Performance Analysis and Degradation of a Large Fleet of PV Systems

Sascha Lindig, Julián Ascencio-Vásquez, Jonathan Leloux, David Moser, and Angèle Reinders

Abstract—**This article presents an initial performance analysis of a database of photovoltaic (PV) system performance time series collected within the European funded COST Action PEARL PV. The database contains monitoring data of over 8400 PV systems with accompanying metadata. The PV plants are small residential systems, primarily installed in Europe, with a high density in Belgium. In this initial study, the annual average performance ratio, the annual energy yield, and the performance loss rate of the systems are determined and evaluated. The systems have an average lifetime of 30.5 months. The annual mean performance ratio across all systems is 76.7% and the average yield is 954.9 kWh/kWp per year. The performance loss rate is calculated using three different statistical approaches and one irradiance data source. Average performance losses between −0.74%/year and −0.86%/year are calculated depending on the used approach. Furthermore, certain weather-dependent correlations are detected, such as decreasing performance ratio and increasing yield values with increasing irradiation. This study is a stepping-stone for further populating the present database, lessons learnt for handling large amounts of PV performance data, and carrying out performance studies of PV system fleets installed across Europe.**

Index Terms—**Big data analytics, degradation, monitoring, PV system fleet, PV systems, system performance.**

I. Introduction

COST Action PEARL PV aims at analyzing data of monitored photovoltaic (PV) systems installed mainly across Europe to quantitatively evaluate the long-term performance and reliability of these PV systems. For this purpose, a databank has been implemented in 2019 that can contain vast amounts of monitoring data of installed PV systems, which will enable systematic performance analyses in combination with simulation [1]–[3]. Data will be analyzed regarding the actual monitored long-term performance to quantitatively determine the absolute influences of:

1) geographic location and as such weather and climate factors;
2) key system design features such as system components rated performance;
3) installation types, including field-based solar parks, rooftop systems, BIPV, VIPV, and floating systems;
4) operation and maintenance practices; and
5) performance degradation over time and failure modes of these PV systems.

The results will be used to increase performance and lower costs of electricity produced by PV systems in Europe via i) obtaining higher energy yields, ii) achieving longer operational life time, and iii) lowering the perceived investment risk in PV projects.

The basic assumption underlying this project is that the analysis of long-term monitoring data—of 1 year or longer—of large amounts of PV systems, also called a fleet, will significantly enhance the statistics of PV performance analytics, which will designate statistically relevant factors which influence either system performance, degradation, or the occurrence of system failures. This approach would avoid detailed and hence time-consuming evaluations of each individual, unique PV system. In general a sample size (portion of the total population which will be studied) of 100 is required to obtain meaningful results [4], [5]. Otherwise stated, a batch of monitoring data of less than 100 PV systems automatically implies that all these PV systems must be analyzed separately, while results of systems above this number could be treated in a statistically relevant manner. Sometimes a small sample size is unavoidable, which will result in wide confidence intervals and risk of errors in hypothesis testing. Using basic statistics approaches, with a population of millions of PV systems installed all over Europe (in Germany only, in the period of 2009 till 2018, close to 1.1 million PV systems were installed) a sample size of approximately 400 PV systems would result in a confidence level of 95% and an error margin of only 5%, provided that the sample is representative without many outliers. Therefore, datasets of 1000 or more PV systems can yield statistically relevant results for the European situation. In this article, a dataset containing monitoring data of more than 8000 PV systems will be analyzed. This dataset contains, besides long-term monitoring data of power, metadata with PV system characteristics, such as location and nominal

Manuscript received March 24, 2021; revised May 10, 2021; accepted June 15, 2021. Date of publication July 30, 2021; date of current version August 20, 2021. This work was supported by European Cooperation in Science and Technology (COST). *(Corresponding author: Sascha Lindig.)*

Sascha Lindig is with EURAC Research, 39100 Bolzano, Italy, and also with the University of Ljubljana, Faculty of Engineering, 1000 Ljubljana, Slovenia (e-mail: sascha.lindig@eurac.edu).

Julián Ascencio-Vásquez is with the 3E sa, B-1000 Brussels, Belgium, and also with the University of Ljubljana, Faculty of Engineering, 1000 Ljubljana, Slovenia (e-mail: julian.ascencio@3e.eu).

Jonathan Leloux is with LuciSun, 1495 Sart-Dames-Avelines, Belgium, and also with the Universidad Politécnica de Madrid, 28040 Madrid, Spain (e-mail: jonathan.leloux@lucisun.com).

David Moser is with EURAC Research, 39100 Bolzano, Italy (e-mail: David.Moser@eurac.edu).

Angèle Reinders is with the University of Twente, 7500 AE Enschede, The Netherlands, and also with the Eindhoven University of Technology, 5600 MB Eindhoven, The Netherlands (e-mail: a.h.m.e.reinders@utwente.nl).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JPHOTOV.2021.3093049.

Digital Object Identifier 10.1109/JPHOTOV.2021.3093049

power of the PV array. In addition, weather variables from the ERA5 [6] complement this data bank adding temperature, wind speed, relative humidity, and irradiance values.

Besides PEARL PV's data bank, previously DuraMAT [7], IEA PVPS Task13 [8], BDPV [9], Sonnenertrag [10], and PVOutput [11] data servers have been established, as well as NREL's PV Fleet Performance Data Initiative [12]. The Initiative aims to collect and evaluate the performance of 100 000 PV systems across the U.S. [13]. Other international performance studies of system fleets include [14]–[16], while a number of other studies have been conducted in Europe [17]–[28]. PV performance analysis of large datasets resulting from analyses of data in these data banks have shown a global picture of the performance of PV systems in the world with particular focus on small-scale residential PV installations.

The main research questions of this article are as follows: 1) what is the average annual PV system performance to be determined by the performance ratio (PR) of this large sample of PV systems and 2) how is PV system performance related to PV system's metadata, climate data, and other variables? The change of system performance in time can be evaluated by the performance loss rate (PLR) which—in this article—is determined by three different methods: a) Seasonal and Trend decomposition using LOESS (STL) [29], b) the Year-on-Year approach (YoY) [14], [30], and c) statistical clear-sky fitting [31]. The next question is how the PLR is related to metadata of PV systems as well as climate data and other variables? Given the size of the sample, uncertainties, confidence, and accuracy of this statistical analysis will be discussed in the context of the total population of PV systems in Europe.

The rest of this article is organized as follows. In Section II, we will first introduce the datasets that have been analyzed as well as the metadata of the associated PV systems. In Sections III–V, respective data processing and the analyses of PV systems' performance ratio and performance loss rate will be presented. Finally, Section VI concludes this article.

II. Cost Action Pearl PV Database

The dataset of the Pearl PV database is stored as time series in 10 min resolution ranging from one to four years from 2010 to 2016. The following metadata are included for a part of the systems: installed capacity, latitude, longitude, azimuth, and tilt. Especially the latter four features are required to understand the operating conditions (location and mounting) and to retrieve weather data over the operation period. In this article, ERA5 reanalysis data are used [6]. Most of the systems are small scale residential PV plants with a median installed capacity of 6 kWp. The majority is installed in Europe, visible in Fig. 1, in particular in Belgium, France, U.K., and Italy. As we will see in the next sections, some of the investigated PV system data did not pass certain data requirements and had to be excluded. For others, the performance ratio and/or the performance loss rate has been calculated.

Given their geographic allocation, the PV systems are divided into the Köppen-Geiger-Photovoltaic (KGPV) climate zones [32]. The majority of systems is installed in a temperature climate with medium irradiation.

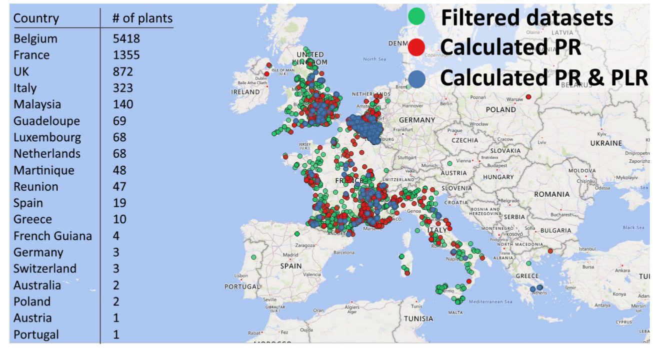

Fig. 1. Distribution of PV systems of Pearl PV database in Europe together with number of systems; green—filtered datasets, red—Calculated PR, blue—Calculated PLR & PR.

TABLE I
Data Quality Grading Criteria. Outliers Include Values Out of Range of $0 < P < 1.05P_{nom}$; Missing Data are Based on Hourly Values and Include Outliers; Longest Gap is Daily. The Dataset Length Needs to Be > Two Years for a P Grade, Otherwise it is Graded F [33]

Letter Grade	Outlier [%]	Missing percentage [%]	Longest Gap [days]	Pass-Fail criterion
A	Below 10	Below 10	Below 15	*P*:
B	10 to 20	10 to 25	15 to 30	x $\geq$ 24 months
C	20 to 30	25 to 40	30 to 90	*F*:
D	Above 30	Above 40	Above 90	x < 24 months

TABLE II
Data Quality Grading by Number of Systems (and Percentage of Total) of Pearl PV Database Before and After Application of Pass-Fail Criterion

All systems				
Letter Grade	Outlier	Missing percentage	Longest Gap	Pass-Fail criterion
A	8,367 (100%)	5,773 (69%)	7,655 (91.5%)	*P*: 4,323
B	0 (0%)	2,216 (26.5%)	280 (3.3%)	(51.7%)
C	0 (0%)	164 (2%)	291 (3.5%)	*F*: 4,044
D	0 (0%)	214 (2.5%)	141 (1.7%)	(48.3%)
Systems which passed Pass-Fail criterion (4,323 systems)				
A	4,323 (51.7%)	2,972 (35.5%)	3929 (47%)	
B	0 (0%)	1238 (14.8%)	155 (1.9%)	
C	0 (0%)	65 (0.8%)	160 (1.9%)	
D	0 (0%)	48 (0.6%)	78 (0.9%)	

III. Experimental Setup—Data Description

The investigated datasets were subject to a preliminary data quality grading, developed in the IEA PVPS Task 13 [33], which is evaluating the dataset based on the relative amount of outliers as well as missing data as shown in Table I. Additionally, a pass-fail criteria is in place for a minimum of 24 months of available data, which is the theoretical minimum for performance loss calculations, as the performance loss describes a year-to-year change.

The data quality grading for the Pearl PV database can be seen in Table II. Missing data are nonexisting entries or measured power equal to 0 during daylight hours. To accurately account for missing data, night entries have been removed using modeled

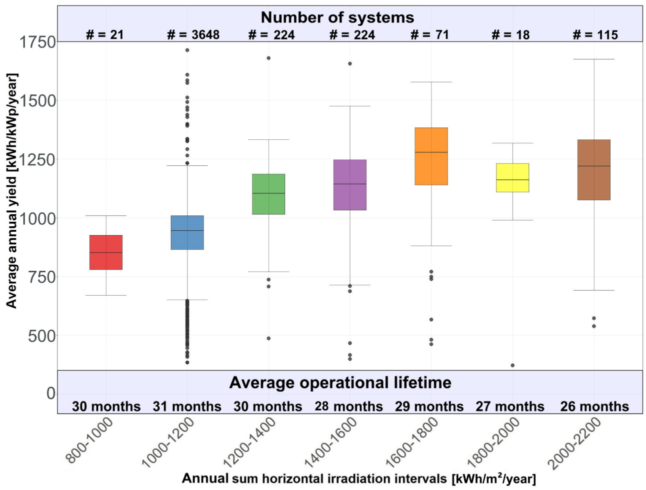

Fig. 2. Boxplots of average annual yield [kWh/kWp/year] versus intervals of annual sum of global horizontal irradiation [kWh/m^2/year]; number of PV plants per interval are provided on top; average lifetime within the interval on the bottom.

clear sky filter to detect and remove instances without irradiance. Data are provided in 10 min resolution without any major power measurement outlier. Instead, roughly 30% of the datasets have more than 10% of missing data in hourly resolution, which is to be expected considering the systems are mainly residential and therefore have simple monitoring systems. For 51.7% of the original 8367 PV systems, at least two years of data are available.

IV. PV System Performance Assessment

In this section, the performance of the PV systems is evaluated. Thereby, the key performance indicators (KPI's) of average annual final yield and performance ratio (PR) are considered. The final yield Y_f describes the produced energy of the PV system at hand divided by the installed capacity P_nom. The PR represents the ratio between the final yield and the reference yield Y_ref of a PV system [34]

$$PR = \frac{Y_\mathrm{f}}{Y_\mathrm{ref}} = \frac{E_\mathrm{AC}/P_\mathrm{nom}}{H_\mathrm{POA}/G_\mathrm{STC}} \tag{1}$$

E_AC is the ac energy of the system (energy equals accumulated power [W to Wh]), H_POA the plane-of-array irradiation (irradiation is accumulated irradiance [W/m^2 to Wh/m^2]), and G_STC is the irradiance under standard test conditions of 1000 W/m^2. The PR is a unitless parameter, which describes the relationship between incoming irradiation and produced energy by a PV system. In the course of this study, irradiation data for the calculation of the PR are retrieved from ERA5 in hourly resolution [6] and modeled to the plane-of-array using the Erbs [35] and Hay-Davies [36] model for decomposition and transposition, respectively.

Fig. 2 presents the annual yield over annual horizontal irradiation intervals. A clear pattern can be observed, where system installed under higher irradiation conditions have higher energy yields. The average annual energy yield across all systems is 954.86 kWh/kWp. This value is driven by the high number of systems within the annual global horizontal irradiation bin of 1000–1200 kWh/m^2/year. Furthermore, it is visible that the energy yield does not further increase for the investigated systems under very high irradiation (>1800 kWh/m^2/year). We believe that this is, at least partially, due to lower achieved efficiencies under higher operating temperatures. The mean annual ambient temperature for systems installed in regions with an irradiation below 1800 kWh/m^2/year is 10.3 °C whereas this value amounts to 25.2 °C in regions with higher irradiation. The dependence of annual horizontal irradiation and ambient temperature is also visible in Fig. 3. Here, the PR in dependence of incoming irradiation and ambient temperature is shown.

Both Fig. 2 as well as Fig. 3 show that most systems are installed in locations with an annual horizontal irradiation of 1000–1200 kWh/m^2/year, where an average yield of 925.01 kWh/kWp/year and an average annual PR of 77.7% is calculated. The number of systems available for yield and PR evaluation varies because required information on tilt and azimuth were not available for all systems under investigation. While the yield of 4206 systems has been calculated, the PR of 3494 systems was determined; this represents 83% of the analyzed data. The mean and median annual PR values per bin are as follows:

- 800-1000 kWh/m^2/year: $\overline{PR} = 80.1\%$ & $\widetilde{PR} = 81.8\%$
- 1000-1200 kWh/m^2/year: $\overline{PR} = 77.7\%$ & $\widetilde{PR} = 79.1\%$
- 1200-1400 kWh/m^2/year: $\overline{PR} = 75.6\%$ & $\widetilde{PR} = 76.4\%$
- 1400-1600 kWh/m^2/year: $\overline{PR} = 67.9\%$ & $\widetilde{PR} = 70.4\%$
- 1600-1800 kWh/m^2/year: $\overline{PR} = 67.0\%$ & $\widetilde{PR} = 71.4\%$
- 1800-2000 kWh/m^2/year: $\overline{PR} = 63.1\%$ & $\widetilde{PR} = 63.1\%$
- Overall: $\overline{PR} = 76.7\%$ & $\widetilde{PR} = 78.4\%$

A steadily decreasing trend both in mean and median PR is evident with an increasing irradiation intensity, due to rising temperatures in locations with higher irradiation. The PR is thereby negatively correlated with the energy yield, which experiences a rise with increasing irradiation intensity.

Most systems analyzed operate within a PR range of 60% to 90%, in line with previous work already mentioned in Section I. The small fraction of PR values higher than 90% is probably not realistic and most likely the result of some inaccuracies on the metadata of some PV systems, in particular their azimuth, tilt, and peak power.

V. Performance Loss Rate Evaluation and Data Requirements

A. Methodology

According to previous research [33], there is currently no single approach to follow to calculate PLR in order to achieve the most reliable results across a number of PV system datasets. The study was based on PV plants with high quality data and available on-site measured irradiance measurements. The lack of measured meteorological data in the COST Action PEARL PV database makes the task of calculating reliable PLR even more nontrivial. That is why we selected and tested three different

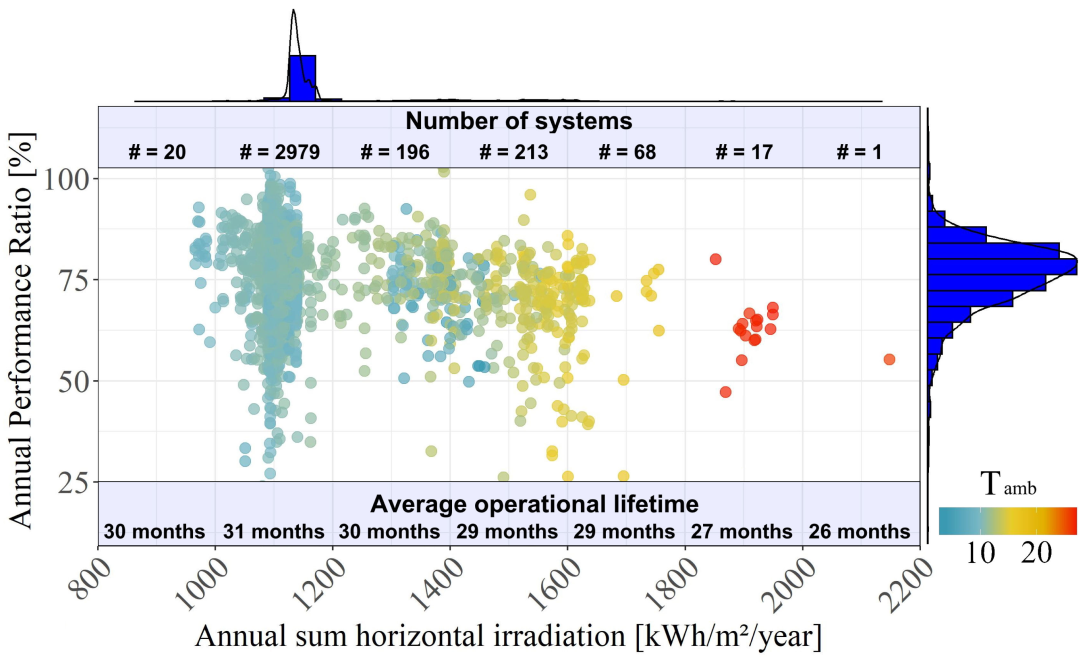

Fig. 3. Annual sum of horizontal irradiation [kWh/m ^{2}/year] versus annual PR [%] categorized by ambient temperature [°C] together with respective histograms; number of PV plants per interval are provided on top; average lifetime on the bottom.

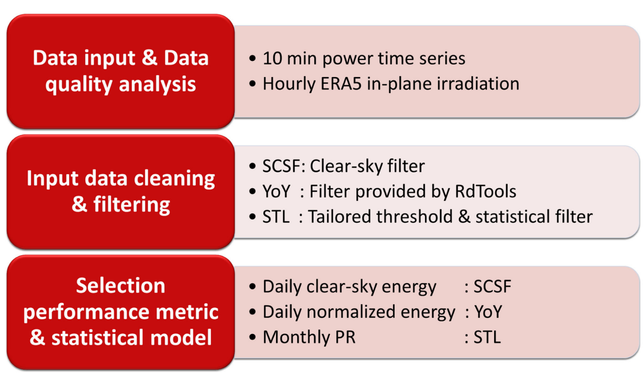

Fig. 4. General performance loss rate calculation steps.

performance loss rate calculation approaches. Fig. 4 summarizes the performed calculation steps.

First, the input data have to be read. From the database, 10-min ac power time series of varying length are available. Additionally, hourly satellite irradiance and temperature data have been retrieved. Using tilt and azimuth of the plants, horizontal irradiance values were transposed to the plane-of-array. After an initial data quality check, discussed in Section III, various filters are applied in dependence of the chosen calculation approach. Table III summarizes the selected filter divided by calculation approach.

Hereby, G_{POA} refers to the plane-of-array irradiance and T_{mod} to the PV module temperature. The PR filter excludes nonrealistic power-irradiance pairs through the nearly linear power-irradiance relationship over a wide irradiance interval.

In the next step, the data are aggregated and the performance metric is selected, which is used for the PLR calculation. As

TABLE III
APPLIED FILTER FOR CALCULATION STEPS

	Filter			
Statistical method	G_{POA} [W/m^2]	T_{mod} [°C]	Power	Performance Ratio
SCSF	Strict clear-sky filter			
YoY	200 - 1200	-50 - 110	P > 0	
STL	100 - 1200		(0.01 - 1.05) $\cdot P_{nom}$	$\pm 2\sigma$ around daily mean PR

shown in Fig. 4, statistical clear-sky fitting (SCSF) uses the clear-sky filtered power/energy time series in daily resolution. The Year-on-Year (YoY) approach is applied on daily normalized energy data, computed with PVWatts [37], and seasonal and trend decomposition using LOESS (STL) on monthly PR time series.

The consecutive steps depend on the chosen statistical method and are discussed below.

Statistical clear-sky fitting: This algorithm was recently developed as an alternative to common PLR calculation methods, which require irradiance time series, and sometimes also temperature and other meteorological data. It only needs a measured power dataset as input, which is used as the performance metric, and fits a clear-sky model to the data. This clear-sky model serves as well as an outlier filter for nonrepresentative data. SCSF consists of two parts. First, it constructs an underlying mathematical model of PV power time series and subsequently fits it to the measurements. The modeled PV data have a perfect clear-sky behavior and are therefore expected to have a certain annual periodicity including a reduction over time. This consistent year-over-year percent change in daily energy becomes the estimate of system PLR [31].

TABLE IV
NUMBER OF PV SYSTEMS AND CORRESPONDING PERFORMANCE LOSS RATES AFTER VARIOUS FILTERING STEPS

Filter	**SCSF**	**YoY**	**STL**
All systems	8,367		
Min. 2 years of data	4,323		
Azimuth & tilt		3,494	3,494
Calculated PLR	4,165	3,342	3,342
Min. 3 years of data	754	613	613
-4%/a < PLR < 1%/a	**661**	**361**	**366**

Year-on-Year approach: The YoY approach was first developed by Sunpower [14], later improved by NREL [30], and is now available in the RdTools package [38] in Python and the PVplr package [39] in R.

YoY is using a loss rate distribution instead of one single value as it averages differences between data-points. Thereby, the differences between a data-point in a calendar year and the data-point at the same position in the subsequent year are collected over the complete dataset. The median of the differences for all data-points available represents the overall system PLR. The PLR of the YoY method is normalized to the first-year's median.

Seasonal and Trend decomposition using LOESS: LOESS stands for locally estimated scatterplot smoothing and STL is based on locally weighted regression [29], [40]. STL decomposes a seasonal time series into three components, namely "trend" (T_t), "seasonal" (S_t), and "remainder" (R_t)) in an additive fashion

$$Y_\mathrm{t} = T_\mathrm{t} + S_\mathrm{t} + R_\mathrm{t}. \tag{2}$$

Afterward, a linear fit of the nonlinear trend is performed. The following equation is used to calculate the final PLR

$$PLR\,[\%/a] = \left(\beta_1 \frac{t}{\beta_0}\right) 100 \tag{3}$$

where β_1 is the gradient and β_0 the y-intercept of the linear trend line of the linearized model. t is a scaling parameter that converts the time scale to a yearly scale (e.g., 365 for daily, 12 for monthly).

B. PLR Results

A first observation was the difference in computation times between the approaches. The application of SCSF as well as YoY has been carried out in Python while STL has been applied in R. While it took SCSF [31] several days to complete the calculations, the application of YoY [38] and STL [41] was finished after roughly 200 min each. For all calculations, a computer with an Intel(R) Core (TM) i5-7200 U CPU using 8 GB of RAM, running Windows 10, has been used. SCSF is a more sophisticated algorithm, which requires by far more computational resources.

Table IV shows the reduction of available PV system datasets for PLR calculations after various filtering and exclusion steps. Initially, 4165 and 3342 PLR have been calculated depending on the selected approach.

TABLE V
CORRELATION COEFFICIENT [42] AND P-VALUE (WELCH TWO SAMPLE T-TEST [43])

	SCSF	**YoY**	**STL**	
SCSF		0.10	0.51	p-value
YoY	0.25		0.38	
STL	0.30	0.42		
	Correlation			

Jordan *et al.* [16] stated that PLR calculated even on five years of data with limited weather data accuracy are expected to have high uncertainties, which can also be assumed for noncorrected ERA5 reanalysis data as we have here. That is why another time-dependent filter was included, looking only at time series with a minimum length of three years. Another threshold filter of -4%/year $<$ PLR $<$ 1%/year was applied to remove strong outlier. The additional filter yield more reasonable results. From originally more than 8000 systems, between 361 and 661 remain part of this analysis depending on the calculation method.

Fig. 5 shows the method-dependent distribution of the PLR after the last filtering step. In the top-left corner, the mean and the median values are given. Comparing mean and median, it is visible that all three distributions are left-skewed. SCSF peaks at the bin of -0.5 to 0%/year, whereas YoY and STL have very wide distributions between -1%/year and 1%/year.

Table V shows the correlation between different PLR calculation approaches. On the left side, the correlation coefficient [42] is given, and the right side provides the p-value between the PLR values [43]. The individually calculated PLR using the three different approaches are weakly correlated with the highest correlation of 0.42 found between values calculated with STL and YoY. It is likely that the stronger correlation stems from the shared irradiance dataset. Furthermore, a possible shortcoming of the SCSF method is that it is based on the assumption that the annual solar resource does not change over time. Year-to-year changes in irradiation, which might be especially pronounced in shorter time-series as we have them here, might affect the results. By including irradiation into the PLR determination, these year-to-year changes are accounted for. The results of the t-test clearly underline an existing correlation between the PLR values calculated with the different approaches. Thereby, the p-value between SCSF and STL is with 0.51 the highest.

A definite statement on which approach yields the most trustful results cannot be made. To achieve this, higher quality irradiance data and longer time series would be needed. The overall results of this study agree well with other PLR distributions given in the literature. Jordan *et al.* [44] collected PLR from multiple sources and calculated a mean value of -0.8 to -0.9%/year and a median value of -0.5 to -0.6%/year. However, most of the results are based on module measurements, thus not on system data, and hence they do not capture system losses. Kiefer *et al.* [25] reported a mean PLR of -0.7%/year for 44 large scale systems.

Nevertheless, mainly small scale PV systems are considered in our study and it is expected that they are subject to higher systems losses, because better operation practices are usually applied to large scale systems. Moreover, residential PV systems do

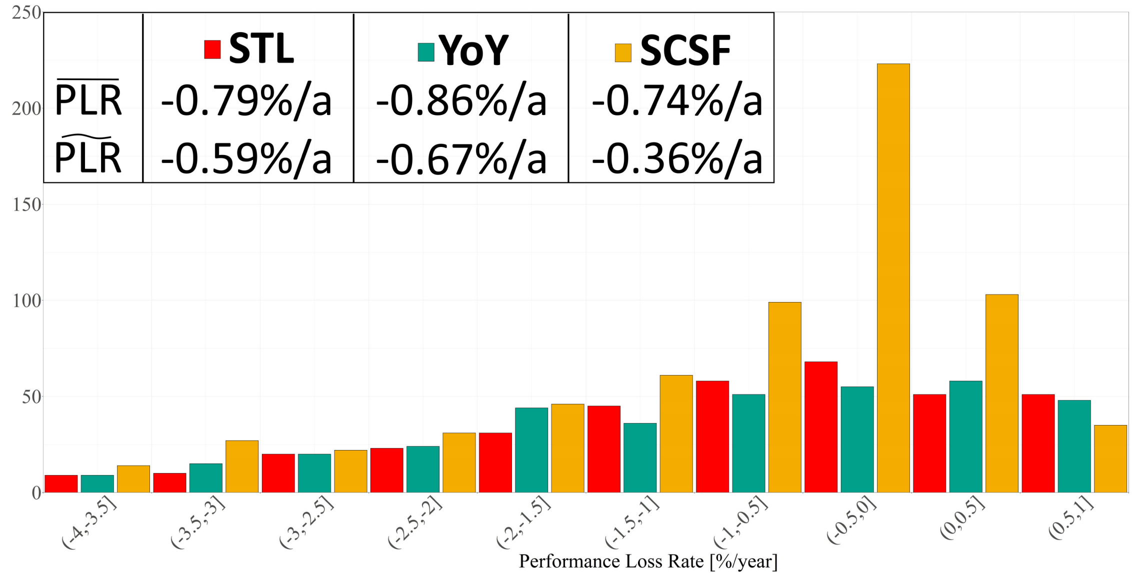

Fig. 5. Performance loss rate distribution of Pearl-PV performance database; red: calculated with STL; turqouise: calculated with YoY; yellow: calculated with SCSF.

not normally benefit from regular O&M surveillance by trained professionals as it occurs in large scale PV systems [16], [45].

VI. Conclusion

In this article, a PV system fleet, collected within the COST Action Pearl PV database, is presented and a first performance evaluation has been carried out. The fleet consists of residential systems primarily installed in central Europe. After the initial presentation of the system fleet, a data quality assessment has been carried out, the performance evaluated and the performance loss rates been estimated.

The system data quality has been assessed using a new data quality grading scheme developed in the IEA PVPS Task 13 showing that the quality of the data is overall quite high. Half of the systems have at least two years of operational data available. The average annual yield is 954.9 kWh/kWp and the average performance ratio 76.7% considering all systems with at least one year of data. Roughly half of the systems are included for the yield estimation while 3500 systems are considered for the PR. The difference in the number of systems is due to the fact that the tilt and azimuth have to be known for the PR calculation, which was not the case for all systems.

The last part of this work focuses on the calculation of performance loss rates. It is a first study showing preliminary results and, more importantly, requirements from the data-side for reliable calculation procedures. Three different approaches have been tested, namely statistical clear-sky fitting (SCSF), seasonal and trend decomposition using LOESS (STL) and the year-on-year approach (YoY). SCSF is a somewhat exotic approach as it does not require any weather data for the PLR calculation. It is based on clear-sky modeling of power time series connected to a year-over-year evaluation of the clear-sky modeled power. Thereby, it does not account for changes in the solar resource, for example, through global brightening effects. STL and YoY instead require as input, apart from the power time series, plane-of-array irradiance. In this study, ERA5 irradiance datasets have been used. From originally roughly 8400 PV systems time series, more than 90% had to be excluded due to being categorized as outlier or having an insufficient length. It was shown that a minimum of three years of power data are required to get somewhat sensible results, keeping in mind that the longer the time series, the more trustworthy the results tend to be. The results are in a range of mean PLR between -0.74 %/year and -0.75 %/year and median between -0.36%/year and -0.67%/year, depending on the chosen approach.

This study serves as a first evaluation of the COST Action Pearl PV database. Performance studies of PV system fleets are scarce but of utmost importance for understanding the correlation between climate stressors and PV system performance and degradation of installed PV systems. These advances are necessary to better design and operate PV systems depending on environmental and installation conditions. It is foreseen to update and enlarge the existing database in order to draw stronger conclusions and gain a better insight into the climate related degradation of this system fleet.

Acknowledgment

The authors are thankful to Adrien Desportes and Fabien Grenier from Rtone for their willingness to share the PV performance data from Rbee Solar. This publication is partially based upon work from European Cooperation in Science and Technology (COST) Action CA16235 PEARL PV,[1] COST is a funding agency for research and innovation networks. COST Actions help connect research initiatives across Europe and enable scientists to grow their ideas by sharing them with their peers. This boosts their research, career, and innovation.

[1] [Online]. Available: https://www. pearlpvcost.eu

References

[1] A. Reinders *et al.*, "Development of a big data bank for PV monitoring data, analysis and simulation of COST action 'PEARL PV,"' in *Proc. IEEE 46th Photovolt. Specialists Conf.*, Jul. 2019, pp. 1388–1392.

[2] A. Reinders *et al.*, "Introducing PEARL-PV: Performance and reliability of photovoltaic systems: Evaluations of large-scale monitoring data," in *Proc. IEEE 7th World Conf. Photovolt. Energy Convers. (Joint Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC)*, Jun. 2018, pp. 762–766.

[3] A. Reinders, D. Moser, and J. Leloux, "PEARL-PV country reportsPEARL-PV," Univ. Twente, Enschede, The Netherlands, Rep. 978-90-365-5108-3, Dec. 2020.

[4] NIST, "NIST/SMATECH e-handbook of statistical methods," 2012, Accessed: Mar. 22, 2021. [Online]. Available: http://www.itl.nist.gov/div898/handbook/

[5] R. Conroy, "The RCSI sample size handbook," 2016, Accessed: Mar. 22, 2021. [Online]. Available: https://www.rcsi.ie/files/research/docs

[6] Copernicus Climate Change Service (C3S) ERA5, "Fifth generation of ECMWF atmospheric reanalyses of the global climate. copernicus climate change service climate data store (CDS)," 2017. [Online]. Available: https://cds.climate.copernicus.eu/cdsapp/home

[7] DuraMAT, "DuraMAT data standards and guidelines," 2013. Accessed: May 30, 2019. [Online]. Available:https://datahub.duramat.org/dataset/

[8] R. Nordmann, L. Clavadetscher, W. Van Sark, and M. Green, "Analysis of long-term performance of PV systems," Int. Energy Agency, Paris, France, Tech. Rep. IEA PVPS T13-05:2014, 2014.

[9] "BDPV—Base de données photovoltaïques," Accessed: Mar. 13, 2019. [Online]. Available: https://www.bdpv.fr/en/index.php/

[10] "Sonnenertrag," Accessed: Mar. 14, 2019. [Online]. Available: http://www.solar-yield.eu/

[11] "PVOutput," Accessed: May 30, 2019. [Online]. Available: https://pvoutput.org/

[12] C. Deline *et al.*, "PV Fleet performance data initiative: Mar. 2020 Methodology Report," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-5K00-76687, Mar. 2020. [Online]. Available: https://www.nrel.gov/docs/fy20osti/76687.pdf

[13] C. Deline *et al.* "PV fleet performance data initiative—2021 update," in *Proc. PV Rel. Workshop*, 2021.

[14] E. Hasselbrink *et al.*, "Validation of the PVLife model using 3 million module-years of live site data," in *Proc. IEEE 39th Photovolt. Specialists Conf.*, Jun. 2013, pp. 7–12.

[15] A. M. Nobre *et al.*, "Performance evaluation of a fleet of photovoltaic systems across India and Southeast Asia," in *Proc. IEEE 46th Photovolt. Specialists Conf.*, 2019, pp. 1372–1376.

[16] D. C. Jordan, B. Marion, C. Deline, T. Barnes, and M. Bolinger, "PV field reliability status-analysis of 100000 solar systems," *Prog. Photovolt., Res. Appl.*, vol. 28, no. 8, pp. 739–754, 2020. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.3262

[17] U. Jahn and W. Nasse, "Operational performance of grid-connected PV systems on buildings in Germany," *Prog. Photovolt., Res. Appl.*, vol. 180, pp. 441–448, 2004. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.550

[18] N. H. Reich, A. Goebel, D. Dirnberger, and K. Kiefer, "System performance analysis and estimation of degradation rates based on 500 years of monitoring data," in *Proc. 38th IEEE Photovolt. Specialists Conf.*, 2012, pp. 001 55 1–001 555.

[19] G. Colantuono, A. Everard, L. Hall, and A. Buckley, "Monitoring nationwide ensembles of PV generators: Limitations and uncertainties. The case of the UK," *Sol. Energy*, vol. 108, pp. 252–263, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X14003296

[20] J. Leloux, L. Narvarte, and D. Trebosc, "Review of the performance of residential PV systems in Belgium," *Renewable Sustain. Energy Rev.*, vol. 16, no. 1, pp. 178–184, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364032111003923

[21] J. Leloux *et al.*, "Monitoring 30000 PV systems in Europe: Performance, faults, and state of the art," in *Proc. 31th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2015, pp. 1574–1582. [Online]. Available: https://www.eupvsec-proceedings.com/proceedings?paper=34857

[22] J. Leloux, L. Narvarte, and D. Trebosc, "Review of the performance of residential PV systems in France," *Renewable Sustain. Energy Rev.*, vol. 16, no. 2, pp. 1369–1376, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S136403211100517X

[23] J. Taylor *et al.*, "Performance of distributed PV in the UK: A statistical analysis of over 7000 systems," in *Proc. 31th Eur. Photovolt. Sol. Energy Conf. Exhib.*, 2015, pp. 2271–2276. [Online]. Available: https://www.eupvsec-proceedings.com/proceedings?paper=33981

[24] S. Lindig, D. Moser, A. J. Curran, and R. H. French, "Performance loss rates of PV systems of task 13 database," in *Proc. IEEE 46th Photovolt. Specialists Conf.*, 2019, pp. 1363–1367.

[25] K. Kiefer *et al.*, "Degradation in PV power plants: Theory and practice," in *Proc. 36th Eur. Photovolt. Sol. Energy Conf. Exhib. Proc.*, Sep. 2019, pp. 1331–1335.

[26] H. te Heesen, V. Herbort, and M. Rumpler, "Performance of roof-top PV systems in Germany from 2012 to 2018," *Sol. Energy*, vol. 194, pp. 128–135, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X19310023

[27] S. Seme, K. Sredensek, B. Stumberger, and M. Hadziselimovic, "Analysis of the performance of photovoltaic systems in Slovenia," *Sol. Energy*, vol. 180, pp. 550–558, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X19300787

[28] J. Schardt and H. T. Heesen, "Performance of roof-top PV systems in selected European countries from 2012 to 2019," *Sol. Energy*, vol. 217, pp. 235–244, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0038092X21001006

[29] R. B. Cleveland, W. S. Cleveland, J. E. McRae, and I. Terpenning, "STL: A seasonal-trend decomposition procedure based on loess," *J. Official Statist.*, vol. 6, no. 1, pp. 3–73, Mar. 1990.

[30] D. C. Jordan, C. Deline, S. R. Kurtz, G. M. Kimball, and M. Anderson, "Robust PV degradation methodology and application," *IEEE J. Photovolt.*, vol. 8, no. 2, pp. 525–531, Dec. 2018.

[31] B. Meyers, M. Deceglie, C. Deline, and D. C. Jordan, "Signal processing on PV time-series data: Robust degradation analysis without physical models," *IEEE J. Photovolt.*, vol. 10, no. 2, pp. 546–553, Mar. 2020.

[32] J. Ascencio-Vasquez, K. Brecl, and M. Topic, "Methodology of Köppen–Geiger-photovoltaic climate classification and implications to worldwide mapping of PV system performance," *Sol. Energy*, vol. 191, pp. 672–685, Nov. 2019.

[33] S. Lindig *et al.*, "International collaboration framework for the calculation of performance loss rates: Data quality, benchmarks, and trends (towards a uniform methodology)," *Prog. Photovolt., Res. Appl.*, vol. 29, no. 6, pp. 573–602, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.3397

[34] *Photovoltaic System Performance, Part 1: Monitoring*, IEC Standard 61724-1:2017, 2017.

[35] D. Erbs, S. Klein, and J. Duffie, "Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation," *Sol. Energy*, vol. 28, no. 4, pp. 293–302, 1982. [Online]. Available: http://www.sciencedirect.com/science/article/pii/0038092X82903024

[36] P. Loutzenhiser *et al.*, "Empirical validation of models to compute solar irradiance on inclined surfaces for building energy simulation," *Sol. Energy*, vol. 81, no. 2, pp. 254–267, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0038092X06000879

[37] "PVWatts version 5 manual," Nat. Renewable Energy Lab., Golden, CO, USA, Tech. Rep. NREL/TP-6A20-62641, 2014. [Online]. Available: https://pvwatts.nrel.gov/downloads/pvwattsv5.pdf

[38] RdTools, "Version 2.0.5." 2020. [Online]. Available: https://github.com/NREL/rdtools

[39] A. Curran, T. Burleyson, S. Lindig, D. Moser, and R. French, "PVplr: Performance loss rate analysis pipeline," R package version 0.1.0. 2020. [Online]. Available: https://CRAN.R-project.org/package=PVplr

[40] R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. Melbourne, VIC, Australia: OTexts, 2013.

[41] S. Seabold and J. Perktold, "Statsmodels: Econometric and statistical modeling with python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 92–96.

[42] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Pearson Correlation Coefficient*. Berlin, Germany: Springer, 2009.

[43] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria: R. Foundation for Statistical Computing, 2020. [Online]. Available: https://www.R-project.org/

[44] D. C. Jordan, S. Kurtz, K. VanSant, and J. Newmiller, "Compendium of photovoltaic degradation rates," *Prog. Photovolt., Res. Appl.*, vol. 24, pp. 978–989, Feb. 2016.

[45] R. Dubey *et al.* "Comprehensive study of performance degradation of field-mounted photovoltaic modules in India," *Energy Sci. Eng.*, vol. 5, no. 1, pp. 51–64, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/ese3.150