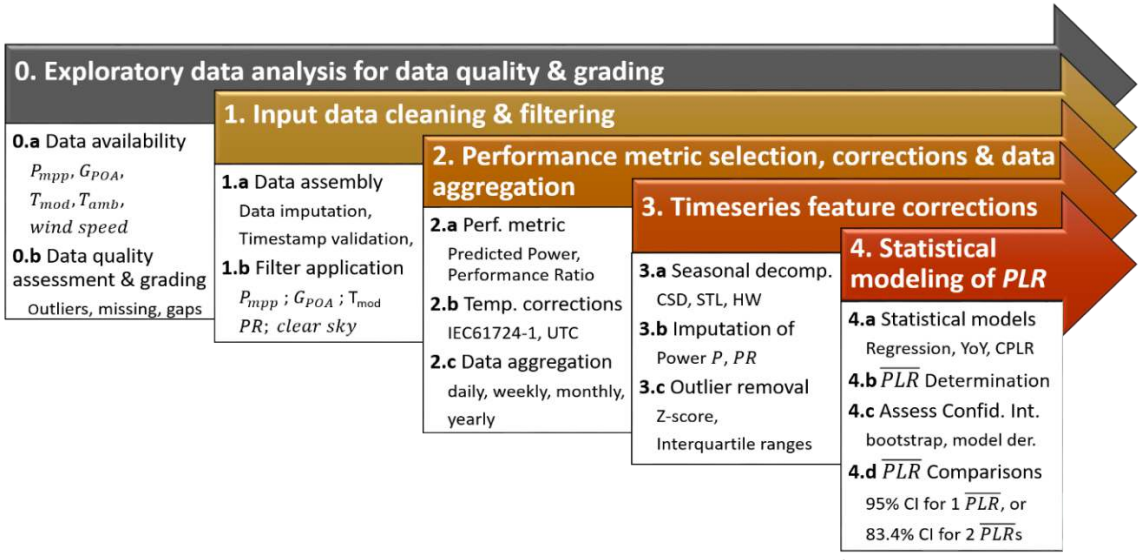


Figure 1: General *PLR* calculation steps using time series data[1].

First, we have to understand which data are available and the format conditions of our raw data. A quality check of the measured data is always recommended and this exploratory data analysis will ensure a smooth application of the steps to follow. To characterize the quality of time series datasets, we use exploratory data analysis, for example visualizing the power dataset as a heatmap (Figure 2) to assess the dataset for outliers, missing data points, and larger gaps in the data and then use a grading scheme to document this information (as discussed in Section 3.4.1 and Section 0). Next, we apply filters to extract the essence of our data. This step is performed to remove anomalous points, measurement errors and non-representative 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 satellite-based weather data. At this point a performance metric has to be selected to account for the instantaneous operating conditions of the systems, most notably irradiance and temperature. These metrics are usually performance ratios (*PR*) but also empirically defined metrics like power predictive models. Correcting for temperature is not required but in most cases suggested. The correction attenuates seasonal variations of the chosen metric. Either the measured or modelled module temperature can be used. If the module temperature has to be modelled, the choice of the model will depend on available climatic input data. Popular representatives are the nominal operating cell temperature (NOCT)[2], the Sandia Photovoltaic Array Performance Model[3], or the weighted moving average temperature model[4]. The temperature correction of the *PR* should be performed according to standard IEC 61724-1:2017[5]. Additionally, the data will be aggregated to a desired time interval, which is usually days, weeks, months, or years. After this step you will be left with a metric of power that is theoretically independent of the variations in weather conditions through time. Quantified performance loss is extracted from the trend between this metric and time.

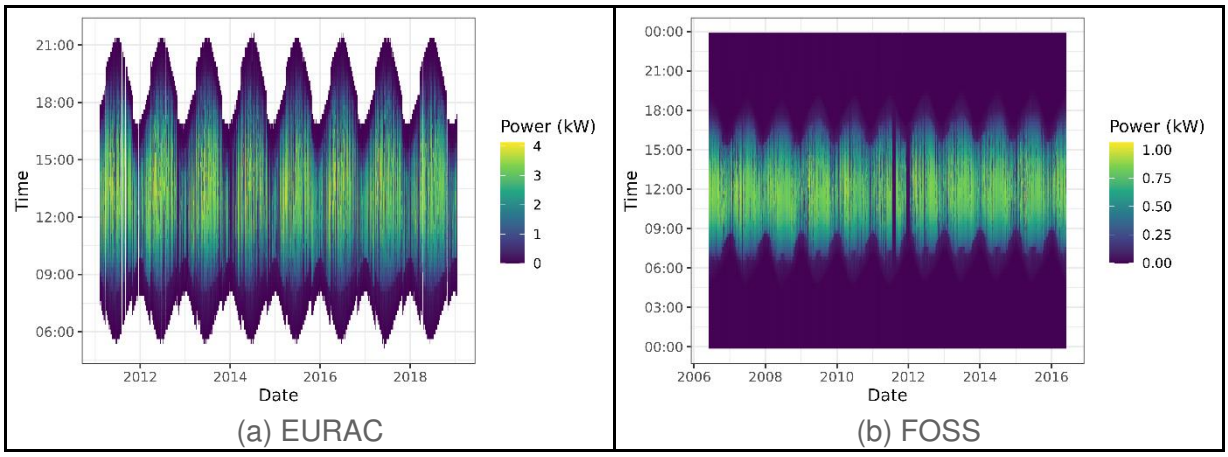


Figure 2: Power heatmap of (a) the EURAC PV System, and (b) the FOSS PV System.

The last step involves the application of a statistical or empirical methodology to receive your systems final *PLR*. Currently, there are numerous methodologies in the literature to choose from. A comparative study of methodologies found in the literature has been performed by Phinikarides[6] et al., and by Lindig et al.[7].

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

Equation 1. $$PLR_{rel.}\left[\frac{\%}{a}\right] = \left(\beta_1 \frac{t}{\beta_0}\right) * 100$$

and has units of %/annum (or %/a). The absolute *PLR* is calculated by:

Equation 2. $PLR_{abs.}\left[\frac{}{a}\right] = \beta_1 t$.

In these equations for the performance loss over time, β_i are the coefficients of the linear additive model's terms between time and the chosen metric (predicted power or *PR*), with β_1 being the coefficient of the slope (units of watts/time for predicted power or 1/time for *PR*) of the line and β_0 being the y-intercept of the model for the *PLR* calculation (units of watts for predicted power and unitless for *PR*). t is a scaling parameter to convert the time scale at which power or *PR* is observed to a yearly scale, as *PLR* is reported 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, and the units are those of the chosen metric/annum. The absolute *PLR* gives an indication of the absolute power loss rate (i.e. watts or *PR* loss, per year) but it is important that the fitting parameter β_0 is also given[8]. 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, correcting for plant size and initial performance and making results comparable between different systems. In the course of this work, the calculated *PLR* refers to the relative Performance Loss Rate.

1.2 Data imputation, filtering and correction approaches

In the 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 exist. A recent study by Livera et al.[9] 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[10,11] is recommended for missing power measurements, multiple imputation by predictive mean matching for missing irradiance measurements[12] and the Sandia module temperature model[11] for temperature measurements. In another study by Lindig et al.[13] 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 as well as machine learning-based models.

Filtering serves to identify and remove data within the time series that are influenced by factors that cannot be modelled[14,15]. 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, shading/soiling/snow coverage or inconsistent irradiance across modules, 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.

1.2.1 Irradiance threshold

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 night time 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 was subset to a high irradiance level, typically 800W/m^2 and above to maintain conditions similar to STC. This has become less popular recently given the massive amount of data removal that occurs from such filters, as opposed to low irradiance cutoffs which keep more of the operational data. Low irradiance cutoffs are generally around 100-200 W/m^2.[24,49]

1.2.2 Power threshold

Power thresholding and irradiance thresholding have strong overlap with each other, given their fundamental link in PV systems. Removing low irradiance values will also remove low power values and vice versa, however power thresholding can still target some specific features that irradiance thresholding cannot. 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.

1.2.3 Inverter saturation & 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. 99 % is commonly used but other values can be applied if needed for different datasets[16]. 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.

1.2.4 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. In keeping with the trend of previous filters which remove features that cannot be captured well by models, 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 and the pyranometer under periods of 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 et al.[17] available in PVlib[18], and a clear sky index (CSI), used by NREL in RdTools[19]. 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 also used a comparison between sensor and 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. Of these two methods the 5 criteria method is stricter 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.

1.2.5 Influence of filtering on *PLR* analysis

A comparison of *PLR* values of the same systems calculated with different power correction models and filter criteria showed that *PLR* magnitude and uncertainty shows a dependence on filtering[24].

1.2.6 Shading, soiling and snow corrections

Shading, soiling, and snow coverage may refer to events 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 performance ratios, soiling events will produce lower performance ratio values than regular operating periods and can be filtered out[20]. Other methods can also be applied which use power and irradiance trends and clustering to detect and remove data influenced by soiling[21]. Shading, soiling, and snow can vary greatly between systems and it is recommended that PV analysts should view power corrected time-series (performance ratio, weather regression, etc.) of their systems to identify any potential areas of concern.

1.2.7 Performance metric IQR filters

P and *PR* are the most common performance metrics used. *PR* is a unit-less parameter, which describes the relationship between incoming irradiation and produced energy 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 non-realistic power-irradiance pairs created through sensor shadowing, alignment or other issues. Usually, statistical thresholds based on interquartile ranges around the median or mode[22] of the performance metric values are used to filter irradiance and power data.

1.2.8 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.

Table 1: Chosen filter and aggregation steps.

#	Models	Metric	Filter				Aggregation
			Irradiance [W/m^2]	Module temperature [°C]	Power	PR	
1	**STL1, YoY1**	PR_{Tcorr}	500-1200	-40 – 100	(0.01 – 1.2) *P_{nom}	±2 standard deviations around monthly *PR* mode	Monthly
2	**STL3, STL4, STL5, STL6, LS-LR4, LS-LR5, LS-LR-6, LS-LR7, STLYoY1, YoY4**	6K PVUSA XbX XbX + UTC	>100		(0.01 – 1.2) *P_{nom}	1.5x inter quartile range	Monthly
3	**VAR1**	*P*	350-850			±2 standard deviations around instantaneous *PR*	Daily then yearly
4	**R-LR1, LS-LR1**	*P*	800-1000	5k bin containing largest share of data points		±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 – 100	0.01W-(98th percentile of P_{ac}*0.99)	*PR* > 0	Monthly
7	**YbY1 or YbY2**	*P*	780-820, 980-1020	18-22 23-27			Yearly
8	**LS-LR8, CSD3, STL7, STL8, HW1, Prophet1**	*PR*	50-1300		(0.1-1.3) *P_{nom}	±3 standard deviations around monthly *PR* mode	Monthly
9	**YoY2**	PVWatts	200-1200	-50-110	*P* > 0		Daily
10	**YoY5**	*PR*	100-1000			1.5x inter quartile range	Daily
11	**SCSF1**	*P*	Strict clear-sky filter				Daily

1.3 Metrics

A metric is a certain measure which 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.

1.3.1 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.

1.3.2 Performance ratio (PR) models

The performance ratio at the DC side is calculated by **Equation 3**[5]:

$$\textbf{Equation 3.} \quad PR_{DC} = \frac{Y_a}{Y_{ref}} = \frac{E/P_{nom}}{H_{POA}/G_{STC}} .$$

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

The *PR* can be corrected for temperature using temperature coefficients as provided by the manufacturers (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 IEC standard procedures. 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[23].

1.3.3 Predicted power models

Generally, a power prediction model is built to predict power as a function of weather over a period of time, then standard or representative weather conditions are applied to all models. This produces a predicted power value, at the given conditions, that is in theory independent of weather. Four Predicted Power models are used in this study to compare the effects of the subsequent time-series they produce on the *PLR* determined. The models are described in detail[24]. Here only a few details are given:

XbX: The XbX model, is a data-driven, multiple regression predictive model[25] with an irradiance (*G*) and a temperature (*T*) term (**Equation 4**) and β_i are this model's coefficients, while ϵ Is the residual error between the model and the data. The flexibility of this model enables non-linear, change point *PLR* and allows for either Plane of Array (*POA*) or Global Horizontal Irradiance (*GHI*) to be used in the irradiance term (*G*) and air or module temperature in the temperature term (*T*).

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

The X in the name refers to a given time step the power prediction model is built over; 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 step is chosen based on the condition of the data being modeled, and what modeling will be performed on the overall dataset.

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 versus 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. Temperature correction coefficients are provided with a given module by the manufacturer, however they can also be

obtained from the time-series data to better reflect the actual outdoor performance of the module.

Equation 5. $P_{cor} = P_{obs}/(1 + \gamma_T(T_{obs} - T_{rep})(G_{obs}/G_{rep}))$, $P_{cor} = \beta_0 + \beta_1 G + \epsilon$

Here, data are subject to a high irradiance G_{rep} of 900 W/m² and the slope of the irradiance over temperature becomes γ_T. In **Equation 5**, *obs* represents observed or measured values and T_{rep} is a representative temperature.

This method is most similar to a temperature corrected performance ratio used in other *PLR* tools such as RdTools[26,27], but structured as a Predicted Power model for better comparison with other models.

PVUSA: The well-known PVUSA model[28] is physics based and described by **Equation 6**:

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

Here, T_{amb} is the ambient temperature [°C], and WS the wind speed [m/s].

The assumption of the model is that the current of a solar panel is a function of the irradiance G_{POA} and the voltage is a function of the irradiance G_{POA} and the module temperature, which is predicted by the ambient temperature T_{amb} and the wind speed *WS*.

6K: The 6K model[29] is the most complicated Predicted Power model used in this study and is summarized in **Equation 7**, **8** and **9**. The name "6K" refers to the coefficients fit by the model.

Equation 7. $G' = G_{POA}/G_{STC}$

Equation 8. $T' = T_{mod} - T_{STC}$

Equation 9. $P = G'(P_{nom} + k_1 ln(G') + k_2 ln(G')^2 + k_3 T' + k_4 T' ln(G') + k_5 T' ln(G')^2 + k_6 T'^2)$

This model uses POA 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 a nameplate power input (P_{nom}) and will always predict P_{nom} at STC conditions.

PVWatts: This simple Predicted Power model (**Equation 10**) follows the irradiance and temperature scaling approach of PVWatts[30] as implemented in the PVLib Python software package[31].

Equation 10. $P = G_{POA}/(1000 * P_{nom})\ (1 + \gamma_T(T_{mod} - 25°C))$

1.4 Statistical methods

Finally, a statistical methodology is applied to compute the *PLR*, given in percentage per year. The methodologies applied in this paper are:

1.4.1 Linear regression (LR)

PLR is commonly assumed as linear, where a single *PLR* value is representative of the entire lifetime of a system. Alternately, non-linear *PLR* methods[23,32] 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 modeling. For regression determined *PLR*, the slope and intercept of the trend directly relates 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[33]. 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 ordinary least squares regression, while being more computationally demanding[34].

1.4.2 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[35]. The step of the seasonal period depends on the data resolution and is usually set to 12 for monthly data. In this case, six 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[36].

1.4.3 Seasonal and trend decomposition using Loess (STL)

The idea behind Seasonal and Trend Decomposition Procedure Based on Locally Weighted Regression (Loess), commonly referred to as STL, is to decompose the *PR* or predicted power time-series into a seasonal part, a remainder and a trend using locally weighted, non-parametric regression[35]. The trend is a nonlinear curve[37], and STL functions are available in R in both the base R stats package and the STL-Plus package[38,39]. 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[7].

STL serves to highlight another important consideration in defining a robust methodology for *PLR* determination, even a single statistical method can give different results, depending on the programming language (R or Python) and the specific implementation. STL was first developed by W. S. Cleveland in 1979[40], 1988[41] and 1990[37]. In 2010 a PhD student of Cleveland's, Ryan Hafen, in his PhD thesis research developed and published the stlplus R package[39]. Loess is non-parametric regression, which is more complex than simple regression. This case of one statistical method demonstrates that to define a robust standard method of *PLR* determination, even a single statistical method, can have varying performance, depending on its implementation. For example we tend to find the best performance from the STL function implemented in the stlplus R package because it is capable of handling more diverse data quality issues successfully when it is applied.

Just as the dataset, the filtering and statistical methods must be defined, even the implementation and coding language of a statistical method can lead to differences in results. In this benchmarking study, STL7 and STL8, were performed using the Python programming language and follow the exact same approach including filtering, metric and STL time series decomposition. The only difference is that STL7 uses STL ported from the STL function in the base R stats package[42] to Python as the rstl package[43], while STL8 uses a STL implementation developed in Python's statsmodels package[44,45]. The stlplus package is currently not ported or available in Python. These two Python implementations of STL, appear to perform differently on the real datasets we are studying here, for reasons that are not currently clear.

1.4.4 Year on year (YoY)

The YoY approach for *PLR* determination was first applied by Hasselbrink[46], and is now available in the RdTools package[19] in Python and the PVplr package in R[21]. In YoY the differences between one data-point in a calendar year with the data-point 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[47,48].

1.4.5 VAR method

The VAR method gives degradation rates from one year to the next, and then by averaging the annual degradation rates we get the *PLR* 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[49]. After filtering, aggregating and transforming the data, it fits a regression model, $\Delta P = f(\Delta G, \Delta T_{Amb}) + d$, 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 one year to the next and by averaging the annual degradation rates we get the *PLR* of a system.

1.4.6 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*.

1.4.7 Statistical clear sky fitting (SCSF)

The SCSF method fits a constrained, non-parametric clear sky model to the data[50]. This model is adaptive and can model sites with complex shade patterns, 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 sub-daily, 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[51]. This approach is unique in that no other information or data is required besides measured power—no irradiance data, no 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 as well as 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.

1.4.8 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 determine how fast the exponential weights decline over the past observations[52,35].

1.4.9 Prophet (Prophet)

The Prophet R package is for forecasting time series datasets using four parameters, namely trend, seasonality, holiday and error[53,54]. The holiday term is used in business applications and is omitted in 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. Prophet 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.[55,23].

1.4.10 Piecewise linear, change point *PLR*

Piecewise linear *PLR* combines the ease of use and interpretability of other regression methods but does not use a linear assumption. It is instead able to quantify non-linear trends in PV time-series. This functionality is available in the R package, PVplr[21]. Instead of a single slope being representative of the entire trend, the piecewise linear *PLR* uses piecewise regression to identify the change point location[56] and then divides the time-series into two (or more) separate linear trends, each with its own *PLR* magnitude and uncertainty. The *PLR* values of the individual trends have all the same interpretability of an assumed linear *PLR*, but non-linear behavior can be observed in the differences in *PLR* between segments.

1.5 Combinations of performance metrics & *PLR* calculation models

Table 2 presents an overview of the metric - statistical method combinations used in this study.

Table 2: Combinations of metrics together with statistical models.

	LR	STL	YOY	VAR	CSD	YbY	SCSF	HW	Prophet
PR	LS-LR2 LS-LR8	STL7 STL8			CSD1 CSD3			HW1	Prophet1
PR_{Tcorr}	LS-LR3	STL1 STL2	YoY1 YoY3		CSD2				
XbX	LS-LR4	STL3							
XbX + UTC	LS-LR5	STL4	YoY4 STLYoY1						
PVUSA	LS-LR6	STL5							
6K	LS-LR7	STL6							
PVWatts			YoY2						
Power	R-LR1 LS-LR1			VAR1		YbY1 YbY2	SCSF1		

2 DESCRIPTION OF *PLR* BENCHMARKING DATASETS

2.1 Data characteristics: time interval, time length, data types

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

- Time interval - The time interval of collected time series data. Typical values range from 1-15 minutes but can vary between 30 seconds to one or more hours, depending on hardware used or user settings. Typically, high resolution data with a low time interval is desired as it gives a more complete look at the performance of a system and a larger amount of data improves model fitting.
- 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, however waiting is the only option for extending time lengths. Typically, systems with less than two years of data have difficulties in *PLR* measurements as seasonal trends in the data are difficult to account for in less time. For reliable PLR evaluation, at least a 3- or 5-year time series should be available[57,58].
- 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 of the inverter, 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, air mass, etc. can all be used in certain types of *PLR* analysis, depending on availability. Lastly, metadata are also important; e.g. module and inverter characteristics, location, scale, orientation etc.
- 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 for some of them. However, *PLR* cannot be calculated reliably in datasets 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 Systems with monthly power only

2.2.1 IEA-PVPS Task 13 database

Figure 3 shows the spatial distribution of the total number of PV systems, which are collected within the IEA PVPS Task 13 database, and the ones which are usable for calculations[59]. One necessary requirement for a PV system to be included in the study is the availability of a time interval of at least 24 months, whereby a small amount of missing months can be approximated

using a rolling mean/moving average. From 173 systems in total, data of 120 were usable for this study (Figure 4). Unfortunately, the database has not been updated since 2016 and it is relatively incomplete.

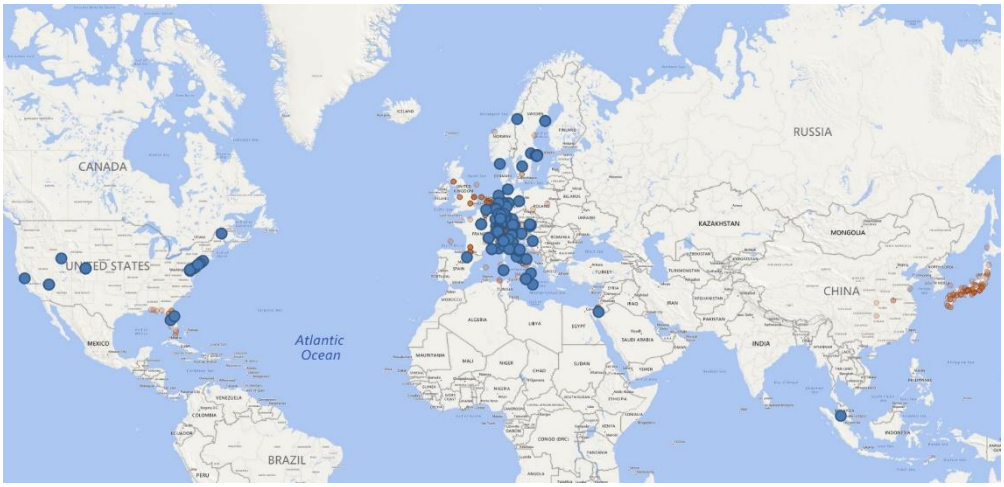


Figure 3: PV systems within Task 13 database: red transparent dots represent all systems available; blue dots represent all usable systems[59].

Little is known about the state of the plants. Several PV parameters such as the power, the performance ratio and others are given as well as the location of the plants. There is no information concerning the history of the plants. No data are provided on how thorough the systems are monitored, if and which degradation modes and other performance reducing effects took place and if countermeasures are carried out. For example, downtime due to inverter issues could negatively affect the *PR*, which is the primary parameter for the calculations, and would artificially influence the results.

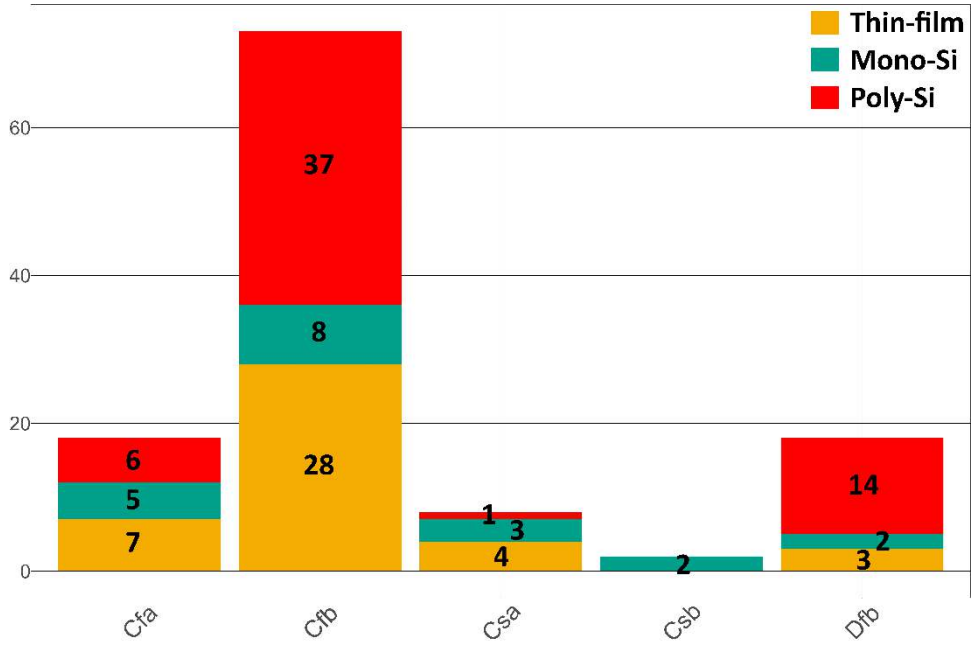


Figure 4: PV systems in the IEA-PVPS Task 13 database, as a function of their Köppen-Geiger climate zone and technology[59].

Table 3. Metadata summary of the PV systems analyzed in this benchmarking study.

Dataset	Tech-nology	Coun-try	P_{nom} (kWp)	Time Period		Azi-muth	Tilt	Measurement available	KGC† [63,60]	KGPV‡ [64,65]
EURAC	pc-Si	Italy	4.20	02/11 - 01/19	8 yrs.	188.5°	30°	G_{POA}, T_{amb}, WS, T_{mod}, P_{DC}, P_{AC}	Dfb, ET	EM, DM
FOSS	mc-Si	Cy-prus	1.03	06/06 - 05/16	10 yrs.	180°	27.5°	G_{POA}, T_{amb}, WS, T_{mod}, P_{DC}	Csa, BSk	DH, CH
RSE CdTe	CdTe	Italy	1.16	06/09 - 12/18	9.5 yrs.	180°	30°	G_{POA}, T_{amb}, P_{AC}	Cfa, Cfb	DM
RSE pc-Si	pc-Si	Italy	1.68	06/09 - 12/18	9.5 yrs.	180°	30°	G_{POA}, T_{amb}, P_{AC}	Cfa, Cfb	DM
Pfaff-staetten A*	pc-Si	Austria	2.11	01/13 - 04/19	6.25 yrs.	220°	22°	G_{POA}, T_{amb}, T_{mod}, P_{DC}, P_{AC}	Cfb, Dfb	DM
Pfaff-staetten B*	pc-Si	Austria	2.06	01/13 - 04/19	6.25 yrs.	220°	22°	G_{POA}, T_{amb}, T_{mod}, P_{DC}, P_{AC}	Cfb, Dfb	DM
Pfaff-staetten C*	CIGS	Austria	2.25	01/13 - 04/19	6.25 yrs.	220°	22°	G_{POA}, T_{amb}, T_{mod}, P_{DC}, P_{AC}	Cfb, Dfb	DM
US DOE c10hov6	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	BSk, Cfb	CK, DH
US DOE kodpi8	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	BWk, BWh	BK, CK
US DOE luemkoy	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	Dfb	EM
US DOE lwcb907	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	Dfb	EM
US DOE t3pg1sv	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	BSk, Cfb	CK, DH
US DOE wca0c5m	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	30°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	Cfa	DH
US DOE wxzsjaf	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	35°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	BWk, BWh	BK, CK
US DOE z0aygry	mc-Si	USA	3.24	11/15 - 05/18	2.5 yrs.	180°	30°	G_{POA}, T_{amb}, T_{mod}, P_{DC}	Cfa	DH
NREL1	mc-Si	USA	2.70	05/16 - 07/19	3.25 yrs.	180°	30°	G_{POA}, T_{amb}, WS, T_{mod}, P_{DC}, P_{AC}	Dfb, BSk	DH, CH
NREL2	HIT	USA	1.00	08/07 - 12/16	9.25 yrs.	180°	40°	G_{POA}, T_{amb}, T_{mod}, P_{DC}, P_{AC}	Dfb, BSk	DH, CH
NREL3	mc-Si	USA	94.00	09/09 - 01/18	8.25 yrs.	175°	10°	G_{POA}, T_{amb}, WS, T_{mod}, P_{AC}	Dfb, BSk	DH, CH
NREL4	mc-Si	USA	524.00	07/11 - 05/18	6.75 yrs.	165°	9.1°	G_{POA}, T_{amb}, WS, P_{DC}, P_{AC}	Dfb, BSk	DH, CH
4 Digital power plants**	c-Si	France	1.82 each		5 yrs.	180°	21°	G_{POA}, T_{amb}, WS, P_{DC}, P_{AC}	Cfb	DM

* Second hand Modules;
** Irradiance data from HelioClim for Rennes/France.
† A-Tropical, B-Arid, C-Temperate, D-Continental, and 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
‡ A-Tropical, B-Desert, C-Steppe, D-Temperate, E-Cold and F-Polar climates;
L-Low, M-Medium, H-High and K-Very High irradiation zones

Using longitude and latitude, which are provided for each dataset, it is possible to organize the plants by Köppen-Geiger classification[60,61,62,63] visible in Figure 4. It is visible that the gross

amount of systems is located in a temperate climate without dry seasons since more than 50 % of the systems are located in Germany and Italy. Most plants are either made out of poly-crystalline silicon or thin-film materials. No information is provided which thin-film material is used.

2.3 Systems with high quality time series power & weather data

19 datasets were made available to the interlab benchmarking participants. An overview together with the most important metadata information on these PV Systems is summarized in Table 3. The climate zone categorization used in Table 3 is based on the well-known Köppen-Geiger classification[60,63]. We also include the new KGPV classification which includes irradiance intensity (L-K) but has reduced weather distinctions (A-F)[64,65].

In the following, the PV systems are briefly introduced and Figure 5 shows their geographical distribution. All datasets are publicly available at doi: 10.17605/OSF.IO/VTR2S[66].

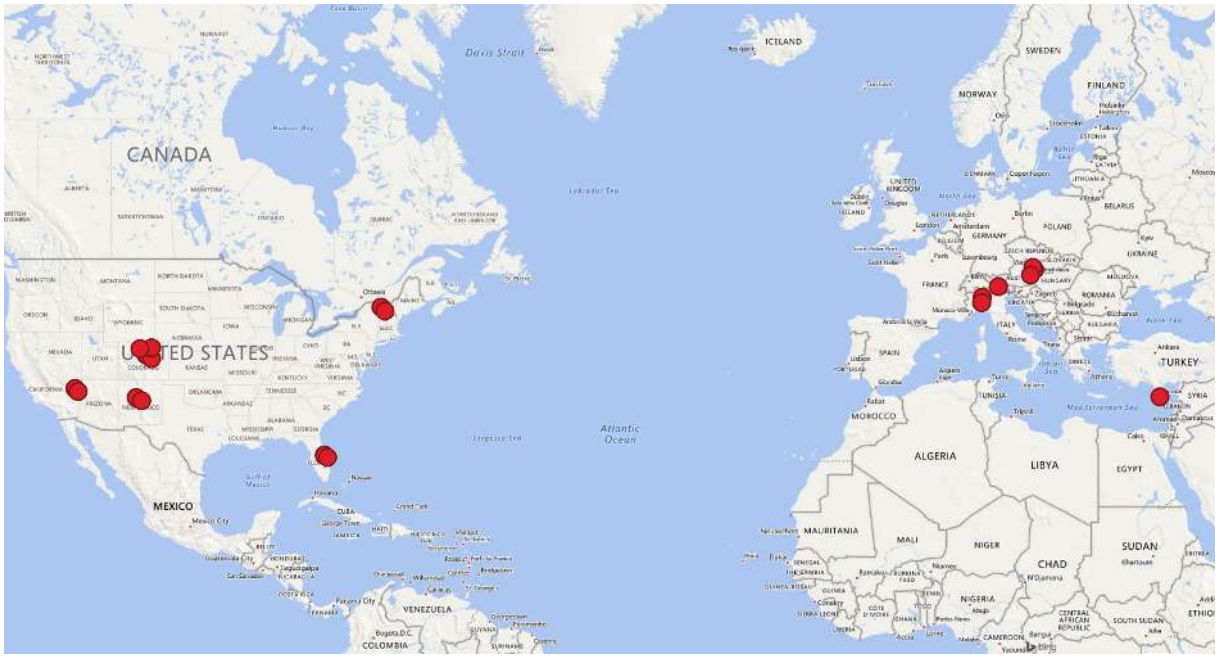


Figure 5: The locations of the PV Systems used for the benchmarking exercise[1].

2.3.1 EURAC PV system

The EURAC PV System was installed at the airport of Bolzano/Italy (ABD) in 2010. The poly-crystalline 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 plane of array 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[5]. The weather data are recorded with a time interval of one minute. Since the electrical measurements are taken at time intervals of 15 minutes, all values are averaged to the same time interval. A period of eight years of power data, shown in the heatmap of Figure 2a 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 six months before the observation time starts. The delayed start of observation was set to exclude initial degradation effects.

2.3.2 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 semi-arid. The PV system dataset used in this investigation, (Figure 2b) is obtained from a ground-mounted mono-crystalline 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 weather data and operational measurements were acquired and stored in a database. 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 61724-1:2017 standard[5], and stored with the use of a robust measurement monitoring system. The monitoring system records plane-of-array 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.

2.3.3 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 report (Figure 6) refers to two PV power plants, respectively with c-Si and CdTe technology. Both systems started to operate in June 2009. The c-Si (polycrystalline silicon) PV plant is a ground-mounted PV plant, full south orientation, and tilt of 30°. The PV plant has a nominal power of 1.61 kW, constituted by a string of 8 PV modules of 210 W. The CdTe (Cadmium telluride) PV plant is a ground-mounted PV plant, full south orientation, and tilt of 30°. The PV plant has a nominal power of 1.16 kW, constituted by 4 string of 4 PV modules of 72.50 W. A weather station is installed close to the test site and allows the acquisition of irradiance parameters (plane-of-array) and air temperature. According to the IEC61724-1:2017[5], sensors are periodically cleaned and calibrated. Operational data are acquired every 10 seconds and sent to the remote unit which stores them as mean or integral values (1 and 15 minutes intervals).

The PV System Pfaffstaetten is a 5 kWp rooftop system, running from 01/2013 until 04/2019. The system has three strings, two with 2nd hand poly-crystalline modules (110 and 120 Wp, dating back to the end of the nineties), and the third with CIGS modules (150 Wp).

The inverter has three separate MPP trackers, and these are connected to

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

The monitoring data is 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 is recorded, as can be seen in Figure 7.

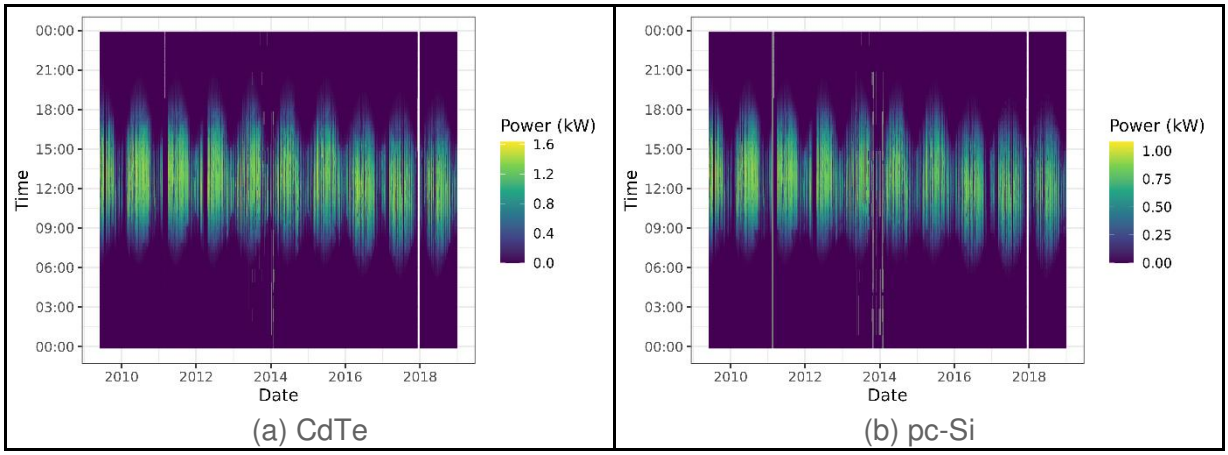


Figure 6: Power heatmap of (a) the RSE CdTe PV System, and (b) the RSE pc-Si PV System.

2.3.4 Pfaffstaetten PV systems

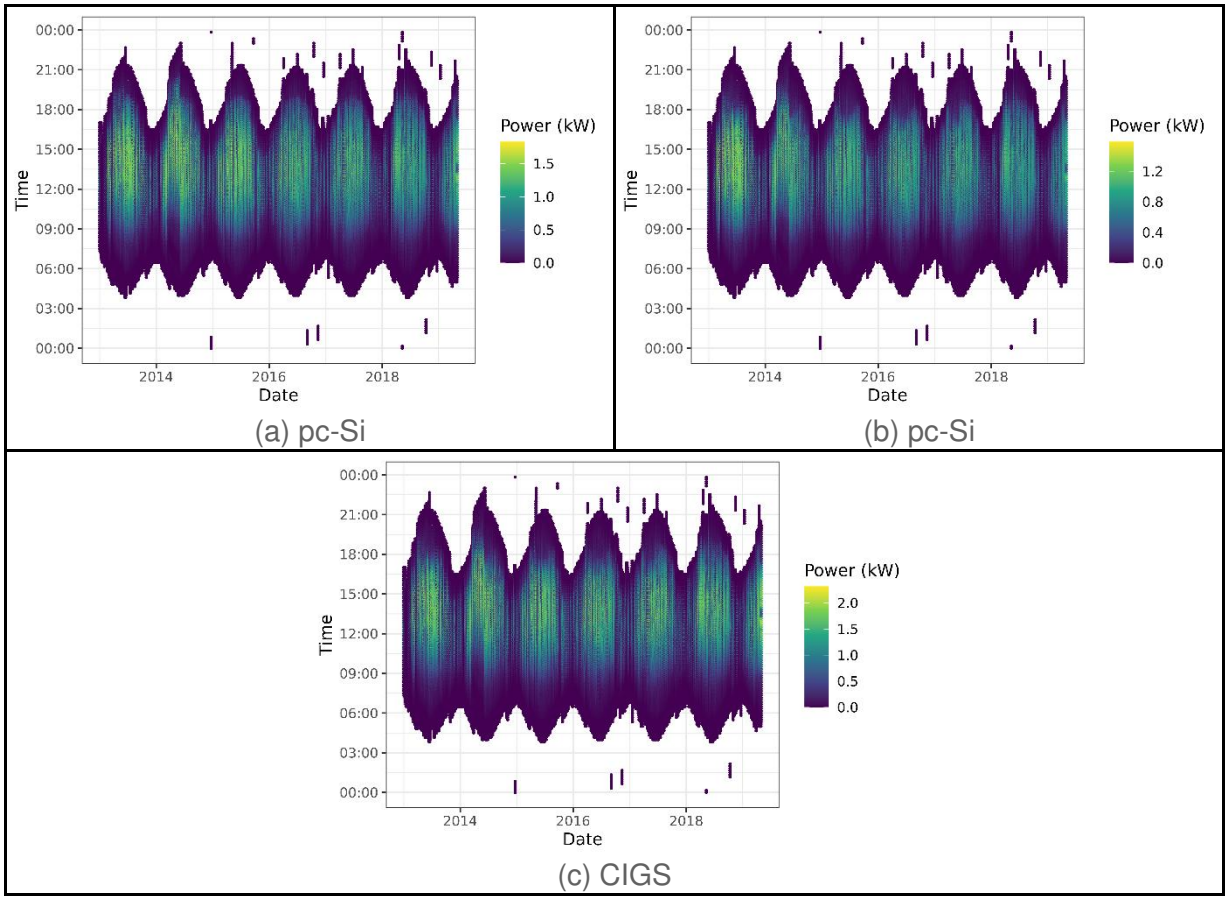


Figure 7: Power heatmap of (a) the Pfaffstaetten A PV System, and (b) the Pfaffstaetten B PV System and (c) the Pfaffstaetten C PV System.

Irradiance values are based on a c-Si reference cell as irradiance sensor, and the ambient temperature sensor is integrated in the case underneath, so the ambient temperature readings

are not the air temperature (in the shadow), but these inner case temperatures in the sun, and therefore follow more or less the temperature of the one module temperature sensor attached on the back side of one of the Kyocera modules.

There are no additional temperature measurements for the MiaSolé CIGS glass/glass modules, which may operate at slightly higher temperatures.

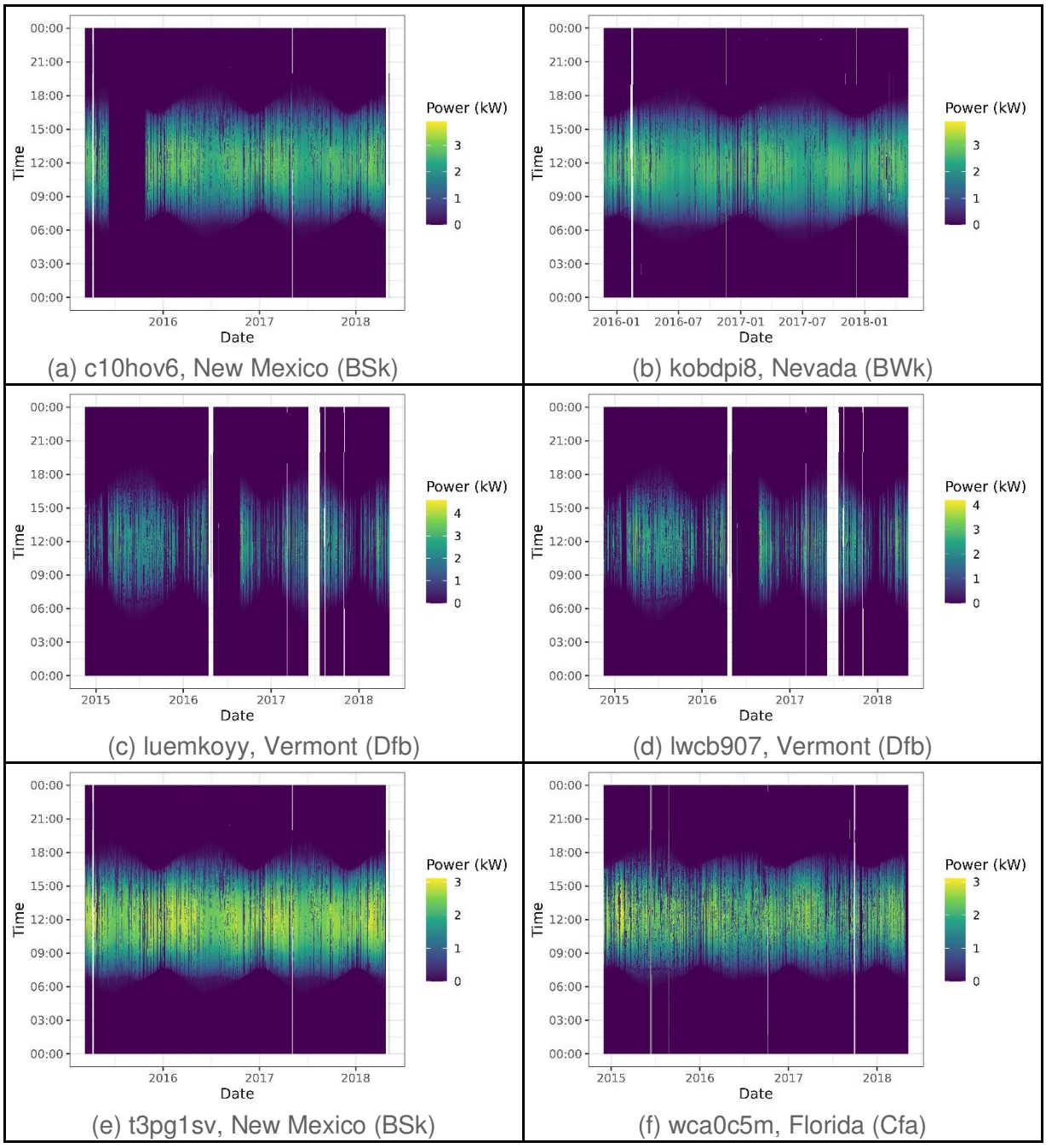
2.3.5 US DOE RTC baseline systems

The US Dept. of Energy Regional Test Center Project[67] has five sites. At each site there are 2 systems that are used as the “Baseline Systems” and the data from these is publically available[68]. These publically available, “open” datasets are currently version 0.2, and consists of a series of 8 identical PV systems, 3.24 kW strings with the same module and inverter, in 4 locations/climate zones. The system Locations and their Köppen-Geiger climate zones[62] include Nevada (BWk), New Mexico (BSk), Florida (Cfa), Vermont (Dfb). The data is 1 minute time series inverter data with ground and satellite weather data. The systems began logging data in 2016, with an identical system in Nevada starting up about 9 months later (Figure 8). The Vermont system has recently been discontinued but the historical data is still available. 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.

2.3.6 NREL PV systems

Four PV systems from the US National Renewable Energy Laboratory (NREL) were included in the benchmark study (Figure 9), all located at the NREL main campus in Golden, Colorado. Short time interval 1-minute data is collected for three systems, with 15-minute 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 kWp, 3 kWp, with data available from April 2016 until July 2019.

System #2 is also a small research system using a string of 5 silicon heterojunction modules, 1 kWdc, 1.8 kWac, with data availability from August 2007 until December 2016. Both of these small research systems have co-located calibrated broadband pyranometer G_POA irradiance measurements, wind-speed and back-of-module temperature measurement, and are mounted on free-standing open-rack structures.



(a) c10hov6, New Mexico (BSk)

(b) kobdpi8, Nevada (BWk)

(c) luemkoyy, Vermont (Dfb)

(d) lwcb907, Vermont (Dfb)

(e) t3pg1sv, New Mexico (BSk)

(f) wca0c5m, Florida (Cfa)

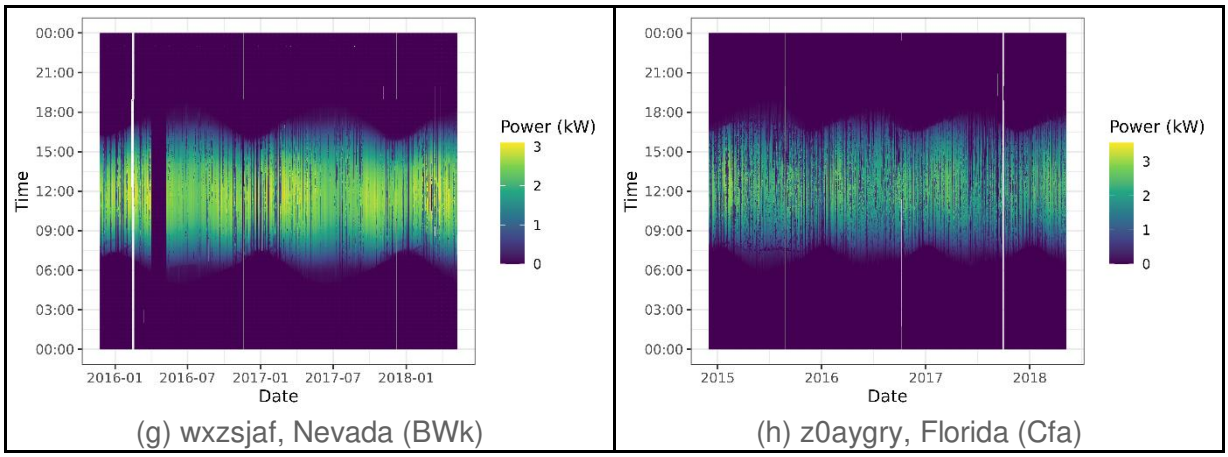


Figure 8: Power heatmap of the eight US DOE RTC PV Systems, labeled (a) to (h) with their 7 digit alphanumeric identifier, location and Köppen-Geiger climate zones.

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

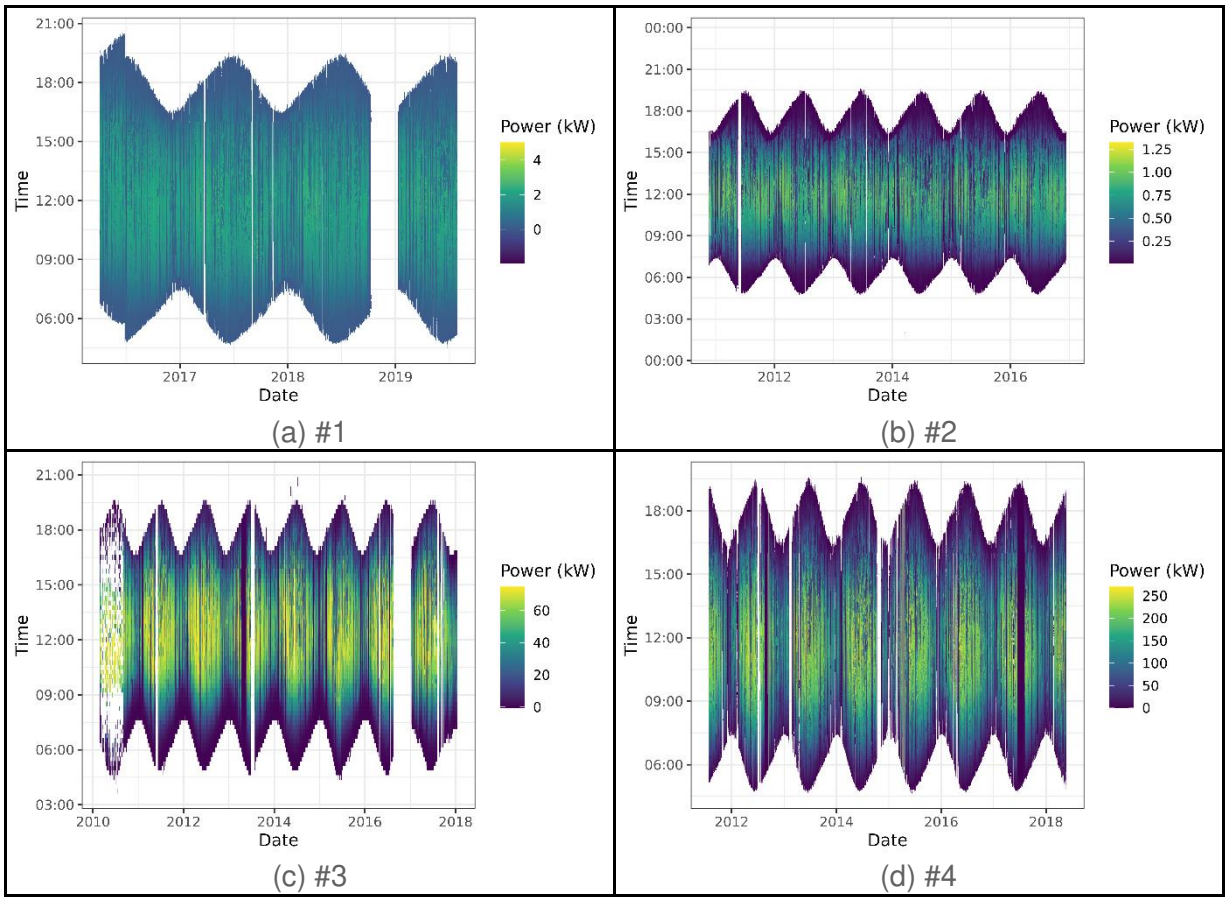


Figure 9: Power heatmap of the NREL PV Systems a) #1, b) #2, c) #3, and d) #4.

2.3.7 EDF digital power plants

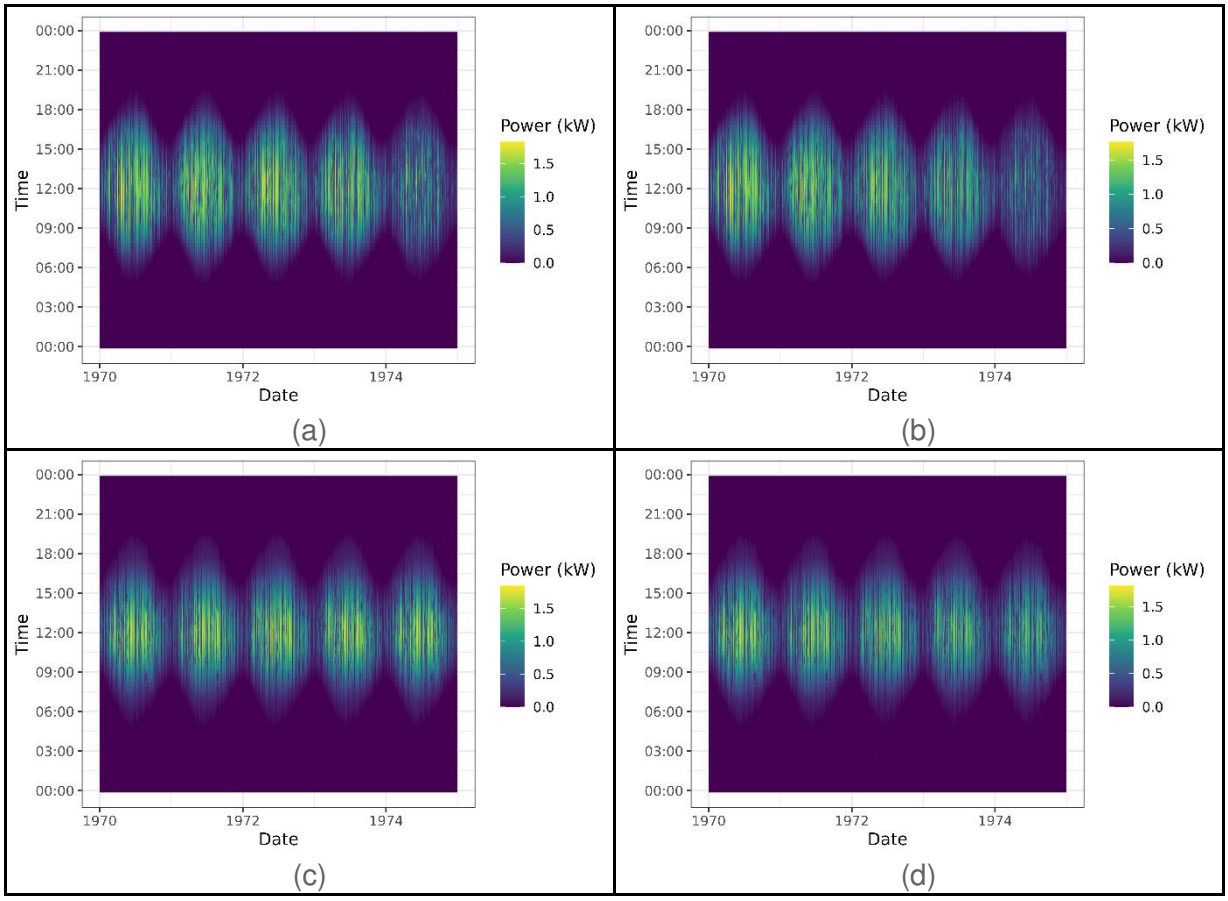


Figure 10: Power heatmap of the EDF Digital Power Plants (a) , (b) , (c) , and (d).

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 Data sheet: 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, Voc = 44.8 V

The string is connected to a 2 kW inverter. The behavior of the plant is simulated with a Dymola/BuildSysPro Software, developed by EDF[69]. The model for the PV modules is a 2 diode model.

Solar data were derived from HelioClim[70] with a temporal resolution of 15min. All in all, four scenarios have been created with the following settings:

- 1 year weather data repeated for 5 years with and without known degradation
- 5 years different weather (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 Amps, variation of -4 %/a), series resistance per cell ($R_{s_{cell}}$, initial value: 0.00012 $\Omega\, m^2$, variation of +7 %/a), shunt resistance per cell ($R_{sh_{cell}}$, initial value : 0.14745 $\Omega\, m^2$, variation of -6 %/a), and the temperature coefficient ($\propto_{I_{sc}}$, initial value : 0.037 %/K, variation of -1.5 %/a). The resistances at the module level, can then be determined

as follows: $R_{s_{mod}} = R_{s_{cell}} * N_s/(N_p * S_{cell})$, roughly 0.6 ohm, and $R_{sh_{mod}} = R_{sh_{cell}} * N_s/(N_p * S_{cell})$, roughly 740 ohms, where N_s and N_p and the number of cells in series and parallel respectively, and S_{cell} is the area of the cells. The values of the parameters change every hour, according to the chosen degradation rate (a decrease for I_{sc}, R_s and γ_T, and an increase for R_s). The theoretical degradation for the simulated dataset with induced power loss was simulated from two viewpoints:

- @STC the P_{mpp} degradation is -4.41 %/a (degradation with a linear variation of parameters: I_{sc}, R_s, R_{sh}, and the temperature coefficient (α) of I_{sc})
- Absolute energy degradation is -4.89 %/a, which has been quantified for the systems with repeating weather data.

The resulting four digital power plant PV systems consist of timestamp, irradiance (G_{POA}), ambient temperature (T_{amb}), wind speed, AC power and DC power, and their DC Power heatmaps are shown in Figure 10.

An additional set of systems was created at a later period by request for the evaluation of multiple low *PLR* values. The creation basis of this additional data is identical to the previous, including repeated weather and the weather with a different final year, however more cases were given across a larger range of *PLR*. In total, 11 cases (ranging from case 0 to case 10) were created with input energy degradation values from 0 %/a to -4.89 %/a (the previous degradation case).

2.4 Systems with higher-order time series data types

Additional data types, such as I-V, Power, Weather, or imaging data, can provide new insights into PV systems performance and degradation.

2.4.1 Time series I-V, P_{mpp} datastreams

The Fraunhofer-ISE dataset[71] contains both time-series P_{mpp} measured at about 1 minutes time interval and time-series current-voltage (I-V) curve data measured at about 5 minutes time interval. The I-V curves are measured by the electronic load ESL-Solar 500[72], with the number of data points in each *I-V* curve varying from 40 to 80. In addition there other variables such as air mass, irradiance, temperature of modules measured and recorded by local sensors at the same time, these variables are very useful for analyzing and modeling the module behavior over time.

The data contains eight commercial PV modules produced by two manufacturers, one brand has a module architecture of glass-backsheet while the other brand are double glass modules. Five out of eight modules started exposure in 2010 while another three started in 2012, four of them are still being traced right now, while others ended exposure earlier, so the system age for these eight modules varies from 3 years to 9 years. They are also located at three different locations and these three locations are classified to be different climates using the kgc package[63] in CRAN. Three of them are located in Gran Canaria (Spain) which is in the BWh climate zone, three of them are located in the Negev desert (Israel) which is in the BSh climate zone and two of them are located in the Bavarian Alps (Germany) which is the ET climate zone.

3 CALCULATION OF *PLR* BY MULTIPLE METHODOLOGIES

This section of the report is divided into eight parts. First, Section 3.1 discusses the steps of calculating *PLR* while Section 3.2 shows a detailed example applied to a high quality time series dataset of a research PV system, the EURAC system. Next in Section 3.3 the ability to calculate *PLR* from low quality, monthly interval, data is demonstrated. In Section 3.4 the statistical characteristics of the 19 PV power time-series datasets are presented. These 19 PV datasets should be considered as predominantly well-tended "research" PV systems, yet they have many anomalous characteristics. And yet we expect these to be of higher "data quality" than commercial PV power plant datasets. Exploratory data analysis allows us to characterize these datasets and a dataset grading schema and tool are presented in Section 3.4.1. The dataset characteristics and grading of the systems used in this benchmarking study are summarized in Section 3.4.2. The contributions to the uncertainty of the *PLR* results are discussed in Section 3.4.3 along with the need for handling uncertainty quantification in regression and YoY methods, so as to arrive at comparable uncertainties. In Section 3.4.4, the calculated *PLR* of the digital plants are investigated. The digital plants used in the benchmarking had large simulated *PLR* values, so an additional study of the impact of *PLR* magnitude on the computed *PLR* of these simulated digital power plants was performed and is addressed in Section 3.4.5. In Section 3.4.6 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. Next the evaluation methodology is presented for how the results based on the real PV power time series datasets and the diverse set of filters and methods are presented in Section 3.4.7. Finally, in Section 3.4.8 benchmarking study, across all filters and modelling methods are compared by quantitative comparison of two groups of results: 1) G7-3-24: Evaluation of 7 systems with 24 approaches and 2) G13-7-17: Evaluation of 13 systems with 17 approaches.

3.1 *PLR* pipeline workflow

Nearly all *PLR* calculation pipelines follow a similar framework of four steps (Figure 1), with the specific application of each step being up to the discretion of the person performing the calculation. A notable exception is the SCSF method[50] which evaluates *PLR* from power data only. The specifics of the data analysis steps and analyst choices are discussed in Section 1 as to how a modeler can make decisions for each step. Here we show examples of how decisions made at each of the four steps of *PLR* determination can influence the final *PLR* result. And unfortunately, a data analyst, knowledgeable of these effects, could bias the results, by rather simple use of filters, anomaly removal, or other factors.

3.2 Example calculation PV system *PLR*

Here, the necessary steps for calculating the *PLR* of a PV system are visualized step-by-step through the analysis of the EURAC PV system dataset. The specifications of the corresponding system can be found in Section 2.3.1 and Table 3.

The steps in *PLR* analysis using STL[37] applied to the monthly temperature-corrected *PR* are shown in Figure 11 . The performance trend, extracted with STL, using the stlplus R package[39], is fit with a simple linear model to determine the assumed linear *PLR*. Additionally, the *PLR* has been evaluated with a one change point, two segment model, to determine the change point position and the PLR of the initial and long-term segments.

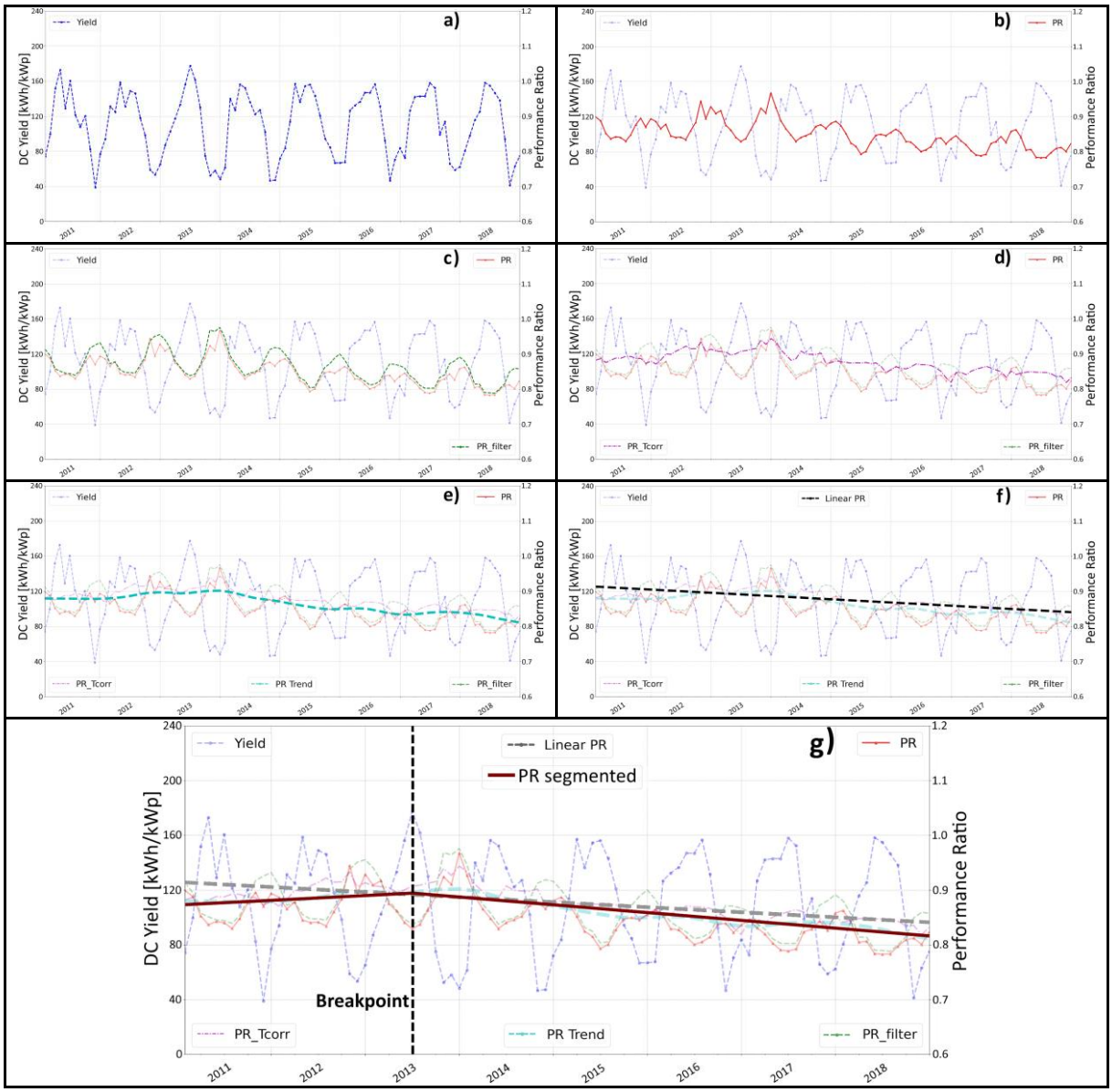


Figure 11: Example calculation steps for retrieving *PLR* value: a) Yield; b) performance ratio (*PR*); c) filtered *PR*; d) temperature corrected performance trend; e) performance trend of temperature corrected *PR*; f) assumed *PLR*; g) change point linear *PLR*[1].

First, the power time series or the yield (the power divided by the installed capacity, in kWh/kWp), is shown in Figure 11a 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 11b, the selected metric (in this case *PR*) is added to the plot. It is shown, that due to the strong temperature dependence of PV modules, especially in crystalline Silicon, 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*, shown in Figure 11c. The applied filters correspond to filter 1 of Table 1 in Section 1.2. It can be seen that the *PR* time series exhibits a sinusoidal shape, which is an indicator of the exclusion of non-representative measurements

or measurement conditions through filtering. Figure 11d shows the effect of temperature correction according to the standard IEC 617214-1:2017[5]. 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 determining the final assumed linear *PLR* was a combination of STL[37] and linear regression. STL is a non-parametric, locally-weighted regression, which extracts a non-linear trend from a dataset by excluding the remaining seasonality and the residuals. This non-linear trend (Figure 11e) can then be fitted with a simple linear model using regression to determine the best fit linear trend line, visible in Figure 11f. The yearly aggregated gradient of the linear function, divided by the intercept, is the final, assumed linear, relative performance loss rate 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 %/a and accounts for one standard deviation.

Additionally, a piecewise linear, change point *PLR* with one change point and two segments is shown in Figure 11g. Here, the segmented non-linear trend is divided into two segments and the change point regression fits the two linear segments and the change point location simultaneously, with the constraint that the two *PLR* segments are continuous. In this case the change point is found to be at 2.5 years, and the first piecewise linear *PLR* is $PLR_1 = 0.95\ \%/a$. The second *PLR* segment has $PLR_2 = -1.49\ \%/a$ for a time span of 5.5 years. Piecewise linear, or change point *PLR* provides a more detailed performance evaluation compared to the simpler assumed linear *PLR* result and is very useful for in-depth characterization activities of PV system performance. For example, many commercial PV systems exhibit stable or slightly increase power over the first 1 or 2 years, which would be the first segment, and then the slow, long-term performance loss of the second segment, begins, and extends over the systems lifetime.

3.3 Low quality data *PLR* results

Here, the performance loss of 120 PV systems included in the PVPS IEA Task13 database was calculated. The systems have been in operation for up to 17 years and are located in the U.S. and Europe. The chosen methodologies for this study are either seasonal-trend decomposition using LOESS (STL) and year-on-year (YoY) statistical modeling to determine *PLR*.

In Figure 12 an overview of the results is given for the two *PLR* methodologies selected. The figure shows the *PLR* distribution for the STL and YoY approach. Both methodologies show similar trends.

The *PLR* peaks at the bin of -1.0 to -0.5 %/a and the distribution is approximately normal considering all 120 systems as we would expect from the central limit theorem for these PV systems[22]. STL delivers a median *PLR* value of -0.71 %/a and YoY a value of -0.63 %/a. Looking at the results, it seems that both methodologies deliver very similar values when applied to monthly data, as we find them here. It became apparent that STL should be used with care if the dataset in question shows a non-seasonal behavior. This non-seasonality can arise from different causes, such as the prevailing climatic conditions, the technology of the system or the application of certain weather dependent corrections. It can be expected that an overestimation of seasonality falsifies the results to a certain degree. Within the database no such cases were found because the climatic zones the systems are located in, have distinct seasons. It is believed that the overall results are relatively accurate.

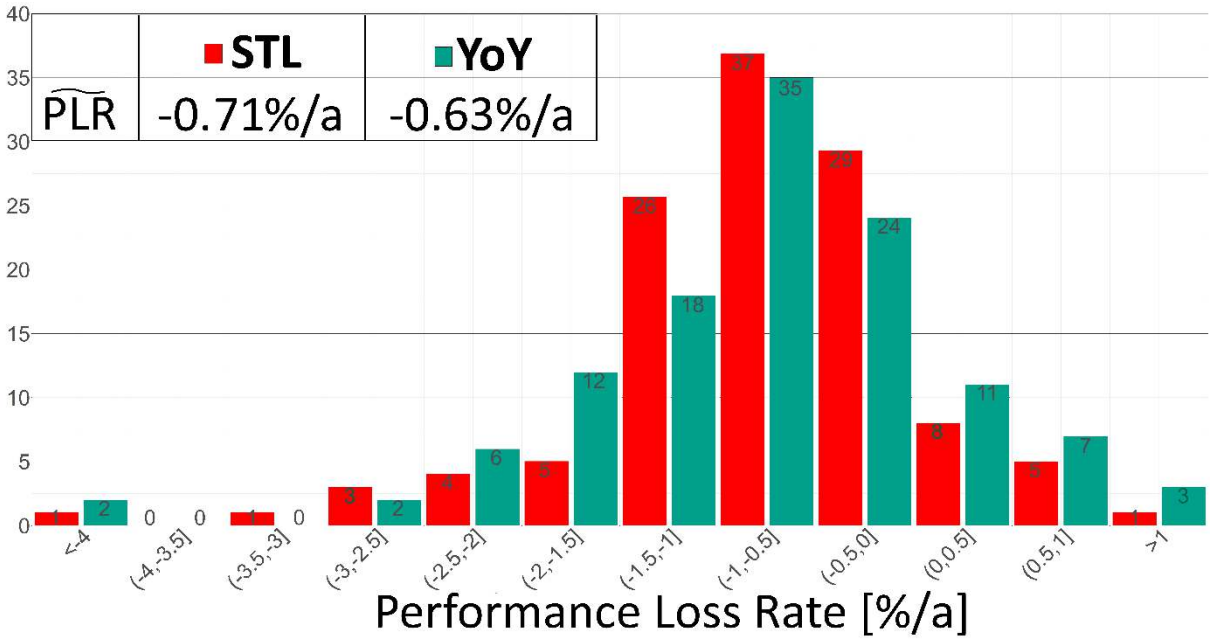


Figure 12: *PLR* distribution of database using STL and the YoY approach[59].

The aforementioned seasonality problem that arose for time-series decomposition is not an issue if YoY is used. It was very interesting to realize that the YoY methodology is highly affected by the amount of digits with which the *PR* is specified. The *PR* is given in the database with two digits, e.g. 0.88. The usage of a *PR* with just two digits results in a very inaccurate representation of the actual value. A high number of *PLR* values from the calculated *PLR* distribution results in a *PLR* of 0.00 %/a. That is why the *PR* was calculated again by dividing the monthly final yield with the reference yield, which are provided in the database as well. Introducing this extra step yields more reasonable final values because the *PR* is reported with greater precision.

Figure 13 shows the *PLR* divided by technology and methodology. Additionally, the amount and the average operational lifetime (median) of the systems is provided. When comparing the trend between the methodologies, again, both methodologies provide similar results. In both cases, mono-crystalline silicon systems (mono-Si) degrade at the slowest rate, while thin-film systems are subject to the highest degradation rate. These results confirm observations of previous studies[73]. On average, the thin-film systems were installed 25-30 months before the crystalline systems. This fact could further result in the observed elevated *PLR* values. The absolute median and mean values, comparing the methodologies, are quite similar, except for the median of the mono-Si systems. Here, YoY provides an almost doubled overall *PLR* compared to STL. One reason might be the relatively small distribution of available systems compared to the other technologies. The distribution of the STL results is non-Gaussian. The $\overline{PLR}$ of the mono-crystalline systems is similar across the methodologies. Due to a stronger left skewness of the distribution, the median is quite low for STL. By comparing to YoY and also considering values from the literature[74], it seems that this value is less trustful.

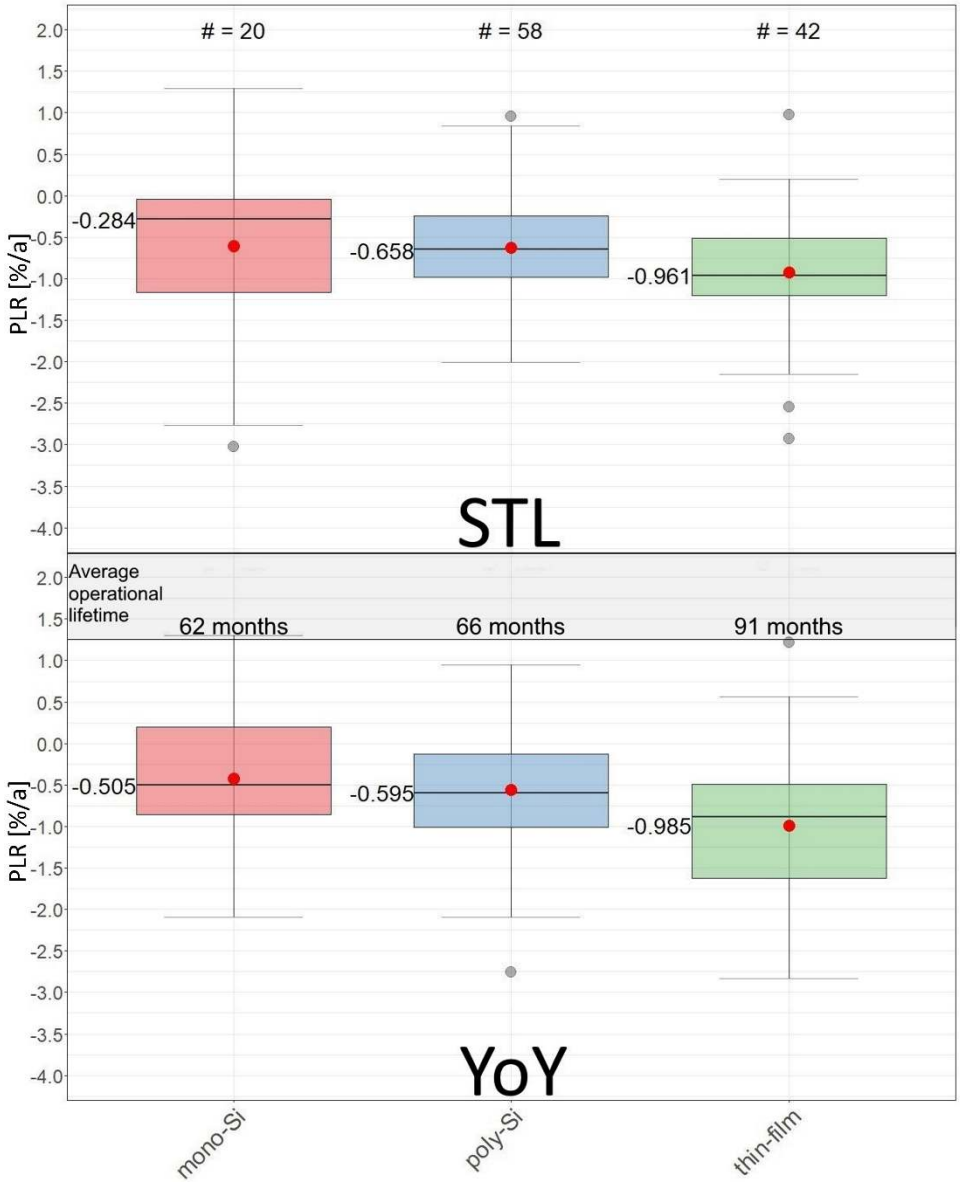


Figure 13: *PLR* divided by technology and methodology; amount and average operational lifetime of systems[59].

It is visible that both *PLR* calculation methodologies, STL as well as YoY, are relatively well suited to calculate the *PLR* of PV systems where low quality data are provided. Under the given circumstances, the application of both methodologies results in relatively reasonable *PLR* rates. The larger the number of studied systems is, the more the *PLR* distributions approach a similar, and more normal, Gaussian distribution.

3.4 High quality data *PLR* results

3.4.1 Exploratory data analysis 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[75],[76]. In addition, these measures can guide the user on the expected uncertainties in comparing among 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[77]. 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 data points, and data gaps. We have developed a grading schema, summarized in Table 4, and the measures and grades are summarized in Table 5.

Table 4: Data quality grading criteria. 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 > two years for a passing "P" grade, otherwise it is graded F.

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

Outliers are typically defined as points which are greater than ±1.5 times the interquartile range (IQR), and these may, or may not, be anomalous datapoints that should be removed[78]. In time series data, such as PV time series, outliers can arise from causes that fall beyond the expectations of a model, so for example of 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[79]. But clouds, being statistically random, are not easily modeled, and the power drops due to cloud shading, would be outlier data points, yet at the same time 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[80]. For outlier detection we use the tsoutliers R package, which identifies time series outliers arising from clouds and anomalous data points such as arise from measurement errors[81,82]. 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 data points 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 3rd important characteristic. These three categories of

statistical properties give us a quantitative sense of the "missingess" 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[83].

3.4.2 Dataset quality issues

Various dataset quality issues are present in the selection of PV system power time series, as can be seen by in Table 5, and are discussed here.

- EURAC System: No major data quality issues have been detected. This system has more than 10 % outliers (Figure 2a), probably due to cloudiness.
- FOSS System: No major data quality issues have been detected. This system had a higher amount of missing datapoints (Figure 2b).
- RSE Systems: No major data quality issues have been detected. These systems had a 25-day long gap in the datasets, as can be seen in the power heatmap (Figure 6).
- Pfaffstaetten Systems: A relatively low amount of measured data has been reported. This can be seen in the missing % and the # of datagaps, and the power heatmap (Figure 7).
- US DOE RTC Baseline Systems: Several data quality problems have been detected and are visualized in the power heatmap (Figure 8). These resulted in some filter-metric-methods to be unable to calculate sensible *PLR* results for some methodologies. System c10hov6 experienced a four month long initial inverter clipping followed by a period of four 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 as is visualized in the power heatmap (Figure 9). 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 was be replaced by modelled clear-sky values (that were provided in the raw data files). The irradiance sensor used for NREL3 is installed a distance of a few hundred meters away from the PV system, and has a different tilt, which was translated to the plane-of-array, and shows decreasing irradiance values over time, possibly a result of a degrading reference cell.

3.4.3 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 inputs for *PLR* determination, are a sample of a real world PV system. Consider two different samples of this one system, a 1 minute interval and a 5 minute 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, e.g. by lagging one measurement by 30 seconds, these should have the same measurement uncertainty σ_{meas}, but they are two independent samples, and their sampling uncertainties will vary when one considers the population mean μ of the *PLR* (i.e. the $\overline{PLR}$) 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 its 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 the dataset's seasonal, trend and residual components, and then the sampling standard deviation can be calculated from the time series residuals.

Table 5: Statistical characteristics of PV datasets used in the *PLR* benchmarking exercise.

Dataset ID	Grade Outlier, Missing, Gaps, P/F	Power Variable [kW]	Length (Years)	Outlier (%)	Missing (%)	Data Gaps: # of gaps	Data Gaps: longest gap (days)
EURAC	BAAP	P_{DC}	7.95	11.5	2.1	2847	7
FOSS*	BCBP	P_{DC}	10.9	13.8	32.9	134	26
RSE CdTe	AABP	P_{AC}	9.59	10	0.3	2	25.3
RSE pc-Si	BABP	P_{AC}	9.59	11	0.3	2	25.3
Pfaffstaetten A*	ADAP	P_{DC}	6.33	2	41.2	2082	0.9
Pfaffstaetten B*	ADAP	P_{DC}	6.33	2.2	40.1	2014	0.9
Pfaffstaetten C*	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	0.5	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*	BACP	P_{DC}	3.31	13	6.3	727	76.7
NREL2*	BABP	P_{DC}	6.06	15.2	4.3	1733	22.2
NREL3**	AADP	P_{DC}	7.88	8.7	10	669	146.1
NREL4*	ABBP	P_{DC}	6.82	1.7	18.5	1999	27.9
Digital power plant 1	AAAP	P_{DC}	5	8.2	0	0	0
Digital power plant 2	AAAP	P_{DC}	5	7.9	0	0	0
Digital power plant 3	AAAP	P_{DC}	5	8.6	0	0	0
Digital power plant 4	AAAP	P_{DC}	5	8.6	0	0	0

*Incomplete cases omitted.
**Shows negative and high values (>60) for P_{DC}, and a high power time series standard deviation.

In many fields, researchers do not clearly distinguish between the standard deviation (sd) of a set of measurements, a descriptive statistic, and the standard error (se) of the mean[84]. 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 "true" value of a population mean, which in this study is the true *PLR* value of the real-world PV system[85]. Here we are benchmarking which filter-metric-method approach can capture, or determine, the "true" voted value for a system. Our goal is for all the researchers, using common approaches for *PLR* determination, such as in the RdTools or PVplr packages, or their own implementations, should produce comparable results that "true" *PLR* of the system. We are benchmarking methods applied to diverse datasets, to identify the most robust approaches for determining the replication means for these 19 systems[86]. 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*, $\overline{PLR}$. Here, since we have no apriori basis to know which of these many methods is "correct", we expand to use multiple methods, and by comparing standard errors, and 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 across the filter-metric-methods of Table 1 and Table 2. n null hypothesis testing, for a significance level, or Type 1 Error Rate, of $\alpha = 0.05$ [33], corresponding 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 $\overline{PLR}$ of these results.

Consider that we want to know the true *PLR* of one PV system, and we use one or many *PLR* determination filter-metric-methods and calculate *PLR* 100 times, the *PLR* will vary around the mean, or "true" $\overline{PLR}$ and the important descriptive statistic is the standard error of this $\overline{PLR}$. And as we calculate more values, our confidence in the mean improves. This is the basis of the preference aggregation, or"voting", method we apply here so that we can determine the most likely mean value of *PLR* for the 19 real-world PV systems[87]. By utilizing many data filtering and statistical modeling methods applied across all 19 datasets, the $\overline{PLR}$ value is probably the most likely. And the standard error of the mean is a measure of the variance of these methods in determining the $\overline{PLR}$. To compare multiple methods of *PLR* determination for a single system, we compare the different results and their 95 % confidence interval (CI) to achieve a p-value of 0.05. With this approach we will find the true population $\overline{PLR}$, within 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[88]. This is the approach we use here, for example as shown for the EURAC system and multiple methods in Figure 21. If in this Figure 95 % CIs for each result were used, then we could define which methods give similar estimates of the PV system's $\overline{PLR}$, and which methods provide distinctly different estimates.

Regression based *PLR* uncertainty can be evaluated from the variance of the slope of the linear model coefficients for the corrected performance metrics with time (Section 1.4.1), by reporting the 95 % confidence interval of the final *PLR* result, as described in Lindig's paper[7]. 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 uncertainty of a regression *PLR* and a *PLR* determined on decomposed time series are directly comparable. Otherwise STL would have an apparent and artificial advantage due to the separation of the residuals' variance 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 density function of the individual *PLR* results to represent the uncertainty in the reported *PLR* and one needs to determine the correct comparative uncertainty measure; should this be the standard deviation, standard error, or the appropriate 95 % or 83.4 % confidence intervals.

Since comparisons of the *PLR* uncertainties of different filter-metric-method combinations are important we seek to address the divide between regression and YoY *PLR* uncertainties. *PLR* uncertainty can also be evaluated using bootstrap resampling of the time series with replacement[24,47,48,33]. For *PLR* estimates obtained using regression[7], 65 % of the days are randomly chosen from the total time series and the *PLR* is recalculated, and this process is repeated, typically, for 1000 iterations. The standard error of the distribution of the $\widehat{PLR}$ estimates obtained from all iterations represents the uncertainty of the $\overline{PLR}$ value. A more stable $\widehat{PLR}$ is expected to have less resampling variance. This process can also be used with YoY as well, however instead of resampling individual days, which would bias YoY, the final $\widehat{PLR}$ distribution is resampled. Bootstrap resampling requires large computational capabilities and it's important to confirm that the bootstrap uncertainty has converged, therefore bootstrap is usually run 1000 times to confirm convergence. And bootstrap resampling does enable a direct comparison between uncertainties determined using regression and YoY *PLR*.

The graphical display of 95 % CIs and visualizing whether the CI ranges overlap, is an effective way to enable "inference by eye" and display the relative uncertainties of the $\widehat{PLR}$ estimates. This graphical confidence interval visualization approach also enables multiple comparisons, whereas the traditional student's t-test is only a pairwise comparison[89,90]. For a 5 % type I error rate and significance level, corresponding to a p-value of 0.05, we can compare methods for one system by plotting the 95 % CIs.

If we want to compare two PV systems and determine if the difference of their $\overline{PLR}$ is significant, then we need to adjust our CI criteria to be appropriate for a t-test of the difference of two means. To determine if the $\widehat{PLR}$ for two different PV systems are statistically similar (the null hypothesis), or significantly different, we modify the size of the CI we visualize so as to achieve the same 5 % significance level α, which is also referred to as a 5 % capture rate. In this case, the comparison of the means of two systems ($H_0: \overline{PLR_1} = \overline{PLR_2}$, $H_a: \overline{PLR_1} \neq \overline{PLR_2}$), to achieve a 5 % capture rate or type I error, we graphically compare the overlap of the confidence intervals of $\overline{PLR_1}$ and $\overline{PLR_1}$ using 83.4 % confidence intervals, as this achieves a 5 % capture rate[91]. If we used 95 % CIs for the comparison of two means this would correspond to a hypothesis test with a p-value and significance level of ~ 0.01. This suggests that researchers who wish to compare on a common basis, different *PLR* determination methods should use 95 % Cis on a single system as we do in this report. A PV system fleet owner wanting to determine which systems are exhibiting similar or distinctly different performance loss, should graphically compare the overlap of 83.4 % Cis, and in Figure 22 we show these results for comparing across the PV systems in this study. When the minimum CI for one system just touches the maximum CI of the other system, you have met a p-value significance level of 0.05 and a Type I error of 5 % for your comparison of two system means hypothesis test.

3.4.4 Benchmarking *PLR* digital datasets

The digital datasets are introduced in Section 2.3.7 along with the real PV systems and Table 6 summarizes the most important characteristics of the plants.

Table 6: 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

Initially four PV systems have been simulated, two plants with five years of repeating weather data and two PV systems with four 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 for these systems the "real" *PLR* values are known, which is not the case for the 19 real PV system 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.

The individual methods used in this comparison have been introduced in Table 1 while Figure 14 shows the calculated 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 line representing the degradation value for P_{mpp} at STC and the green line the energy degradation. For these two systems, with no induced degradation, these values are very similar to one another.

Looking at the results, it is clear that the calculation accuracy varies depending on the source of the weather data. While all filter-metric-method approaches used yield the correct value within a ±0.05 % interval for the system with repeating weather data, greater deviations can be detected looking at Figure 15 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 a high irradiance threshold filter (VAR1, R-LR1, LS-LR1, STL1) or predicted power approaches as metrics (i.e. XbX + UTC) seem to be able to remove this induced effect and provide reliable results.

Figure 15 shows the results for the digital plants with simulated degradation. Here, the indicated degradation values (orange and green line) vary distinctively from one another. Among other things, this has to do with the irradiance distribution for this particular simulated site. In Figure 25 in the appendix, the yearly plane-of-array irradiation per irradiance interval is shown for the dataset measured at the test site in Rennes/France, where more than 60 % of the solar resource arises from irradiance values below 600 W/m². Therefore a large amount of the PV electricity is produced at irradiances far away from STC conditions and therefore both degradation values differ. It is expected that such degradation estimates are closer in areas with a high amounts of irradiation from higher irradiances.

For the simulated digital power plant 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 *PLR*. 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 the larger *PLR* values.

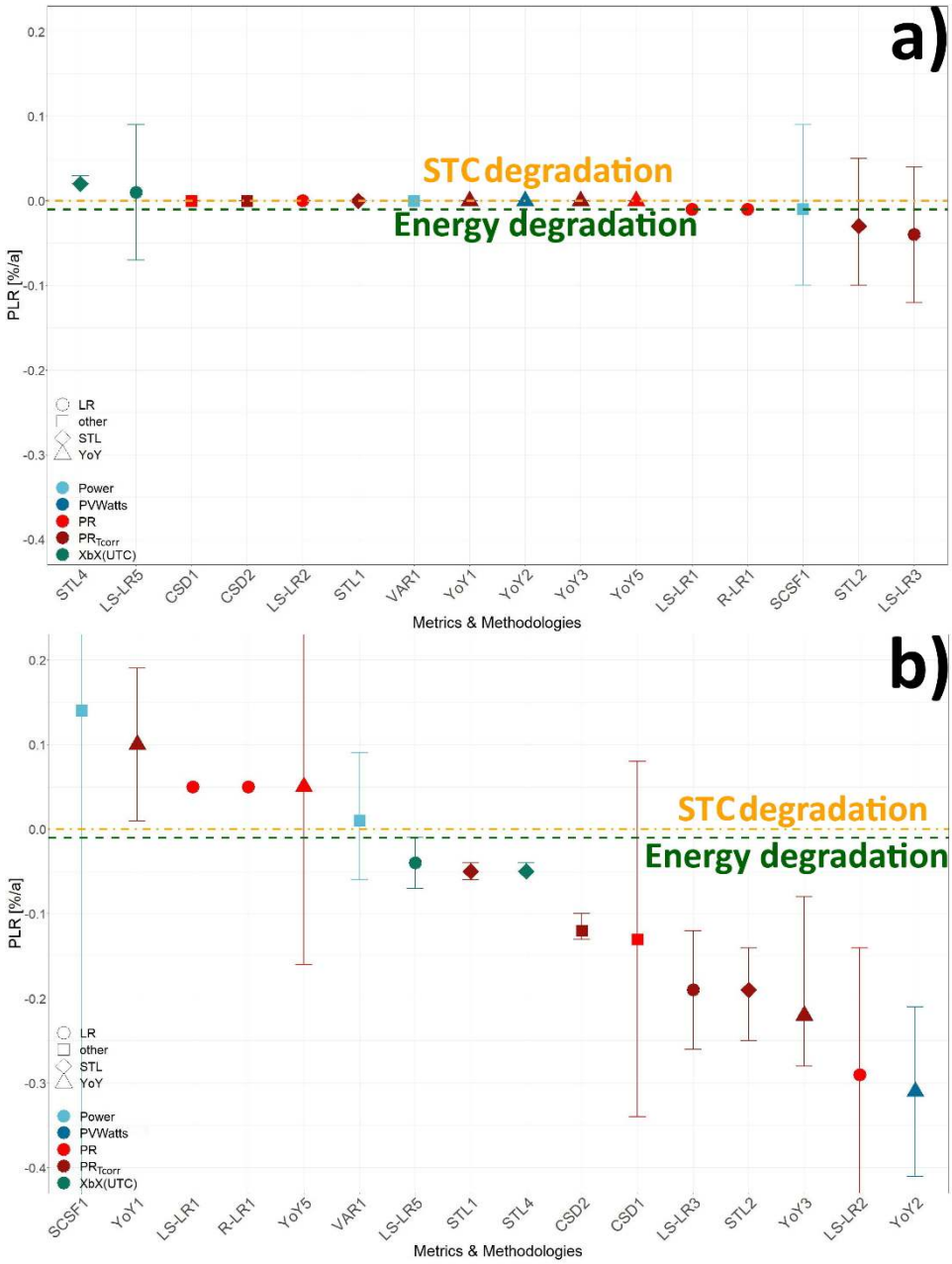


Figure 14: a) Calculated *PLR* of digital plant with no degradation & repeating weather data; b) Calculated *PLR* of digital plant with no degradation & real weather data; orange line: *PLR* at STC; green line: *PLR* in absolute energy. The "error bars", are as reported by the analyst[1].

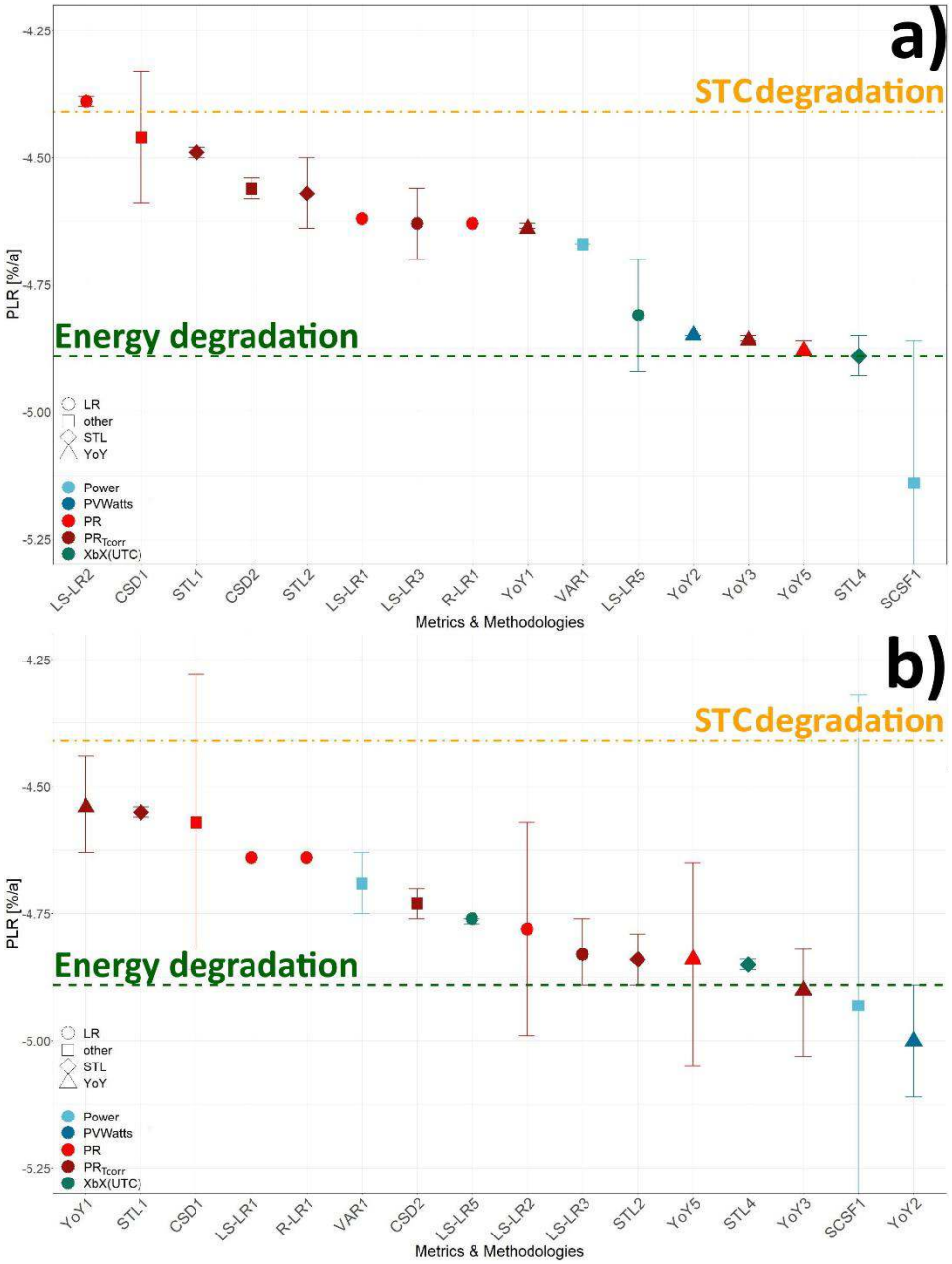


Figure 15: a) Calculated *PLR* of digital plant with degradation & repeating weather data; b) Calculated *PLR* of digital plant with degradation & real weather data; orange line: *PLR* at STC; green line: *PLR* in absolute energy. The "error bars", are as reported by the analyst[1].

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 anomaly removal, is unnecessary as the raw datasets do not have any anomalous values.

3.4.5 Method validation for low *PLR* digital datasets

As a continuation of the work with the digital power plant *PLR* comparisons, additional data was requested which had a range of lower *PLR* results to compare consistency between different *PLR* methods in predicting the actual values. For these systems, the PVplr R package was used to evaluate the *PLR* with three available predicted power models (XbX, XbX + UTC, PVUSA), and these were done with, and without, STL decomposition, and with year-on-year and linear regression evaluated *PLR*.

Data filtering was kept at a minimum, with only an irradiance filter of 100 W/m^2 applied to the systems' datasets. As the data is simulated there were no gaps or corrections that required addressing beyond the irradiance filter. All predicted power models were built with a weekly data aggregation scale, with the coefficients of each weekly model being used to evaluate the long term degradation. The *PLR* results of all the calculated cases is shown in Figure 16. It is clear that the weather data used has a strong impact on the calculated *PLR* values for the PVUSA and XbX cases. In the systems with the 5th year having colder weather data (Diff. Meteo.), there is a distinct shift upwards in the results. The XbX + UTC model, which uses a universal temperature correction as defined by IEC61724-1:2017[5], does not show this shift to nearly the same degree. Both the PVUSA and XbX methods model temperature within each weekly segment, as opposed to an overall temperature correction, however the result show the temperature correction performed in these models cannot be extrapolated and will give different results within different temperature ranges. As the last year has a distinctly lower temperature than the other years, these models cannot adequately adjust the temperature influence between the different weather conditions, leading to higher predictions of power in colder weather. This leads to an artificially increase of the magnitude of the *PLR* result. The universal temperature correction enables correction for larger ranges of temperature and the power predictions in the 5th year are comparable to the others. It should be noted that the PVUSA and XbX models show greater accuracy under constant weather conditions (Same Meteo.), indicating they can still perform well in situations where there is not a significant temperature variation between years.

A more detailed view of the low degradation EDF digital power plant datasets is given in Figure 17 where the difference between the real (the imposed performance loss used in the simulation) and calculated *PLR* results are plotted as a function of the real *PLR*. The shift in *PLR* caused by the differing meteorological data case is again observed in the PVUSA and XbX models, with the XbX + UTC model maintaining predicted *PLR* values much closer to the real results. In all cases there is a trend for a larger difference in *PLR* values at a higher magnitude of *PLR*, in particular for the PVUSA and XbX models, with a trend towards zero as the *PLR* decreases. The PVUSA model shows a significant amount of spread between *PLR* results calculated using year-on-year versus regression for the different weather case and between decomposed and non-decomposed values for the same weather case, leading to an overall greater variance in results than the other two models. The XbX + UTC model shows a trend of over predicting the *PLR* results compared to the other two models, which consistently under predict the results, even in the same yearly weather case (exceptions for both cases occur at low *PLR* values). At this time it is not clear what would lead to a model having consistent trends in over or under predicting *PLR*.

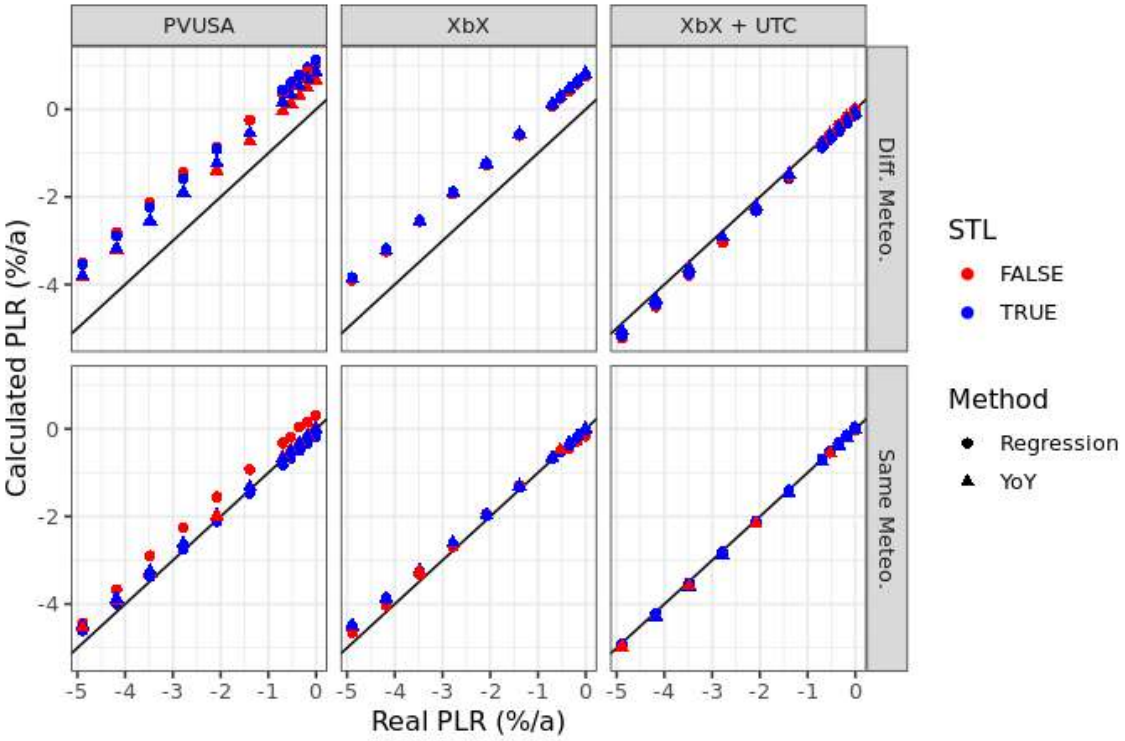


Figure 16: Calculated *PLR* values versus real *PLR* values for all cases from the EDF low degradation digital power plants. The black line indicates equality between results.

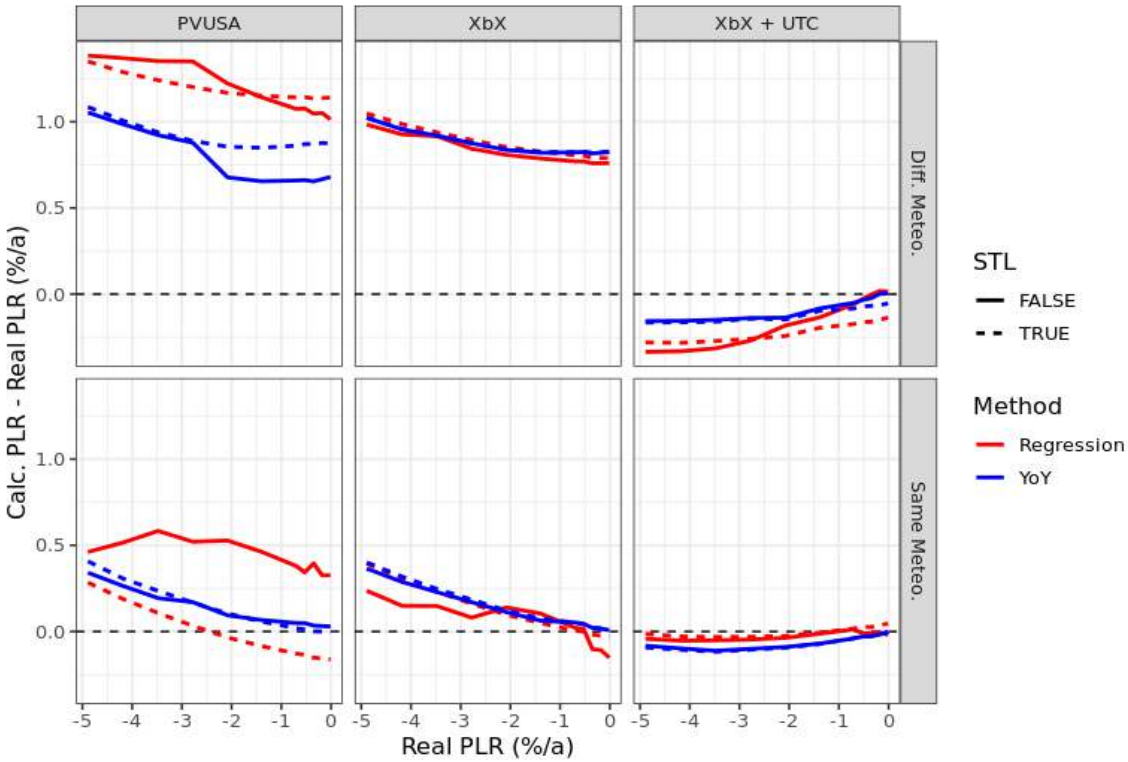


Figure 17: Difference between calculated and real *PLR* values versus real *PLR* values for all cases from the EDF low degradation digital power plants.

Despite the differences shown between these different models it must be noted that while the digital power plants are an excellent tool to demonstrate the accuracy of a *PLR* methodology, these datasets do not contain many of the essential features that are found in datasets from real PV systems. Real systems exhibit differing variance or noise in their data, the annual weather variations, module soiling, sensor inaccuracy, etcetera. Therefore the digital power plants while able to simulate a known *PLR*, do not give a complete picture of the accuracy of our *PLR* determination filter-metric-models for real PV system datasets. The variations in the meteorological data datasets do show the clear benefit of a universal temperature correction and suggests that a wide temperature range is needed to properly account for temperature influences in a PV system on an annual scale.

3.4.6 Impact of data filtering on *PLR* determination

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 filters and their name labels are given in Table 1. Temperature corrected performance ratio (PR_{Tcorr}) is used along with monthly data aggregation, and STL from the stlplus R package 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 systems with degradation and real weather data was studied along with the EURAC system. Figure 18 shows an example power versus POA irradiance plot for filter #3 applied to the EURAC PV system. The power along the y-axis has been normalized to the nominal power of the system and a plane-of-array irradiance interval from 0 to 1250 W/m² is depicted along the x-axis. The blue points represent the raw data, the green points a first threshold filter and the red points the final filtered dataset, which is then used for the subsequent *PLR* calculation. Some filters only include a threshold filter, in which case the in-between filter step is omitted and only the final filtered data are shown.

Figure 19 shows both the calculated *PLR* values of the digital plants and their dependence on the different applied filters, together with power versus plane-of-array irradiance plots in order demonstrate the strong impact of the individual filters. Similar results are shown for the EURAC PV system in Figure 20.

Digital plant with degradation and real weather data:

Figure 19 demonstrates clearly that the choice of filter strongly affect the magnitude of the reported *PLR*. The calculated *PLR* values range from -4.48 %/a to -5.47 %/a. It appears that filters with similar irradiance cut-off thresholds cluster together when using the same metric-method combination. This correlation is strong for the digital power plant datasets, which are not subject to anomalous points or abnormal points. At the same time the outlier grade or filtering (mainly *PR* related) does not have any effect on the outcome. Stricter irradiance thresholds, that remove more low irradiance datapoints, yields lower *PLR*s, an observation that could be exploited by an analyst to arrive at "desirable" *PLR* results. This suggests that *PLR* determined using different data filters, are not comparable results. The application of filters #3, #1, #4 and #7 results in *PLR* values with the lowest calculated values, particularly close to the STC *PLR*, while the irradiance thresholds stretch from 350 to 800 W/m². Filter #6 and #9 apply a 200 W/m² cut-off 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/m², 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/m², 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. Consider PV plant datasets free of any data anomalies (i.e. Grade A in outliers), such as the simulated digital power plants presented here, and the approach of applying STL together with PR_{Tcorr}. For these systems it appears that low irradiance cut-offs between 50 and 200 W/m² yield the most accurate results based on the energy related *PLR* described in Section 3.4.4 for the digital plants.

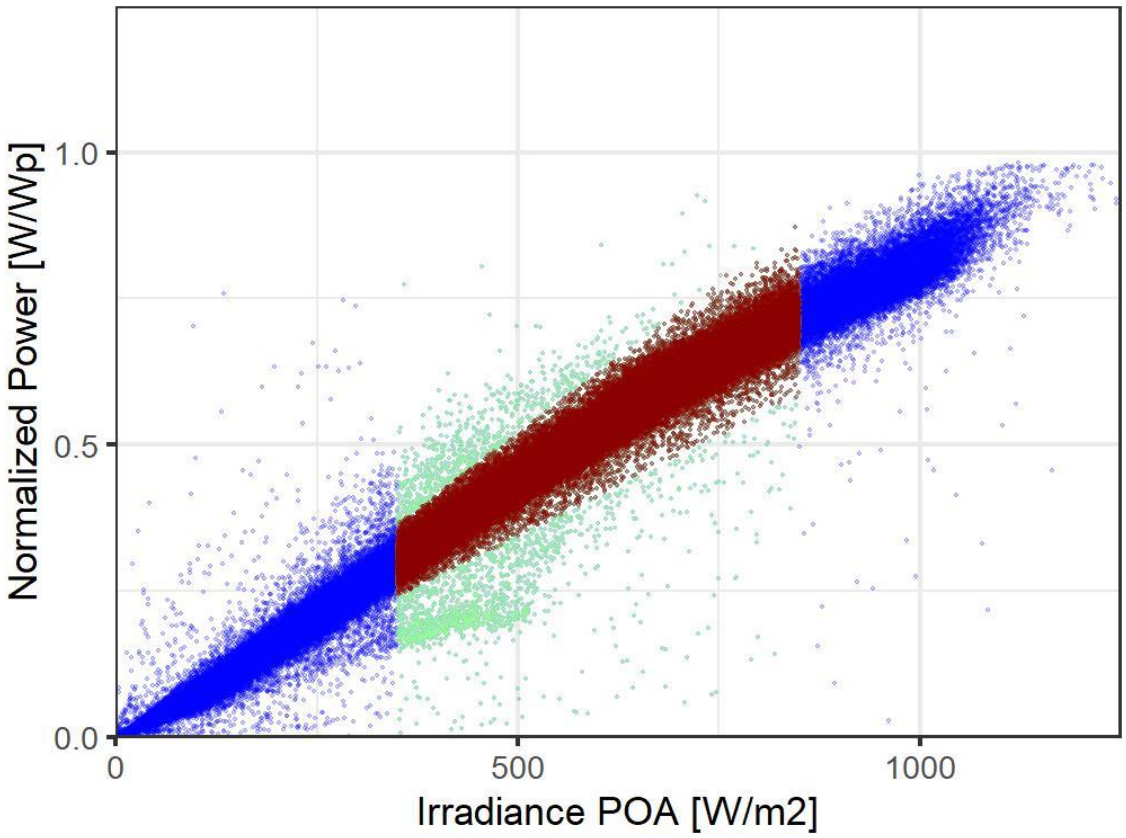


Figure 18: Normalized power versus irradiance plot for EURAC dataset with applied filter #3; blue - raw data, green - threshold filter; red - final filtered data used for *PLR* determination[1].

EURAC System:

Looking at Figure 20, one can see that similar observations can be made 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.

The filter #7, #4, #10, #3, #1 and #2 yield *PLR* very close to the mean *PLR* 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 %

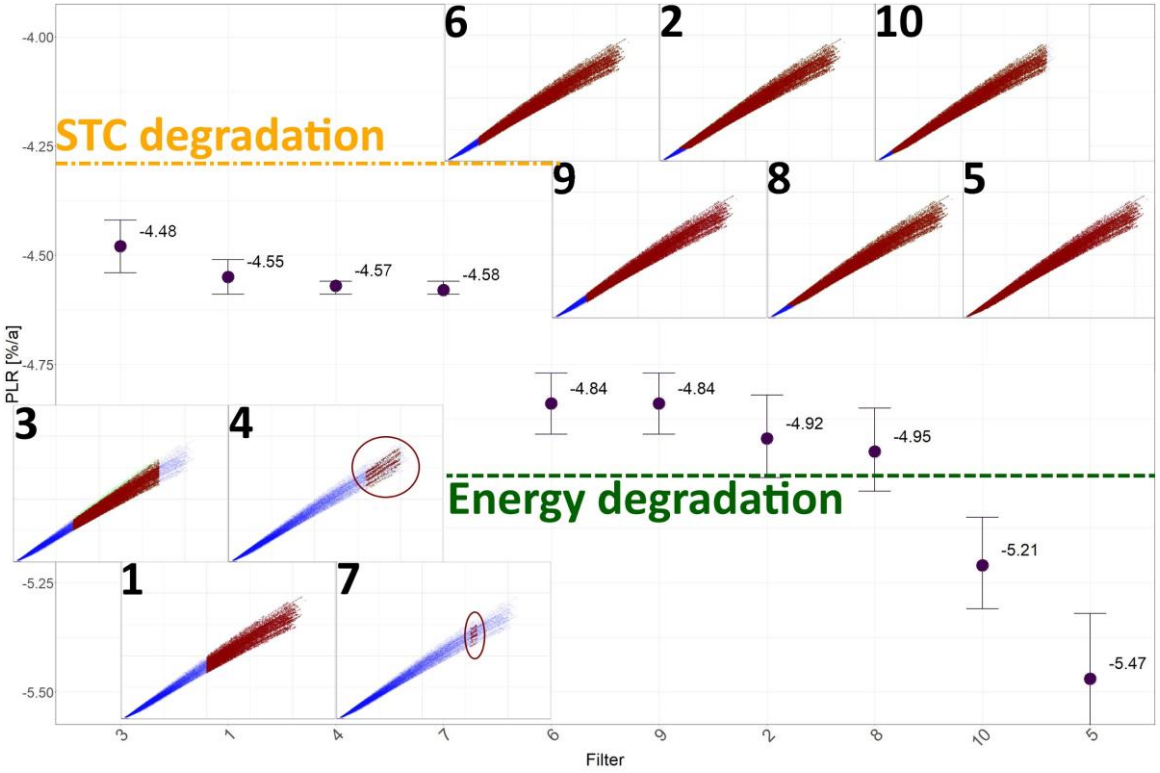


Figure 19: Calculated *PLR* using all ten studied filters with PR_{Tcorr} as metric and STL as calculation statistical method for digital plant with degradation and real weather data; blue - raw data, green - threshold filter; red - final filtered data used for *PLR* determination[1].

Filter #1 already applies a strict irradiance threshold at 500 W/m² but the data within the considered irradiance interval, although 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 therefore 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 3.4.8, 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 even 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, YoY2 uses filter #9 and yields quite accurate results for the EURAC system (see Figure 11) 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 again the strong dependency between all calculation steps from filtering up until the choice of a statistical method.

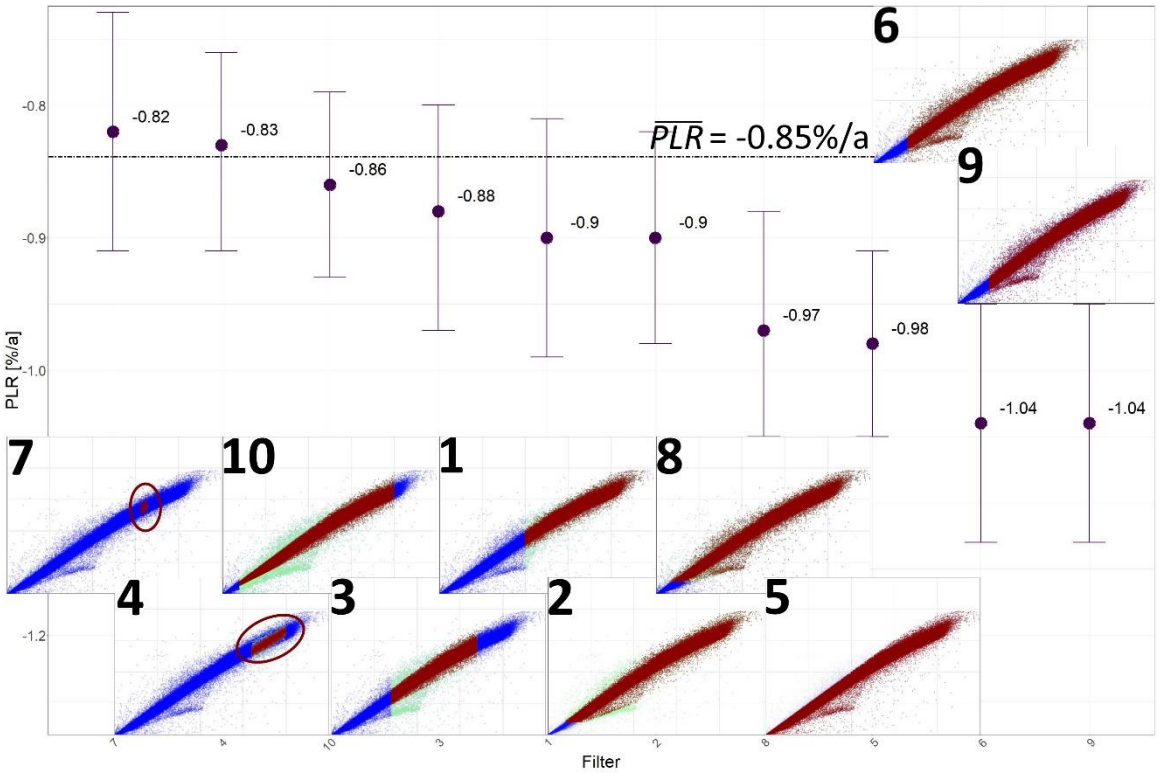


Figure 20: Calculated *PLR* using all proposed filter with PR_{Tcorr} as metric and STL as calculation statistical method for EURAC system[1].

3.4.7 Evaluation methodology of benchmark *PLR* results of real PV datasets

Evaluation of the benchmarking results across the 19 real datasets is complicated by the fact that the "true" value of the performance loss rate 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 $\overline{PLR}$ as the mostly reasonable value for each PV system. First, the ($\overline{PLR_i}$) for an individual PV system is calculated using all calculated *PLR* values (Figure 21a)). Next, the relative difference of all methodologies from the $\overline{PLR_i}$, which is set at 0 %, is calculated for this particular "*i*" system (Figure 21b)). 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 plane-of-array 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.

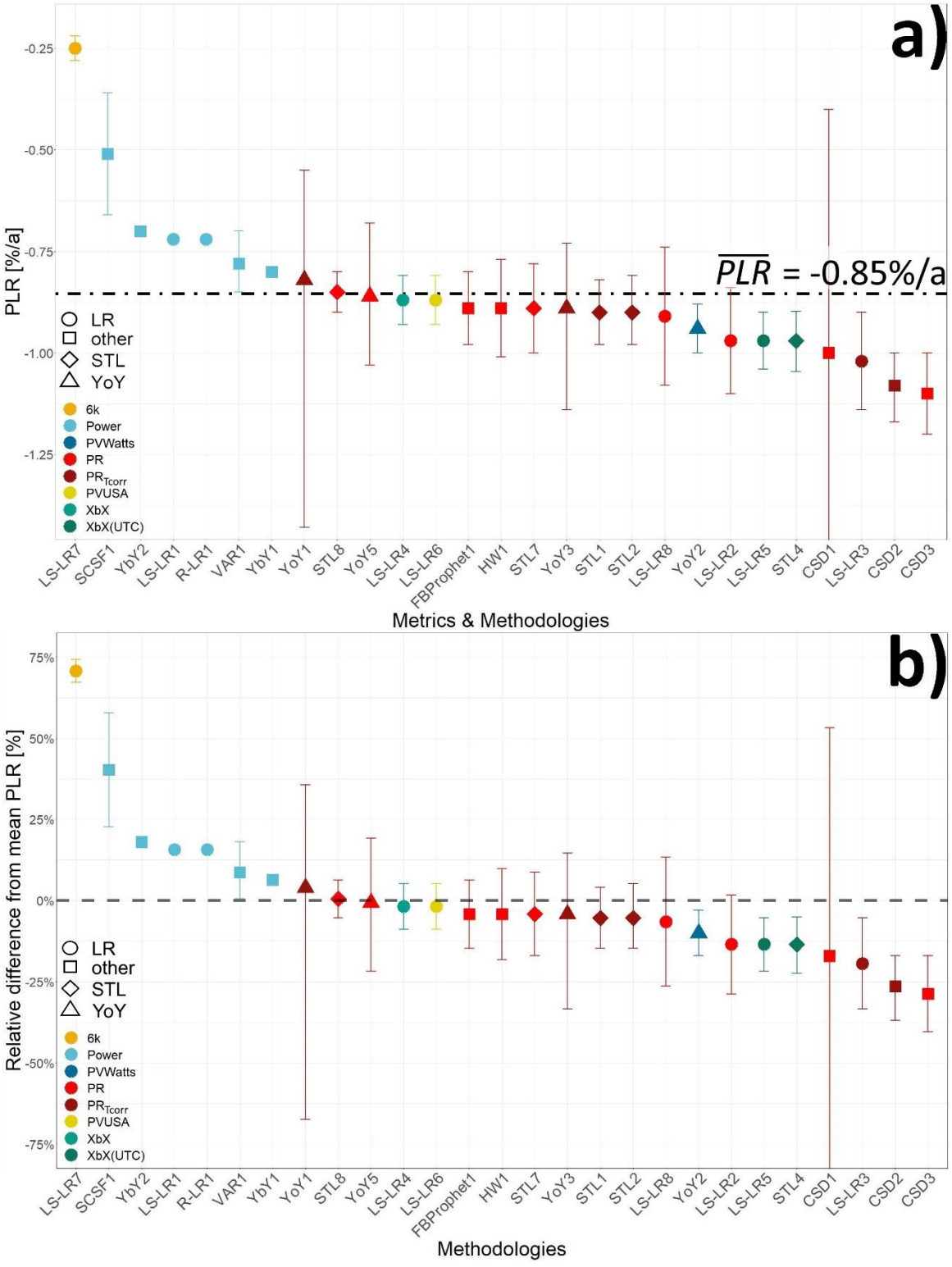


Figure 21: a) Calculated *PLR* of EURAC system; b) Relative calculated *PLR* values of EURAC system. The "error bars", are as reported by the analyst[1].

The difference from the normalized $\overline{PLR}$ 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

Equation 11.

$$\frac{\left|\sum_{i=1}^{n}\left(\frac{\overline{PLR}_i - PLR_i}{\overline{PLR}_i}\right)\right|}{n}$$

- Standard error of the average deviation

Equation 12.

$$\sqrt{\frac{\sum_{i=1}^{n}\left(\frac{\overline{PLR}_i - PLR_i}{\overline{PLR}_i} - \mu\right)^2}{n-1}} \Bigg/ \sqrt{n}$$

Here, *n* is thereby the number of PV systems on which a particular approach was used, and μ is the mean of the numerator of **Equation 11** over all systems for one filter-metric-method. The first KPI (**Equation 11**) 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 ($\overline{PLR_i}$) and the *PLR* for each statistical method is calculated where *i* refers to a specific PV system of the *n* PV systems analyzed in this report. The second KPI (**Equation 12**) provides an indication on how the average value deviates from dataset to dataset. Finally, all results are averaged in a target plot to see which methodologies perform the best across all systems. The results are discussed in detail in Section 3.4.8. Figure 21 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 confidence intervals, 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 7, the $\overline{PLR}$ for all systems are depicted to get an overall impression of the degradation of the systems under evaluation. The calculated $\overline{PLR_i}$ and the 83.4 % confidence intervals for inference by eye of the difference of two means is shown in Figure 22 using all *n* filter-metric-methods that were successfully applied to each PV system.

Unfortunately, some of the datasets under investigation had certain dataset quality issues, as discussed in Section 3.4.2 and Table 5 and individual methodologies failed to yield *PLR* results for these datasets. Particularly affected systems are the ones belonging to the US DOE and NREL datasets. *PLR* of US DOE wca0c5m and US DOE z0aygry are marked with a star as a shift in the power output measurements has been detected and just methodologies, in which the shift was detected, were included for the calculation of the $\overline{PLR}$.

3.4.8 *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 *PLR* 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 *PLR* values decreases

the average *PLR* variability and therefore increases the accuracy of the estimated reference *PLR*. Thereby, the trustworthiness of this "voted" *PLR* as being close to the "real" *PLR* is higher if more *PLR* values are included.

With the voted results of our interlab/multi-method comparison shown in Table 7 and Figure 22, we may actually have in hand an ensemble learning approach to determine the $\overline{PLR_i}$ that is accurate and reproducible. Ensemble models in machine learning, are where different modeling approaches are all used and the final result is a voted result across all models[92]. An example of this is how Random Forest machine learning, is the result of a "forest" of decision tree models all averaged together, and this ensemble averaging allows the different approaches to counterbalance their uncertainties. In the case of *PLR* determination, with tool such as RdTools and PVplr, it becomes easier to perform an ensemble of *PLR* results on a system and then calculate the PV systems' $\overline{PLR_i}$.

Table 7: The mean *PLR* ($\overline{PLR}$), across all filter/methods used, for all systems included in the benchmarking study.

System	$\overline{PLR}$	System	$\overline{PLR}$	System	$\overline{PLR}$
EURAC	-0.85 %/a	NREL1	-0.33 %/a	US DOE luemkoy**	0.95 %/a
FOSS	-0.71 %/a	NREL2	-0.54 %/a	US DOE lwcb907	-0.03 %/a
RSE CdTe	-1.75 %/a	NREL3*	0.06 %/a	US DOE t3pg1sv	-0.75 %/a
RSE pc-Si	-0.96 %/a	NREL4	-0.25 %/a	US DOE wca0c5m***	-1.00 %/a
Pfaffstaetten A	-3.57 %/a	US DOE c10hov6	-0.50 %/a	US DOE wxysjaf	-0.97 %/a
Pfaffstaetten B	-3.96 %/a	US DOE kob-dpi8	-0.73 %/a	US DOE z0aygry***	-2.32 %/a
Pfaffstaetten C	-1.29 %/a				

*The provided modelled 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 29).

**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 true *PLR*.

***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 $\overline{PLR}$, see Figure 30.

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 7 systems with a minimum of 3 calculated *PLR* values per filter-metric-method. This includes 24 filter-metric-method combinations.
- **G13-7-17**: evaluate 13 systems with a minimum of 7 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 7 systems with 24 approaches:

The first evaluation is based on the results calculated for the 7 most pristine 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 7 systems ranges from 20 to 27.

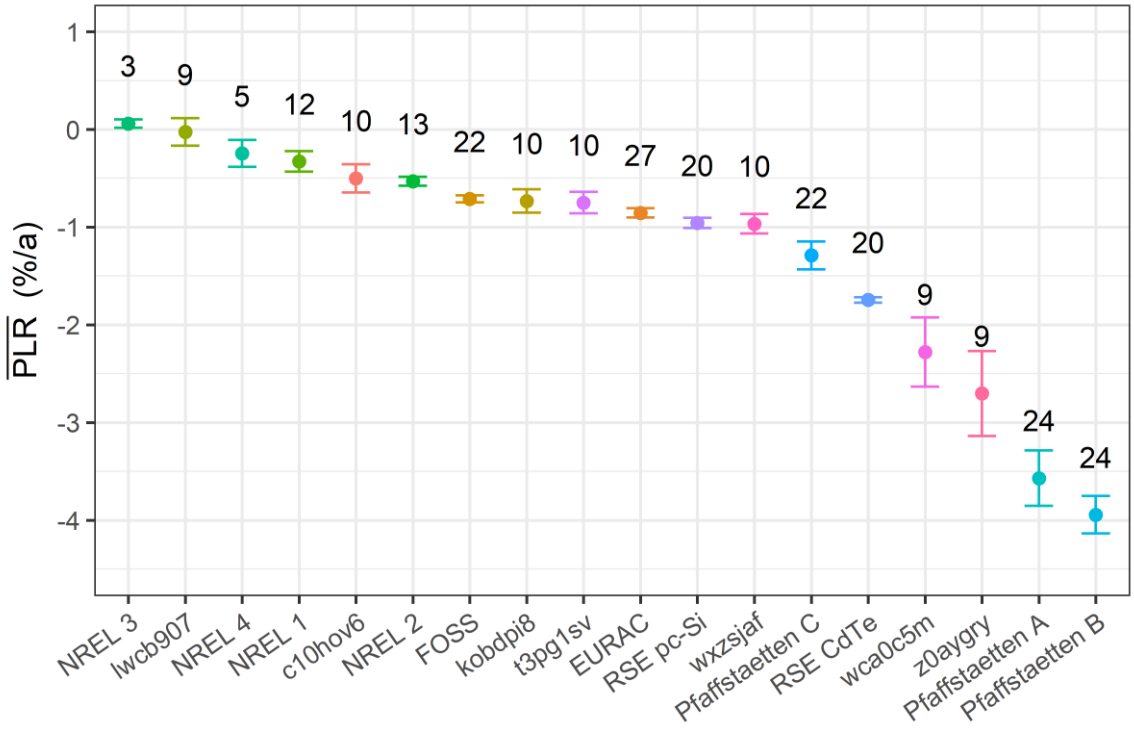


Figure 22: The mean *PLR* and 83.4 % confidence intervals (CI) for each of the 18 PV systems in the benchmark study. One system (luemkoy) that had a large positive *PLR* is not shown. If two systems have CIs that touch without overlapping, then they are different at a 5 % significance level. The number (*n*) above each system's CI is the number of *PLR* results in the sample, and corresponds to the *n* used in the standard error calculation.

The *PLR* statistical method ratings when considering these 7 different PV systems are shown in Figure 23. The deviance, a goodness-of-fit statistic for a statistical model[33], shown along the x-axis and describes the absolute overall difference from the reference $\overline{PLR}$, and along the y-axis the standard error of the average differences from the reference *PLR* across the systems under consideration is shown. Uncertainties are omitted because no consistent approach to report the uncertainties (such as confidence intervals) was used as we collected the benchmark results from the participating analysts 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 % intervals, are a guide for the eye to categorize the results into different groups of accuracy.

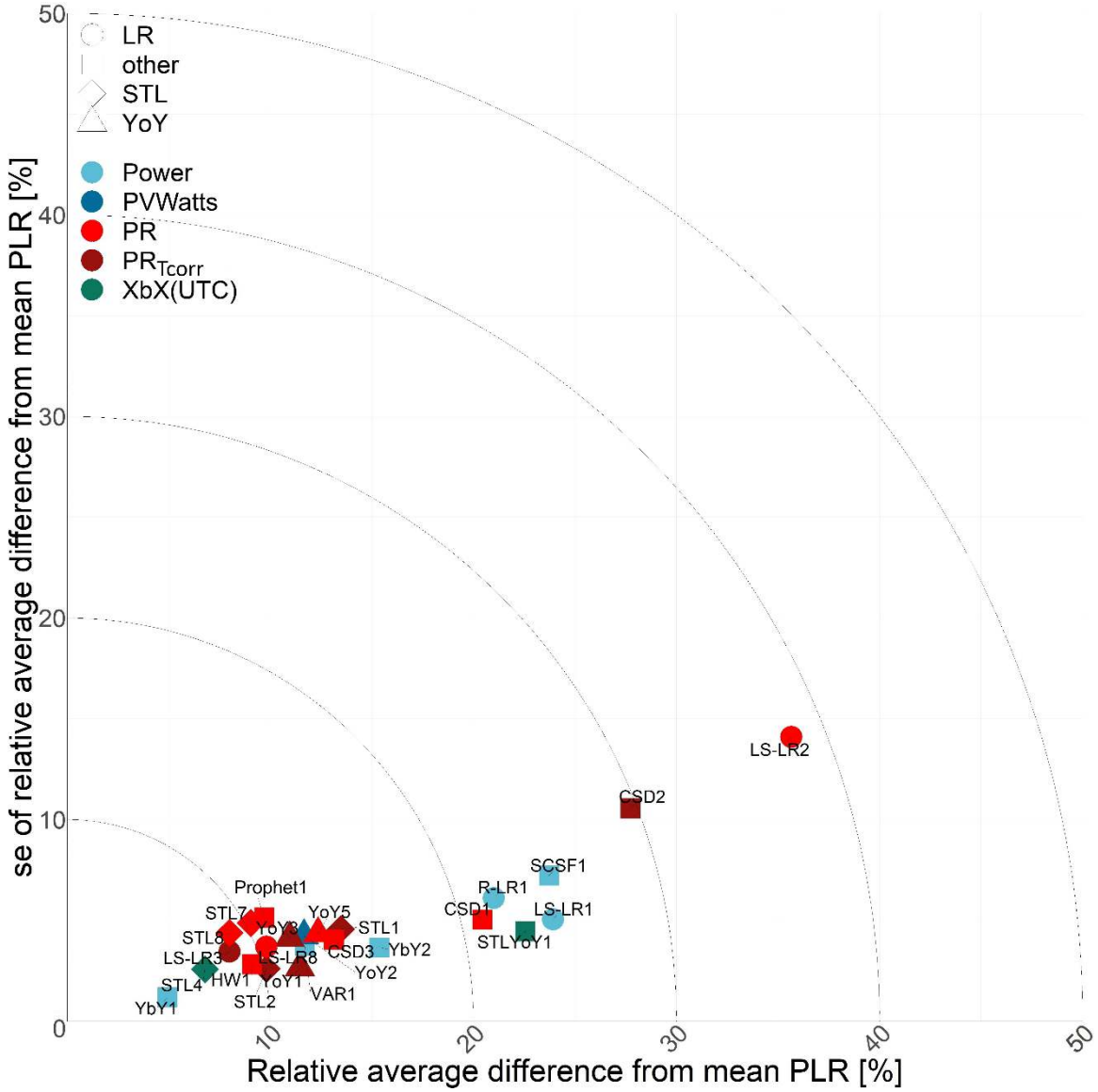


Figure 23: Target plot with absolute average deviations from $\overline{PLR}$ value and standard error considering 7 PV systems (excluding all NREL and US DOE datasets)[1].

In Figure 23, a minimum of three calculated *PLR* was set for a statistical method to be included in the benchmark evaluation of these 7 different PV systems. These datasets are considered as high quality research PV systems without serious data quality issues apart from the minor ones reported in Section 3.4.7. 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 $\overline{PLR}$ of up to 17 % and a corresponding standard error of 1 to 6 %. According to the results evaluating 7 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/m² around NOCT conditions.

From 24 tested filter-metric-method approaches, seven are not in the first and second isoband with deviance values from the $\overline{PLR}$ greater than 20 % and these perform with lower accuracy. YbY2 is at the edge of the cluster of 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. In addition LR, when the metric was not subject to temperature correction is subject to high variance compared to other approaches. 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 2) used for LS-LR2 only applies a *PR* threshold of 0 to 100 % and thereby may not sufficiently exclude outliers and anomalous points, which are ultimately affecting the final result since LR is strongly affected by non-valid data-points.

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

It is interesting to observe the deviance in the results looking at YbY1 and YbY2. For both of these 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/m²), YbY2 filters around STC conditions (980-1020 W/m²). Just considering this small sample of datasets, it seems that NOCT conditions represent the calculated $\overline{PLR}$ better in direct comparison. Furthermore, both methodologies, together with LS-LR1 and R-LR2, use narrow irradiance bands and temperature filtering and exclude thereby the vast majority of data-points. Three out of four (except the mentioned YbY1) of these “heavy filtering” approaches yield results with higher uncertainties, possible because of filtering out large amounts of data. The impact of filtering on the calculated *PLR* is further discussed in Section 3.4.6.

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 3.4.7). 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 $\overline{PLR}$ 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 $\overline{PLR}$, 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 inter-annual 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[93]. For example, Kiefer et al.[94] saw an average increase in irradiance of +1.1 % per year while studying the performance of several PV plants in Germany. These effects may 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 data-points are removed by the use of 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 outliers of power-irradiance pairs.

Overall the majority of the test approaches calculate *PLR* with relatively low uncertainties considering these research PV system datasets. This study serves thereby as a first indicator of *PLR* estimation accuracy for high quality datasets without major measurement and operation issues. But for broad application to commercial PV systems, the impact of data quality, missingness, filters, metrics, temperature corrections and statistical methods could lead to these “best” methods performing very differently on commercial PV systems.

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

The second evaluation considers 13 PV systems while excluding NREL3, NREL4, and US DOE systems luemkoy, lwcb907, wca0c5m and z0aygry. The latter two systems were excluded because of the detected data shift in the power output, which can be seen in the power heatmap (Figure 8). For those two systems, 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 7 calculated *PLR* values 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, and the results can be seen in Figure 24.

The inclusion of methodologies in itself is already a first quality characteristic by being applicable to this wider set of PV system performance datasets, 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 24 increased substantially.

While the majority of methodologies in Figure 24 yields results corresponding to values in the first and second isoband, seven of these now 17 methodologies are remaining in this area and the other 10 stretch over into the other isobands. The approaches with the highest accuracy are LS-LR8 and Prophet1. 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 quality issues are present, led to those results being omitted; so while relatively accurate, these two methodologies may not be robust. Four more statistical methods have been tested using the same metric and filter.

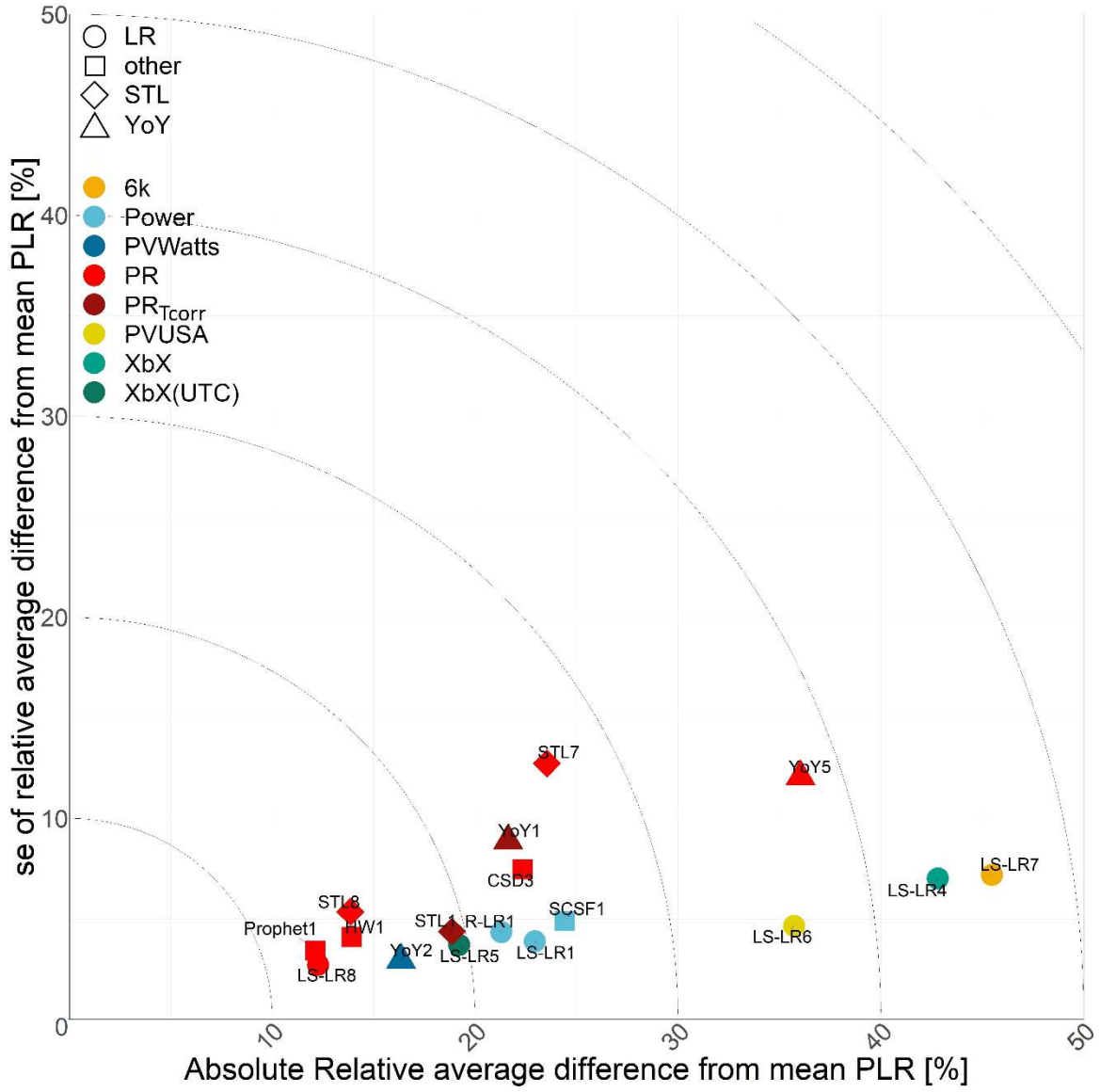


Figure 24: Target plot with absolute average deviations from $\overline{PLR}$ value and standard error considering 13 PV systems (excluding US DOE luemkoy, lwcb907, wca0c5m & z0aygry, and NREL3, NREL4) [1].

It is interesting to note that two of these use STL as the 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 standard error in Figure 24. A cross-comparison of STL7 and STL8 shows that both approaches yield very similar results, except for system NREL1. Here, STL7 overestimates the $\overline{PLR}$ 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. Apart from LS-LR8, Prophet1 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 as well as LS-LR7, three methodologies with high deviations from the $\overline{PLR}$. For all four methods, filter #2 from Table 1 and LS-LR as statistical method are used together with

different Predicted Power models. Thereby, 6k, PVUSA as well as 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 7 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 suitable statistical methods are recommended. Although LS-LR did yield some good results, 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 suitable filters and metrics (performance ratio or power) have been applied. And finally temperature correction of the chosen power metric is beneficial.

3.5 New opportunities from analysis of time-series *I-V*, P_{mpp} datasets

Recently with advances in data science and distributed computing it has become possible to analyze very large time series datasets of combined *I-V* and P_{mpp} data. For example 3.6 million *I-V* curves and 42 million P_{mpp} datapoints from 8 modules, of two different brands of which one is a glass/backsheet module and the other brand is a double glass module, over an 8 year period. First the performance *PLR* for each module is calculated using the DbD + UTC predicted power metric and the YoY regression statistical model on the outdoor time-series max power (P_{mpp}) data, the plane of array irradiance (POA) and the module temperature, [71]. In this research, the *PLR* results show that for both brands of modules, the BSh climate zone is more aggressive than the BWh climate zone, and both of them are much more aggressive than the ET climate zone. And the two modules brands show similar *PLR*s with only an apparent difference in the BWh climate zone, where the double glass module architecture shows a lower *PLR*. The time-series *I-V* curves are then analyzed to give us direct insight into the activated degradation mechanisms for these modules such as uniform current loss, corrosion, shunting and so on, rather than a generalized performance loss rate from analysis of P_{mpp} time-series data.

Two approaches for studying time series I-V curves have been demonstrated in two papers, both of which require extraction of I-V features from I-V curves[95]. The first approach[96] is similar to the analysis of time series P_{mpp}, but instead, the predicted I-V features are obtained at fixed conditions for comparison using purely data-driven models which are different for each I-V feature, thereby making the degradation behavior and loss factors comparable between different modules, and then modeling these time series I-V features change over time. The second approach[97,98] uses an outdoor I_{sc}-V_{oc} , or Suns-V_{oc}, approach to analyze the time series *I-V* curves, so as to construct outdoor I_{sc}-V_{oc} curves from *I-V* curves measured as a function of irradiance throughout the day, and then using a power loss, or loss factor, conversion to determine the loss factors associated with different degradation mechanisms[97]. The results show strong climate zone dependent degradation behavior. More detailed outdoor Suns-V_{oc} and power loss factors approach is in M. Wang's paper[98].

3.6 The role of *PLR* in PV system long term yield assessments

Long term changes in system performance are still among the most unexplored effects occurring in PV systems. From a survey regarding degradation effects in long term yield predictions (LTYP) carried out among PVPS Task 13 experts, a large variety of assumptions was documented. A participant used variable degradation rate for the first 5 years, then fixed from years 5–30. Generally, they assumed a 1 % to 2 % drop in the first year, 0.7 % to 0.5 % to year 5, then 0.3 % to 0.5 % up to year 30.

Some Yield Assessors also add an initial light induced degradation (LID) term of 0.3 % to 1.0 %, depending on module technology and location (increased degradation caused by unfavorable environmental conditions). In other cases, the value considered for the overall *PLR* is quite low using 0.25 %/a for crystalline Silicon based PV systems. Values are based on extensive (publicly available) literature reviews, but there is no consensus on what values should be used as a simple proxy for actual PLR determination.

The overall uncertainty associated with the *PLR* is a combination of the uncertainty related to the performance metric, which is a result of the uncertainty of the field measurements, and to the application of the data cleaning, filtering, and the metric and statistical modeling method as we have shown in this report.

3.7 Critical factors in *PLR* determination

Based on the results of this *PLR* benchmarking study, one uniform “best practice” approach for calculation of reliable *PLR* values, including choices of the “best” filtering approaches, the “best” metric and the “best” 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 quality related issues. In terms of calculation approaches, the most popular ones, STL and YoY, but also newly developed or less common ones, such as the VAR method, HW or Prophet, demonstrated reliable results. The complexity of the data and analysis interactions, may suggest that using many filter-metric-method choices, and determining the voted mean *PLR* ($\overline{PLR}$) may at this point produce the most reproducible and accurate results.

While we cannot recommend a single method to apply to any system, there are steps that can be taken to improve *PLR* calculation and reporting. First, time series features including, outliers, gaps, seasonality, etc. are common in PV data. Proper exploratory data analysis (EDA) before applying any *PLR* evaluation methodology can help a researcher identify and be alerted to potential sources of data analysis bias. Useful tools for PV data EDA are heatmaps, correlation matrices, and system dataset grading. Second, high irradiance filters show a tendency to reflect STC degradation in PV modules, while lower irradiance filters tend to reflect energy degradation. It is up to the individual to determine their modelling requirements. Third, it has been demonstrated (specifically in the low degradation digital power plants) that a universal temperature correction (such as in XbX+UTC or PR_{Tcorr}) provides a better extrapolation of temperature trends in PV power output than modelling within discrete time segments, (such as in XbX or PVUSA approaches). Discrete temperature corrections may introduce additionally yearly seasonality into a corrected PV timeseries trend. Lastly, proper reporting of *PLR* uncertainty helps present a more complete view of the performance of a system. *PLR* uncertainty can be evaluated through regression fitting, year-on-year distributions, multiple-method voting or bootstrap resampling of results, however with bootstrap uncertainties can be lower since resampling can’t capture the total variance. Individual *PLR* uncertainties should be reported

as 95 % confidence intervals. When comparing *PLR* values for two or more PV systems using the 83.4 % confidence intervals enables visualization of the *PLR* uncertainty of multiple systems at a significance level of 5 %. In cases where seasonal decomposition is used, the residual component should be added back into the trend component to prevent artificial smoothing of results and reporting artificially low uncertainties. Existing *PLR* evaluation software, including RdTools[19] and PVplr[21] have existing functionality for calculating *PLR* uncertainty.

4 CONCLUSIONS

The Performance Loss Rate (*PLR*) of a research or commercial PV power plant system quantifies the decline of the power output over time either as a single assumed linear rate in units of %/a, or %/year, or more recently as a rate over multiple time segments over the lifetime of the system. The *PLR* captures both irreversible physical degradation of PV modules and performance-reducing events, which may be reversible or even preventable through good O&M practices. The goal of this report is to define a framework of analytical steps for *PLR* determination, and assess the reliability and reproducibility of these steps and approaches when applied to datasets of both research and commercial PV systems. With this goal, we undertook an extensive Interlaboratory comparison, or benchmarking exercise on 19 real and 4 simulated PV systems, and used multiple filtering, metrics and statistical methods, to find by voting, or preference aggregation, the mean *PLR* (($\overline{PLR_i}$) of these 23 systems. This also enabled us to identify critical aspects that play important roles in PLR determination. These include the critical importance of exploratory data analysis (EDA) to quantify and grade the statistical dataset characteristics of each system; this turned out to be a largely neglected aspect that actually determines the quality of the reported *PLR*.

We establish a framework for *PLR* determination consisting of four basic steps common to all group's *PLR* analyses. The four steps are 1) input data cleaning and filtering, 2) performance metric selection (performance ratio (*PR*) or predicted power (*P*) based), corrections and data aggregation, 3) time series feature corrections and finally 4) application of a statistical modeling methods to determine the Performance Loss Rate value and its uncertainty. The statistical methods are evaluated in terms of a) their deviation from the mean *PLR* value and b) their confidence intervals. Using the one digital power plant and one real PV system we compared the impact of ten different filtering approaches with an otherwise identical *PLR* calculation approach. Filtering can be divided into two categories, threshold filters and statistical filters used to remove outliers in power-irradiance pairs. High irradiance threshold filters tend to lower the reported *PLR* which is not necessarily representative of real system performance. Additionally, statistical filtering approaches which remove the anomalous power-irradiance data pairs, in combination with a low to medium irradiance thresholds (to retain a larger amount of the system's data) provide the most reliable datasets for the next steps in *PLR* determination and consequently result with the highest accuracy results.

Based on the results of our Interlaboratory benchmarking results, one uniform way of calculating a reliable *PLR* value of a PV system, including prescriptions for a fixed filtering approaches, the same metric and statistical method, does not seem to exist at this time. Instead, it was shown that a thorough data quality check together with careful filtering approaches are absolutely crucial steps in calculating *PLR*; since as we know PV system datasets are subject to broad range of data related issues such missingness, gaps, offsets etcetera. In terms of calculation approaches, the most popular ones, STL and YoY, and also newly developed or less common ones, such as the XBX+UTC, VAR, HW or Prophet, demonstrated reliable results. While a standardized way of calculating *PLR* would be the desirable outcome of this study in order to reliably inter-compare results across PV systems and operators, it was highlighted that a sensitive combination of filtering / metrics and statistical methods is an important step forward. Often, dataset EDA is neglected, and the filtering step is either performed insufficiently or not explicitly reported in the corresponding literature. Therefore, when calculating and reporting *PLR*, an exhaustive report on filter selection and data cleaning is vital to better com-

prehend the steps in the *PLR* calculation. In addition reported *PLR* values need to be reproducible by others, and have clearly reported confidence intervals, so that results among systems are comparable at a 5 % significance level.

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. By calculating *PLR* using many filters, performance metrics corrections and data aggregation, corrections and statistical modeling approaches does appear to provide consistent and robust estimates of $\overline{PLR_i}$ for PV 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 $\overline{PLR_i}$.

REFERENCES

[1] S. Lindig et al., “International collaboration framework for the calculation of performance loss rates: Data quality, benchmarks, and trends (towards a uniform methodology),” Progress in Photovoltaics: Research and Applications, doi: 10.1002/pip.3397.

[2] R. G. Ross Jr., “Flat-plate photovoltaic array design optimization,” in IEEE 14th PVSC, San Diego CA USA, 1980, pp. 1126–1132 [Online]. Available: http://adsabs.harvard.edu/abs/1980pvsp.conf.1126R.

[3] J. A. Kratochvil, W. E. Boyson, and D. L. King, “Photovoltaic array performance model.,” Sandia National Laboratories, SAND2004-3535, Aug. 2004 [Online]. Available: https://www.osti.gov/biblio/919131.

[4] M. Prilliman, J. S. Stein, D. Riley, and G. Tamizhmani, “Transient Weighted Moving-Average Model of Photovoltaic Module Back-Surface Temperature,” IEEE J. Photovoltaics, pp. 1–8, 2020, doi: 10.1109/JPHOTOV.2020.2992351 .

[5] “IEC 61724-1:2017: Photovoltaic system performance - Part 1: Monitoring.,” IEC, 2017 [Online]. Available: https://webstore.iec.ch/publication/33622.

[6] A. Phinikarides, N. Kindyni, G. Makrides, and G. E. Georghiou, “Review of photovoltaic degradation rate methodologies,” *Renewable and Sustainable Energy Reviews*, vol. 40, pp. 143–152, Dec. 2014, doi: 10.1016/j.rser.2014.07.155.

[7] S. Lindig, I. Kaaya, K.-A. Weiss, D. Moser, and M. Topic, “Review of Statistical and Analytical Degradation Models for Photovoltaic Modules and Systems as Well as Related Improvements,” IEEE J. Photovoltaics, vol. 8, no. 6, pp. 1773–1786, Nov. 2018, doi: 10.1109/JPHOTOV.2018.2870532.

[8] P. Ingenhoven, G. Belluardo, and D. Moser, “Comparison of Statistical and Deterministic Smoothing Methods to Reduce the Uncertainty of Performance Loss Rate Estimates,” IEEE Journal of Photovoltaics, vol. 8, no. 1, pp. 224–232, Jan. 2018, doi: 10.1109/JPHOTOV.2017.2762523.

[9] A. Livera et al., “Data processing and quality verification for improved photovoltaic performance and reliability analytics,” Prog Photovolt Res Appl, p. pip.3349, Oct. 2020, doi: 10.1002/pip.3349.

[10] D. L. King, W. E. Boyson, and J. A. Kratochvil, “Photovoltaic array performance model.,” Sandia National Laboratories, SAND2004-3535, Aug. 2004 [Online]. Available: https://www.osti.gov/biblio/919131.

[11] T. Huld, G. Friesen, A. Skoczek, "A power-rating model for crystalline silicon PV modules," Sol Energy Mater and Cells. Vol. 95, no.12, pp. 3359-3369, 2011, doi: 10.1016/j.solmat.2011.07.026.

[12] S. van Buuren and K. Groothuis-Oudshoorn, “mice: Multivariate Imputation by Chained Equations in R,” Journal of Statistical Software, vol. 45, no. 1, pp. 1–67, Dec. 2011, doi: 10.18637/jss.v045.i03.

[13] S. Lindig, A. Louwen, D. Moser, and M. Topic, “Outdoor PV System Monitoring—Input Data Quality, Data Imputation and Filtering Approaches,” Energies, vol. 13, no. 19, p. 5099, Jan. 2020, doi: 10.3390/en13195099.

[14] T. Carlsson, K. Åström, P. Konttinen, and P. Lund, “Data filtering methods for determining performance parameters in photovoltaic module field tests,” *Progress in Photovoltaics: Research and Applications*, vol. 14, no. 4, pp. 329–340, 2006, doi: 10.1002/pip.669.

[15] D. C. Jordan and S. R. Kurtz, “The Dark Horse of Evaluating Long-Term Field Performance—Data Filtering,” *IEEE Journal of Photovoltaics*, vol. 4, no. 1, pp. 317–323, Jan. 2014, doi: 10.1109/JPHOTOV.2013.2282741.

[16] M. B. Øgaard, H. N. Riise, H. Haug, S. Sartori, and J. H. Selj, “Photovoltaic system monitoring for high latitude locations,” *Solar Energy*, vol. 207, pp. 1045–1054, Sep. 2020, doi: 10.1016/j.solener.2020.07.043.

[17] M. J. Reno and C. W. Hansen, "Identification of periods of clear sky irradiance in time series of GHI measurements," Renewable Energy, vol. 90, pp. 520–531, May 2016, doi: 10.1016/j.renene.2015.12.031.

[18] W. Holmgren, et al., *pvlib/pvlib-python: v0.8.0*. Zenodo, 2020 [Online]. Available: https://zenodo.org/record/4019830.

[19] M. G. Deceglie, D. Jordan, A. Nag, C. A. Deline, and A. Shinn, "RdTools: An Open Source Python Library for PV Degradation Analysis," National Renewable Energy Lab. (NREL), Golden, CO (United States), NREL/PR-5K00-71468, May 2018 [Online]. Available: https://www.osti.gov/biblio/1436856.

[20] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV Degradation Methodology and Application," IEEE Journal of Photovoltaics, vol. 8, no. 2, pp. 525–531, Mar. 2018, doi: 10.1109/JPHOTOV.2017.2779779.

[21] A. J. Curran, T. J. Burleyson, S. Lindig, D. Moser, R. H. French, *PVplr: Performance Loss Rate Analysis Pipeline*. 2020. Available: https://CRAN.R-project.org/package=PVplr.

[22]D. M. Diez, M. Çetinkaya-Rundel, and C. D. Barr, OpenIntro Statistics: Fourth Edition, 4th Edition. S.l.: OpenIntro, Inc., 2019 [Online]. Available: https://www.openintro.org/book/stat/.

[23] M. Theristis, A. Livera, L. Micheli, and C. B. Jones, "Modeling nonlinear photovoltaic degradation rates," IEEE PVSC 47 Proceedings, vol. 47, p. 6, 2020.

[24] A. J. Curran, C. B. Jones, S. Lindig, J. Stein, D. Moser, and R. H. French, "Performance Loss Rate Consistency and Uncertainty Across Multiple Methods and Filtering Criteria," in 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), Chicago, IL, USA, 2019, pp. 1328–1334, doi: 10.1109/PVSC40753.2019.8980928.

[25] A. J. Curran, Y. Hu, R. Haddadian, J. L. Braid, D. Meakin, T. J. Peshek, and R. H. French, "Determining the Power Rate of Change of 353 Plant Inverters Time-Series Data Across Multiple Climate Zones, Using a Month-By-Month Data Science Analysis," in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)*, 2017, pp. 1927–1932, doi: 10.1109/PVSC.2017.8366477.

[26] D. C. Jordan, M. G. Deceglie, and S. R. Kurtz, "PV degradation methodology comparison — A basis for a standard," in 2016 IEEE 43rd Photovoltaic Specialists Conference (PVSC), 2016, pp. 0273–0278, doi: 10.1109/PVSC.2016.7749593.

[27] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV Degradation Methodology and Application," IEEE Journal of Photovoltaics, vol. 8, no. 2, pp. 525–531, Mar. 2018, doi: 10.1109/JPHOTOV.2017.2779779.

[28] D. L. King, J. A. Kratochvil, and W. E. Boyson, "Field experience with a new performance characterization procedure for photovoltaic arrays," Sandia National Labs., Albuquerque, NM (US), SAND-98-3147C; CONF-980735-, Dec. 1997 [Online]. Available: https://www.osti.gov/biblio/629484.

[29] T. Huld, et al., "A power-rating model for crystalline silicon PV modules," *Solar Energy Materials and Solar Cells*, vol. 95, no. 12, pp. 3359–3369, Dec. 2011, doi: 10.1016/j.solmat.2011.07.026.

[30] A. P. Dobos, "PVWatts Version 5 Manual," National Renewable Energy Lab. (NREL), Golden, CO (United States), NREL/TP-6A20-62641, Sep. 2014 [Online]. Available: https://www.osti.gov/biblio/1158421.

[31] W. F. Holmgren, C. W. Hansen, and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," JOSS, vol. 3, no. 29, p. 884, Sep. 2018, doi: 10.21105/joss.00884.

[32] S. Lindig, D. Moser, B. Muller, and M. Topic, "Application of Dynamic Multi-Step Performance Loss Algorithm," IEEE PVSC 47, p. 6, 2020.

[33] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning: with Applications in R, 1st ed. 2013, Corr. 5th printing 2015 edition. New York: Springer, 2013 [Online]. Available: http://www-bcf.usc.edu/gareth/ISL/index.html.

[34] J. J. Faraway, Linear Models with R, 2nd Ed., Boca Raton: Chapman and Hall/CRC, 2014. https://www.routledge.com/Linear-Models-with-R/Faraway/p/book/9781439887332.

[35] R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 2nd Ed., 2 edition. Heathmont, Victoria: OTexts, 2018 [Online]. Available: https://otexts.com/fpp2/.

[36] D. C. Jordan, S. R. Kurtz, Analytical improvements in PV degradation rate determination. In: 35th IEEE Photovoltaic Specialists Conference; 2010. p.688–693, doi: 10.1109/PVSC.2010.5617074.

[37] R. B. Cleveland, J. E. McRae, and I. Terpenning, "STL: A Seasonal-Trend Decomposition Procedure Based on Loess," Journal of Official Statistics; Stockholm, vol. 6, no. 1, pp. 3–73, 1990. https://search.proquest.com/openview/cc5001e8a0978a6c029ae9a41af00f21/1?cbl=105444&pq.

[38] R Core Team, R: The R Project for Statistical Computing. 2019 [Online]. Available: https://www.r-project.org/. [Accessed: 12-Jan-2016]

[39] R. Hafen, stlplus: Enhanced Seasonal Decomposition of Time Series by Loess. 2016 [Online]. Available: https://CRAN.R-project.org/package=stlplus. [Accessed: 19-Jun-2020]

[40] W. S. Cleveland, "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, vol. 74, no. 368, p. 829, Dec. 1979, doi: 10.2307/2286407.

[41] W. S. Cleveland and S. J. Devlin, "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *Journal of the American Statistical Association*, vol. 83, no. 403, pp. 596–610, 1988, doi: 10.2307/2289282.

[42] *R: The R Stats Package*. 2020 [Online]. Available: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/stats-package.html.

[43] E. Rauch, *rstl: A Python port of R's stl function*. [Online]. Available: https://github.com/ericist/rstl.

[44] *statsmodels: Statistical computations and models for Python*. [Online]. Available: https://www.statsmodels.org/. [Accessed: 31-Oct-2020]

[45] S. Seabold and J. Perktold, "statsmodels: Econometric and statistical modeling with python," in *Proc. of the 9th Python in Science Conf. (SCIPY 2010)*, Austin, Texas, 2010, pp. 92–96, doi: 10.25080/Majora-92bf1922-012 Available: http://conference.scipy.org/proceedings/scipy2010/.

[46] E. Hasselbrink, et al., "Validation of the PVLife model using 3 million module-years of live site data," in *IEEE 39th photovoltaic specialists conference (PVSC)*, 2013, pp. 0007–0012, doi: 10.1109/PVSC.2013.6744087

[47] B. Efron and R. J Tibshirani, An introduction to the bootstrap, vol. 57. Chapman & Hall/CRC Monographs on Statistics and Applied Probability, 1993. Available: https://www.routledge.com/An-Introduction-to-the-Bootstrap/Efron-Tibshirani/p/book/9780412042317.

[48] B. Efron and R. Tibshirani, "Improvements on Cross-Validation: The 632+ Bootstrap Method," Journal of the American Statistical Association, vol. 92, no. 438, pp. 548–560, Jun. 1997, doi: 10.1080/01621459.1997.10474007.

[49] M. Meftah, E. Lajoie-Mazenc, M. Van Iseghem, R. Perrin, D. Boublil, and K. Radouane, "A Less Environment-Sensitive and Data-Based Approach to Evaluate the Performance Loss Rate of PV Power Plants," in 36th European Photovoltaic Solar Energy Conference, 2019, pp. 1554–1559, doi: 10.4229/EUPVSEC20192019-5CV.4.3

[50] B. Meyers, M. Tabone, and E. C. Kara, “Statistical Clear Sky Fitting Algorithm,” arXiv:1907.08279 [cs, eess], Jul. 2019 [Online]. Available: http://arxiv.org/abs/1907.08279.

[51] B. Meyers, M. Deceglie, C. Deline, and D. Jordan, “Signal Processing on PV Time-Series Data: Robust Degradation Analysis Without Physical Models,” IEEE Journal of Photovoltaics, vol. 10, no. 2, pp. 546–553, Mar. 2020, doi: 10.1109/JPHOTOV.2019.2957646.

[52] C. C. Holt, “Forecasting seasonals and trends by exponentially weighted moving averages,” International Journal of Forecasting, vol. 20, no. 1, pp. 5–10, Jan. 2004, doi: 10.1016/j.ijforecast.2003.09.015.

[53] S. Taylor and B. Letham, “prophet: Automatic forecasting procedure,” manual, 2020. Available: https://CRAN.R-project.org/package=prophet

[54] S. J. Taylor and B. Letham, “Forecasting at Scale,” *The American Statistician*, vol. 72, no. 1, pp. 37–45, Jan. 2018, doi: 10.1080/00031305.2017.1380080.

[55] M. Theristis, A. Livera, C. B. Jones, G. Makrides, G. E. Georghiou, and J. S. Stein, “Nonlinear Photovoltaic Degradation Rates: Modeling and Comparison Against Conventional Methods,” IEEE Journal of Photovoltaics, vol. 10, no. 4, pp. 1112–1118, Jul. 2020, doi: 10.1109/JPHOTOV.2020.2992432.

[56] P. Fearnhead and G. Rigaill, “Changepoint Detection in the Presence of Outliers,” *Journal of the American Statistical Association*, vol. 114, no. 525, pp. 169–183, Jan. 2019, doi: 10.1080/01621459.2017.1385466.

[57] C. R. Osterwald, J. Adelstein, J. A. del Cueto, B. Kroposki, D. Trudell, and T. Moriarty, “Comparison of Degradation Rates of Individual Modules Held at Maximum Power,” in *2006 IEEE 4th World Conference on Photovoltaic Energy Conference*, 2006, vol. 2, pp. 2085–2088, doi: 10.1109/WCPEC.2006.279914.

[58] G. Makrides, M. Theristis, J. Bratcher, J. Pratt, and G. E. Georghiou, “Five-year performance and reliability analysis of monocrystalline photovoltaic modules with different backsheet materials,” *Solar Energy*, vol. 171, pp. 491–499, Sep. 2018, doi: 10.1016/j.solener.2018.06.110.

[59] S. Lindig, D. Moser, A. J. Curran, and R. H. French, “Performance Loss Rates of PV systems of Task 13 database,” in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, Chicago, IL, USA, 2019, pp. 1363–1367, doi: 10.1109/PVSC40753.2019.8980638.

[60] M. Kottek J. Grieser, C. Beck, B. Rudolf, F. Rubel. World Map of the Köppen-Geiger Climate Classification Updated. Meteorologische Zeitschrift 2006 May; 15: 259–263, doi: 10.1127/0941-2948/2006/0130.

[61] F. Rubel, M. Kottek. Observed and Projected Climate Shifts 1901-2100 Depicted by World Maps of the Koppen-Geiger Climate Classification. Meteorologische Zeitschrift 2010 Apr; 19(2): 135–141, doi: 10.1127/0941-2948/2010/0430.

[62] F. Rubel, K. Brugger, K. Haslinger, I. Auer. The Climate of the European Alps: Shift of Very High Resolution Köppen- Geiger Climate Zones1800–2100. Meteorologische Zeitschrift 2017, doi: 10.1127/metz/2016/0816

[63] C. Bryant, N. R. Wheeler, F. Rubel, R. H. French. kgc: Köppen-Geiger Climatic Zones; 2017, https://CRAN.R-project.org/package=kgc, R package version 1.0.0.2.

[64] J. Ascencio-Vásquez, K. Brecl, and M. Topič, “Methodology of Köppen-Geiger-Photovoltaic climate classification and implications to worldwide mapping of PV system performance,” Solar Energy, vol. 191, pp. 672–685, Oct. 2019, doi: 10.1016/j.solener.2019.08.072.

[65] J. Ascencio-Vásquez, I. Kaaya, K. Brecl, K.-A. Weiss, and M. Topič, “Global Climate Data Processing and Mapping of Degradation Mechanisms and Degradation Rates of PV Modules,” *Energies*, vol. 12, no. 24, p. 4749, Jan. 2019, doi: 10.3390/en12244749.

[66] S. Lindig, A. Curran, K. Rath, A. Khalilnejad, D. Moser, and R. H. French, "IEA PVPS Task 13-ST2.5: *PLR* Determination Benchmark Study." Case Western Reserve University, 30-Aug-2020 [Online]. Available: https://osf.io/vtr2s/. [Accessed: 04-Oct-2020]

[67] L. Burnham, B. H. King, C. Deline, S. Barkaszi, A. Sahm, and J. Stein, "The US DOE Regional Test Center Program: Driving Innovation Quality and Reliability.," Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 1279686, 2015 [Online]. Available: https://www.osti.gov/servlets/purl/1279686.

[68] A. Curran, C. B. Jones, J. L. Braid, J. S. Stein, and R. H. French, "US DOE-RTC-Baseline Systems." 29-Mar-2019 [Online]. Available: https://osf.io/yvzhk/. [Accessed: 05-Jul-2019]

[69] B. Braisaz, P. Dupeyrat, A. Lindsay, and K. Radouane, "An Advanced Model of PV Power Plants Based on Modelica®," in EU PVSEC Proceedings, 2013, pp. 3644–3648, doi: 10.4229/28thEUPVSEC2013-5AO.9.4.

[70] "HelioClim-3 solar radiation database." [Online]. Available: http://www.soda-pro.com/home . [Accessed: 10-Nov-2020]

[71] J. Liu, et al., "Real-world PV Module Degradation across Climate Zones Determined from Suns-Voc, Loss Factors and I-V Steps Analysis of Eight Years of I-V, Pmp Time-series Datastreams," in *2019 IEEE 46th Photovoltaic Specialists Conference (PVSC)*, 2019, pp. 0680–0686, doi: 10.1109/PVSC40753.2019.8980541.

[72] ET Instrumente, "Series ESL-Solar 500," *ET Instrumente*, 2018. [Online]. Available: https://et-instrumente.de/index.php/en/products2/overview-electronic-loads/pv-modul-test-unit. [Accessed: 18-Nov-2018]

[73] D. C. Jordan and S. R. Kurtz, "Photovoltaic Degradation Rates-an Analytical Review: Photovoltaic degradation rates," *Progress in Photovoltaics: Research and Applications*, vol. 21, no. 1, pp. 12–29, Jan. 2013, doi: 10.1002/pip.1182.

[74] D. C. Jordan, S. R. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates: Photovoltaic degradation rates," Prog. Photovolt: Res. Appl., vol. 24, no. 7, pp. 978–989, Jul. 2016, doi: 10.1002/pip.2744.

[75] S. K. Kwak and J. H. Kim, "Statistical data preparation: management of missing values and outliers," Korean J Anesthesiol, vol. 70, no. 4, pp. 407–411, Aug. 2017, doi: 10.4097/kjae.2017.70.4.407.

[76] N. J. Tierney and D. H. Cook, "Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations," arXiv:1809.02264 [stat], May 2020 [Online]. Available: http://arxiv.org/abs/1809.02264.

[77] A. Khalilnejad, A. M. Karimi, S. Kamath, R. Haddadian, R. H. French, and A. R. Abramson, "Automated Pipeline Framework for Processing of Large-Scale Building Energy Time Series Data," *PLOS ONE*, vol. 15, no. 12, p. e0240461, Dec. 2020, doi: 10.1371/journal.pone.0240461. [Online]. Available: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0240461.

[78] S. K. Kwak and J. H. Kim, "Statistical data preparation: management of missing values and outliers," Korean J Anesthesiol, vol. 70, no. 4, pp. 407–411, Aug. 2017, doi: 10.4097/kjae.2017.70.4.407.

[79] C. C. Aggarwal, Outlier Analysis, 2nd ed. 2017 edition. New York, NY: Springer, 2016. https://www.springer.com/gp/book/9783319475776.

[80] V. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies," Artificial Intelligence Review, vol. 22, no. 2, pp. 85–126, Oct. 2004, doi: 10.1023/B:AIRE.0000045502.10941.a9. Available: https://doi.org/10.1023/B:AIRE.0000045502.10941.a9.

[81] C. Chen and L.-M. Liu, "Joint Estimation of Model Parameters and Outlier Effects in Time Series," Journal of the American Statistical Association, vol. 88, no. 421, pp. 284–297, 1993, doi: 10.2307/2290724.

[82] J. López-de-Lacalle, tsoutliers: Detection of Outliers in Time Series. 2019 [Online]. Available: https://CRAN.R-project.org/package=tsoutliers. [Accessed: 08-Aug-2020]

[83] A. Gelman and J. Hill, "Chapter 25. Missing Data Imputation," in Data Analysis Using Regression and Multilevel/Hierarchical Models, 1 edition., Cambridge; New York: Cambridge University Press, 2006, pp. 529-543. Available: https://www.cambridge.org/us/academic/subjects/statistics-probability/statistical-theory-and-methods/data-analysis-using-regression-and-multilevelhierarchical-models?format=PB.

[84] B. Horan, "Standard deviation, or standard error of the mean?," Anaesth Itens Care, vol. 10, no. 4, p. 297, 1982, doi: 10.1177/0310057X8201000430.

[85] D. Curran-Everett, "Explorations in statistics: standard deviations and standard errors," Advances in Physiology Education, vol. 32, no. 3, pp. 203–208, Sep. 2008, doi: 10.1152/advan.90123.2008.

[86] G. Cumming, J. Williams, and F. Fidler, "Replication and researchers' understanding of confidence intervals and standard error bars," Understanding Statistics, vol. 3, no. 4, pp. 299–311, 2004, doi: 10.1207/s15328031us0304_5.

[87] S. V. Muravyov and I. A. Marinushkina, "Processing Data from Interlaboratory Comparisons by the Method of Preference Aggregation," Meas Tech, vol. 58, no. 12, pp. 1285–1291, Mar. 2016, doi: 10.1007/s11018-016-0886-4.

[88] H. Goldstein and M. J. R. Healy, "The Graphical Presentation of a Collection of Means," Journal of the Royal Statistical Society. Series A (Statistics in Society), vol. 158, no. 1, p. 175, 1995, doi: 10.2307/2983411.

[89] D. Kalpić, N. Hlupić, and M. Lovrić, "Student's t-Tests," in International encyclopedia of statistical science, M. Lovric, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 1559–1563 [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_41

[90] M. E. Payton, M. H. Greenstone, and N. Schenker, "Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance?," J Insect Sci, vol. 3, Oct. 2003 [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC524673/.

[91] S. M. Kamath, "Energy Use Intensities across Building Use Types and Climate Zones using the CBECS dataset," Case Western Reserve University, 2020 [Online]. Available: https://etd.ohiolink.edu/pg_10?::NO:10:P10_ETD_SUBID:184439. [Accessed: 10-Jun-2020]

[92] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, 2nd ed. New York: Springer-Verlag, 2009. Available: https://web.stanford.edu/~hastie/ElemStatLearn/

[93] M. Wild, et al., "From Dimming to Brightening: Decadal Changes in Solar Radiation at Earth's Surface," *Science*, vol. 308, no. 5723, pp. 847–850, May 2005, doi: 10.1126/science.1103215.

[94] K. Kiefer, B. Farnung, B. Müller, K. Reinartz, I. Rauschen, C. Klünter. Degradation in PV Power Plants: Theory and Practice. In: 36th EUPVSEC Proceedings Marseille, France; 2019. p.9–13, doi: 10.4229/EUPVSEC20192019-5BO.7.5.

[95] W. Huang, et al., *ddiv: Data Driven I-V Feature Extraction*. 2018. Available: https://CRAN.R-project.org/package=ddiv.

[96] J. Liu, et al., "Cross-correlation Analysis of the Indoor Accelerated and Real World Exposed Photovoltaic Systems Across Multiple Climate Zones," in *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)*, 2018, pp. 3949–3954, doi: 10.1109/PVSC.2018.8547840.

[97] M. Wang, et al., *SunsVoc: Constructing Suns-Voc from Outdoor Time-series I-V Curves*. 2020. Available: https://CRAN.R-project.org/package=SunsVoc

[98] M. Wang, et al., “Evaluation of Photovoltaic Module Performance Using Novel Data-driven I-V Feature Extraction and Suns-VOC Determined from Outdoor Time-Series I-V Curves,” in *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)*, 2018, pp. 0778–0783, doi: 10.1109/PVSC.2018.8547772.

APPENDICES

A. Irradiance distribution for digital power plant (location: Rennes/France)

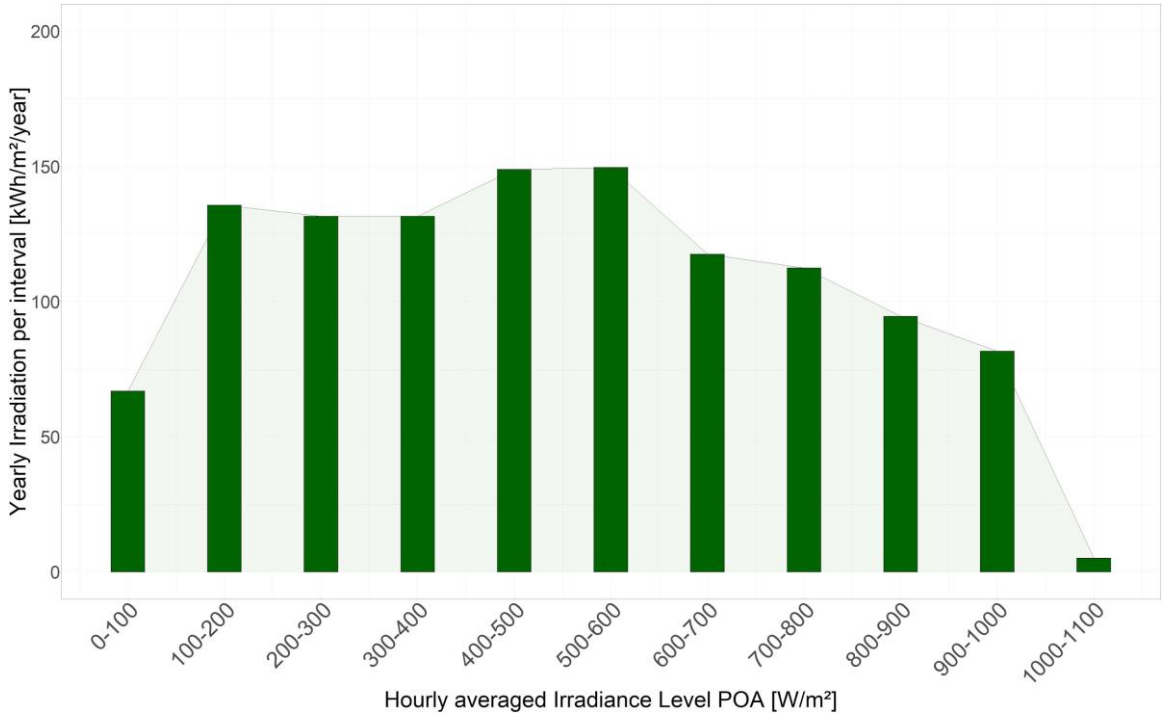


Figure 25: Plane-of-array (POA) irradiance distribution per 100W/m² interval[1].

B. Data quality issues of PV system datasets

B.1 US DOE RTC baseline PV systems

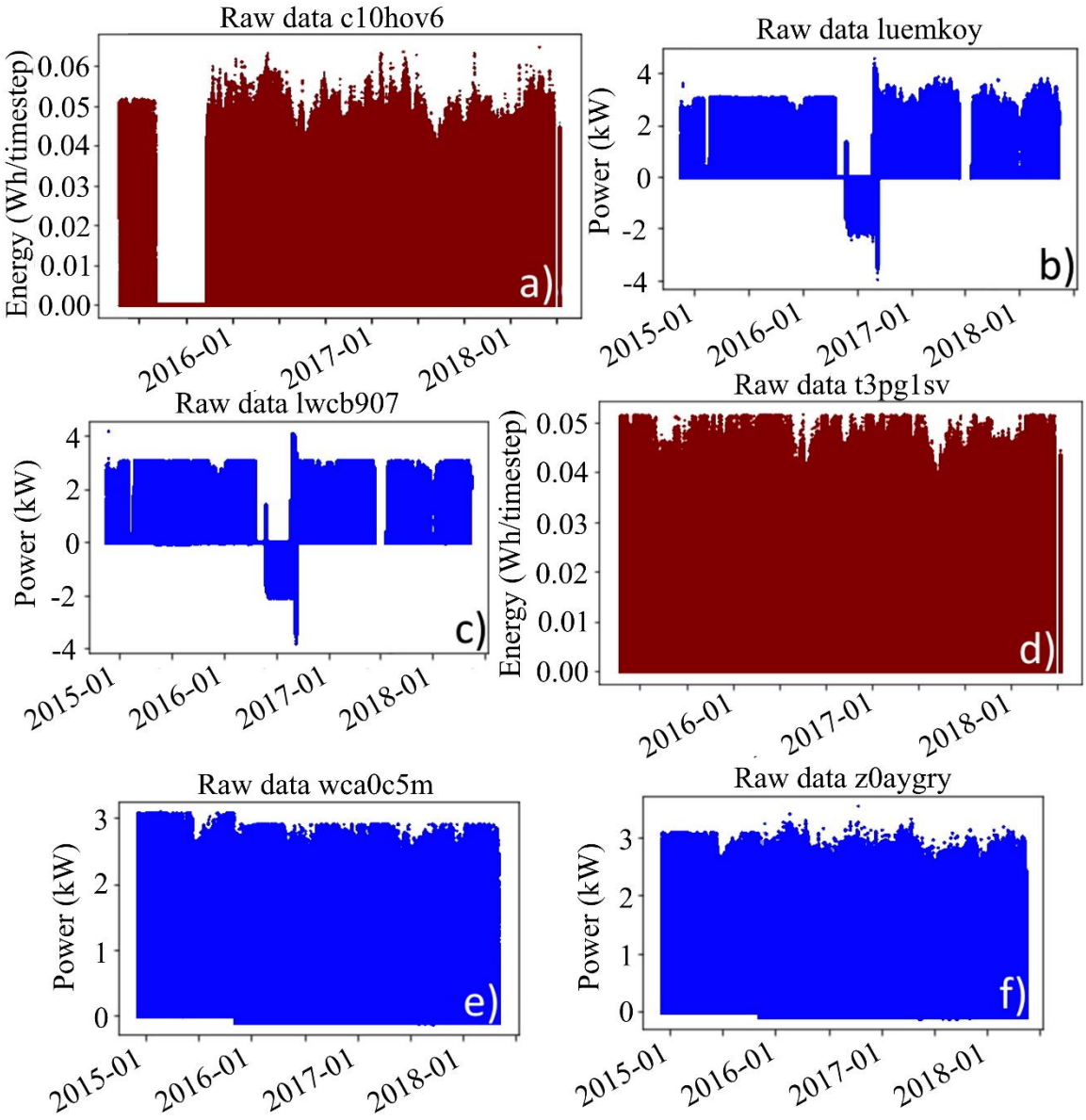


Figure 26: Energy (red) / power (blue) over time for US DOE datasets: a) US DOE c10hov6 - initial inverter clipping and missing data; b) US DOE luemkoy - negative power values and initial inverter clipping c) US DOE lwcb907 - negative power values and inverter clipping; d) US DOE t3pg1sv - inverter clipping; e) US DOE wca0c5m - data shift after 1 year and inverter clipping; f) US DOE z0aygry - data shift after 1 year and initial inverter clipping[1].

B.2 NREL PV systems

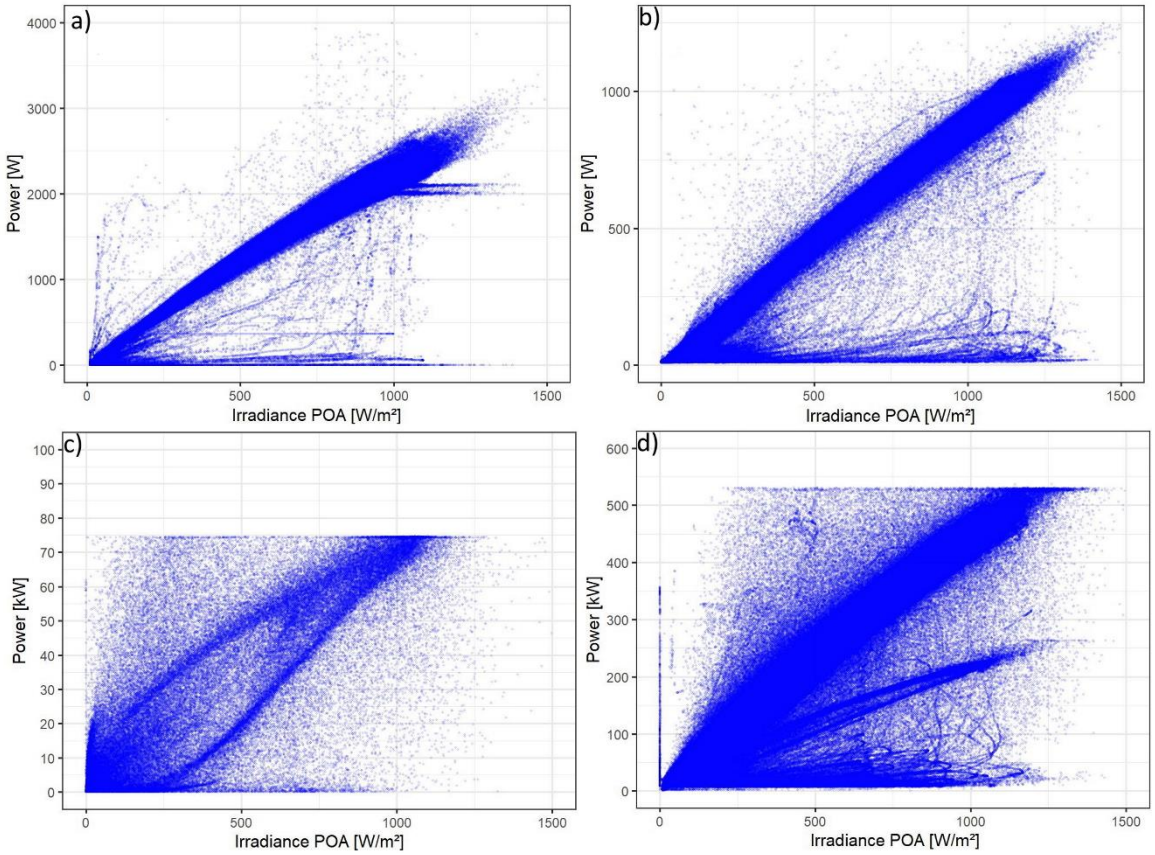


Figure 27: Power over plane-of-array irradiance for NREL datasets: a) measured irradiance - NREL1 - numerous outlier detected; b) measured irradiance - NREL2 - numerous outlier detected c) modelled irradiance - NREL3 - extreme outlier and inverter clipping detected; d) modelled irradiance - NREL4 - extreme outlier and inverter clipping detected[1].

C. *PLR* results for all PV systems

C.1 EURAC, FOSS, RSE, Pfaffstaetten systems

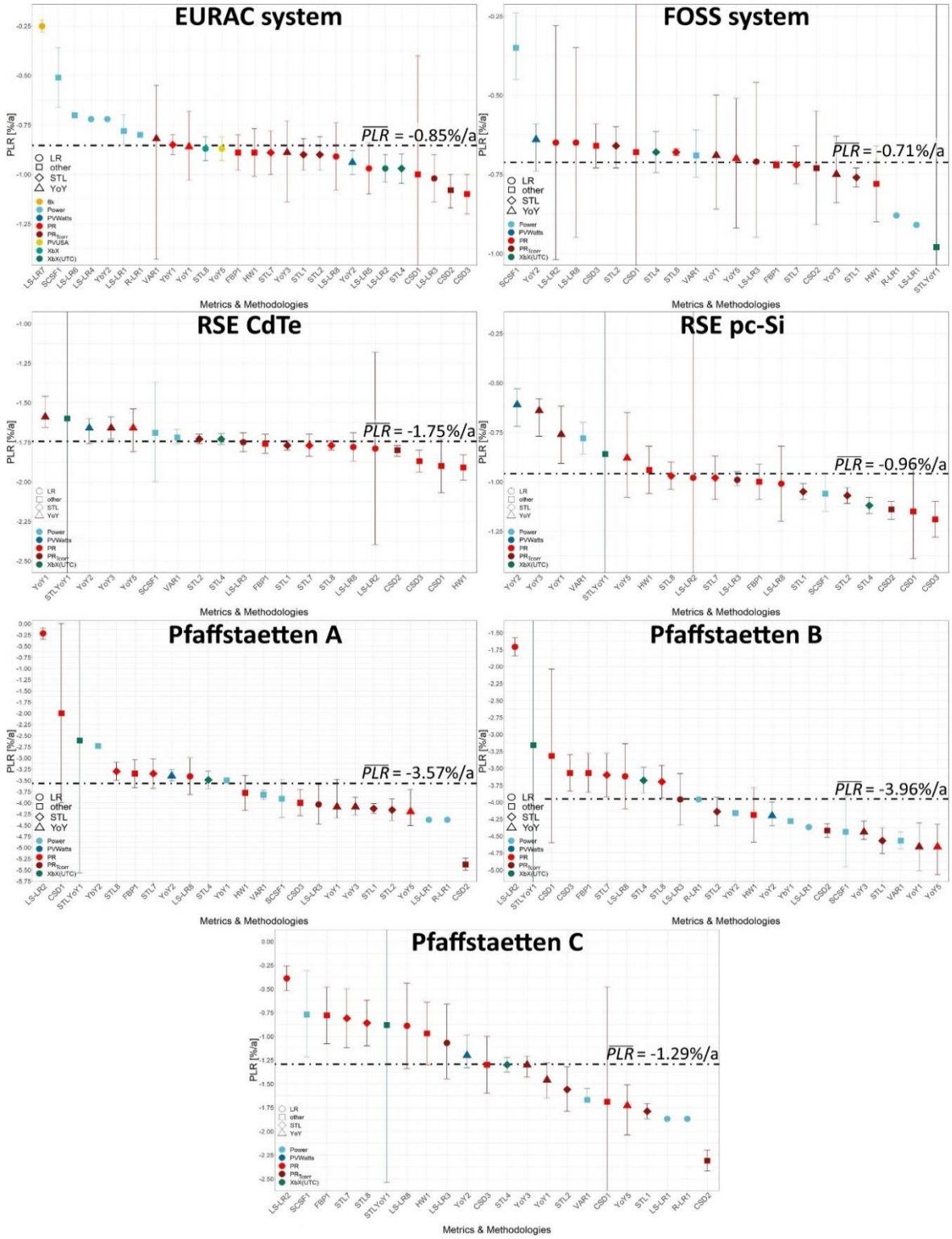


Figure 28: Calculated *PLR* of EURAC system; FOSS system, RSE CdTe & pc-Si systems; Pfaffstaetten A, B & C systems. The "error bars", are as reported by the analyst[1].

C.2 *PLR* results for NREL PV systems

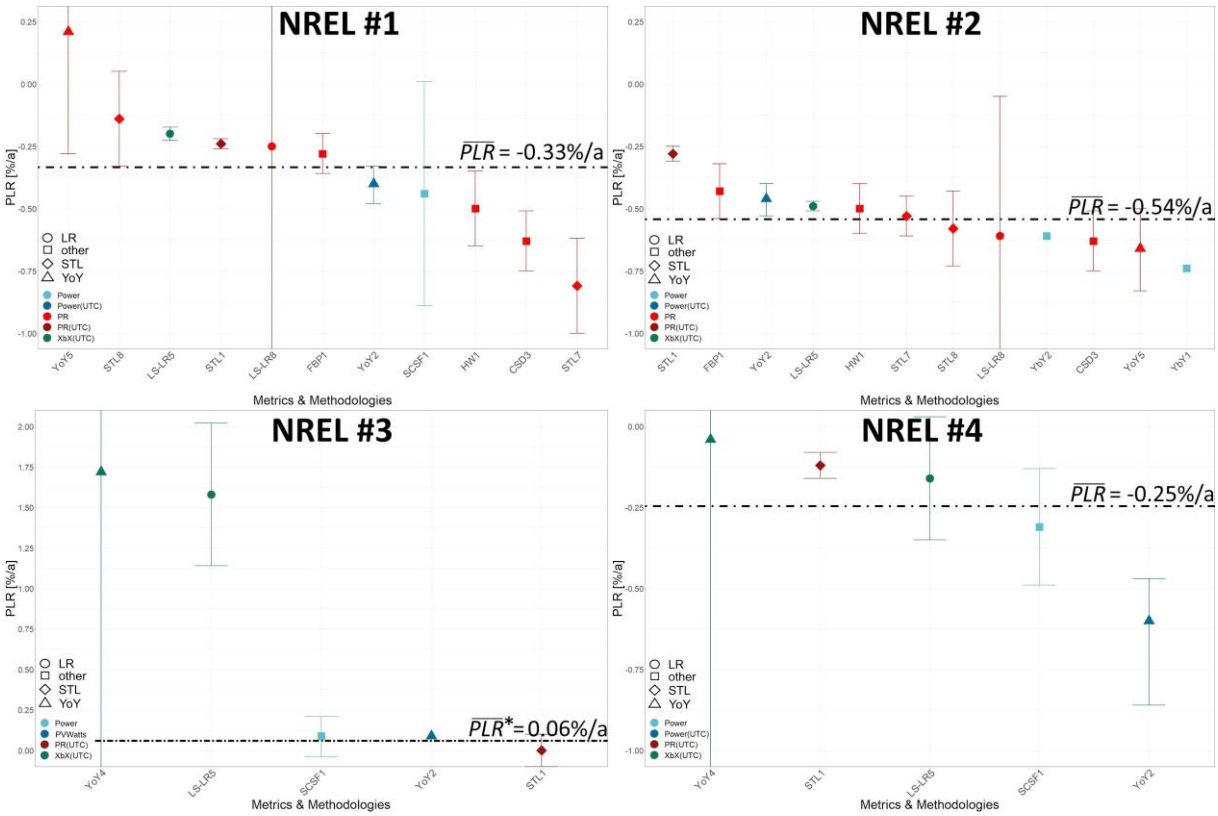


Figure 29: Calculated *PLR* of the NREL systems. The "error bars", are as reported by the analyst[1].

Here, the calculated *PLR* results 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 3.4.8. 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 $\overline{PLR}$ for the respective system. Exceptions are the US DOE systems wca0c5m and z0aygry as well as NREL3. For the US DOE systems, the $\overline{PLR}$ 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 3.4.2). This shift should have been detected and excluded for the *PLR* calculation. This was done only with statistical method STL1 as well as YoY2. For NREL3, the $\overline{PLR}$ corresponds to the average *PLR* of the methodologies SCSF1, YoY2 and STL1. YoY2 and STL1 used the provided modelled 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 inputs and the corresponding results deviate thereby substantially from the "true" *PLR*.

C.3 *PLR* results of US DOE RTC baseline PV systems

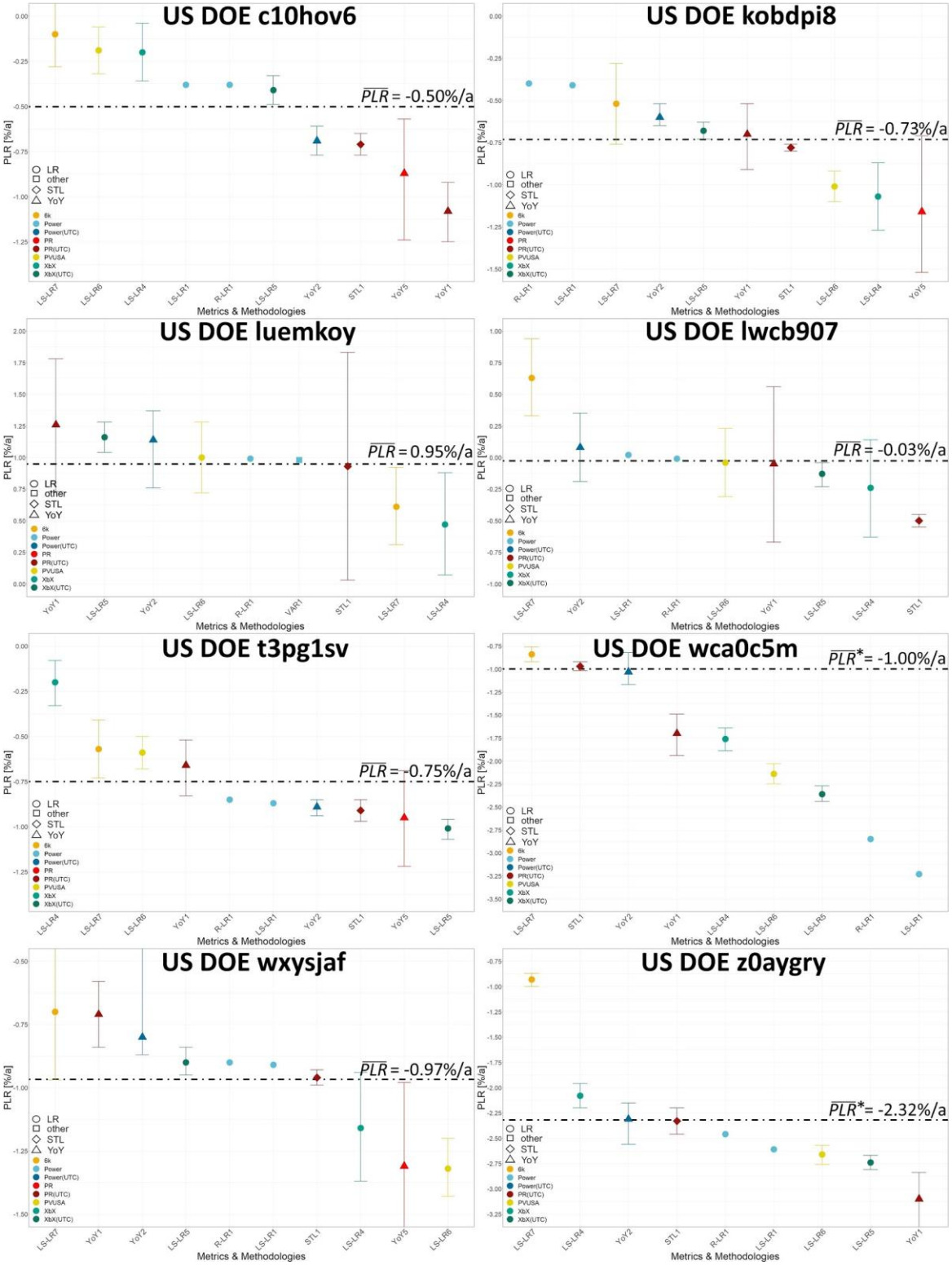


Figure 30: Calculated *PLR* of US DOE systems; $\overline{PLR}$ for US DOE wca0c5m & z0aygry corresponds to average *PLR* of STL1 and YoY2. The "error bars", are as reported by the analyst[1].

ISBN 978-3-907281-10-9