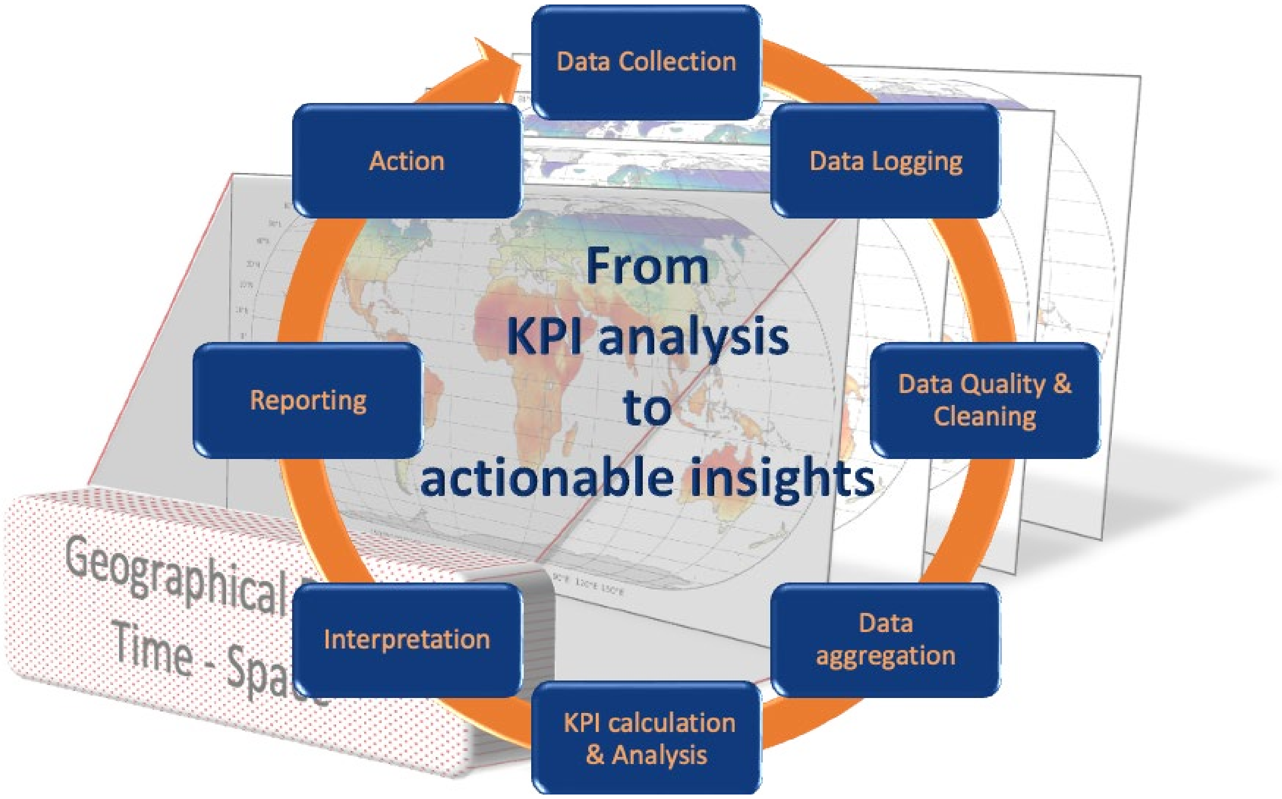
Technology Collaboration Programme
by iea

International Energy Agency
Photovoltaic Power Systems Programme



Task 13 Reliability and Performance of Photovoltaic Systems

PVPS

Best practice guidelines for the use of economic and technical KPIs

2024

Report IEA-PVPS T13-28:2024

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 Programmes (TCP) were created with a belief that the future of energy security and sustainability starts with global collaboration. The programmes are made up of 6.000 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 TCPs 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.” 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 organisation member, which designates distinct ‘Tasks,’ that may be research projects or activity areas.

The IEA PVPS participating countries are Australia, Austria, Belgium, Canada, China, Denmark, Finland, France, Germany, Israel, Italy, Japan, Korea, Malaysia, Mexico, Morocco, the Netherlands, Norway, Portugal, South Africa, Spain, Sweden, Switzerland, Thailand, Türkiye, and the United States of America. The European Commission, Enercity, Solar Energy Research Institute of Singapore and Solar Power Europe 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.

IEA PVPS Task 13 has so far managed to create a framework for the calculations of various parameters that can indicate the quality of PV components, systems, and applications. The framework is available and can be used by the PV industry which 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, the United States of America, and the Solar Energy Research Institute of Singapore.

COVER PICTURE

The data processing cycle backed by an illustration of mapped PV specific climate stressors. Source: Univers/Julián Ascencio Vásquez, Sascha Lindig

ISBN 978-3-907281-69-7 Task 13 Report: Best practice guidelines for the use of economic and technical KPIs

INTERNATIONAL ENERGY AGENCY

PHOTOVOLTAIC POWER SYSTEMS PROGRAMME

IEA PVPS Task 13

Reliability and Performance of Photovoltaic Systems

Best Practice Guidelines for the Use of Economic and Technical KPIs

Report IEA-PVPS T13-28:2024

December 2024

ISBN 978-3-907281-69-7

AUTHORS

Main Authors

Sascha Lindig, Univers SAS, Courbevoie, France
Julien Deckx, 3E, Brussels, Belgium
Magnus Herz, TÜV Rheinland, Cologne, Germany
Julián Ascencio Vásquez, Univers SAS, Courbevoie, France
Marios Theristis, Sandia National Laboratories, Albuquerque, USA
Bert Herteleer, KU Leuven, Ghent, Belgium
Kevin Anderson, Sandia National Laboratories, Albuquerque, USA

Contributing Authors

David Moser, EURAC Research, Bolzano, Italy
Atse Louwen, EURAC Research, Bolzano, Italy
Erik Stensrud Marstein, IFE, Norway

Editors

Sascha Lindig, Univers SAS, Courbevoie, France
Ulrike Jahn, Fraunhofer CSP, Halle, Germany

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

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

Funding agency statements:

This report is supported by the German Federal Ministry for Economic Affairs and Climate Action (BMWK) under contract no. 03EE1120B.

This work was 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 38267. Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This report describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the report do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

The work was supported by funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement N952957, project "TRUST-PV".

LIST OF ABBREVIATIONS

ACOE	Actual cost of electricity
AM	Asset manager
AO	Asset owner
BIF	Bifaciality factor
CAPEX	Capital expenditure
CFD	Cumulative distribution function
CI	Confidence intervals
CUF	Capacity utilization factor
DHI	Direct diffuse irradiance
DNI	Direct normal irradiance
EF	Environmental Footprint
EPBT	Energy payback time
EPC	Engineering, procurement, and construction
EPD	Environmental product declaration
EPI	Energy performance index
EROI	Energy Return on (Energy) Investment
IEA	International Energy Agency
IR	Infrared
IRR	Internal return of investment
I-V	Current-Voltage
GHI	Global horizontal irradiance
GWP	Global warming potential
KPI	Key performance indicator
LCA	Lifecycle assessment
LCOE	Levelized cost of electricity
NPV	Net present value
NREPBT	Non-renewable energy payback time
O&M	Operation & maintenance
OEF	Organization environmental footprint
OPEX	Operational expenditure
PEF	Product environmental footprint
PLF	Plant load factor
PLR	Performance loss rate
POA	Plane-of-array

PPA	Power purchase agreement
PR	Performance ratio
Pt	Platinum
PV	Photovoltaic
Rd	Degradation rate
RH	Relative humidity
ROI	Return on investment
SAPM	Sandia PV array performance model
SCADA	Supervisory control and data acquisition
SL	Soiling level
SR	Soiling ratio
STC	Standard test conditions
TMY	Typical meteorological year
TRY	Typical reference year
UAV	Unmanned aerial vehicle
UPS	Uninterruptible power supply
γ	power temperature coefficient

EXECUTIVE SUMMARY

Key Performance Indicators (KPIs) are an important set of metrics used to assess various aspects of photovoltaic (PV) systems, including their long-term performance, economic viability and carbon footprint. Technical KPIs support data-driven and informed decision-making when optimizing PV systems and provide a comprehensive overview of how PV systems operate across different conditions and climates. Different KPIs are commonly employed throughout the entire value chain of PV projects and can be categorized into technical, economic and sustainability aspects.

In this work, a set of best practices for handling PV system data to reliably calculate relevant KPIs is discussed. While most technical KPIs are generally well-known among asset owners, EPCs, O&M providers and consultants, not all stakeholders in the financing-to-operation chain are equally aware of the nuances and consequences of certain decisions, which are based on *how* technical KPIs are operationalized, i.e. translated from contracts to *how* and *where* raw data are stored, *which* data cleaning and imputation techniques are used, to *how* the technical KPIs are calculated and used for subsequent decision-making. In many cases, the decisions made in the development-to-construction phase, will affect the asset for a significant part of its lifetime. For example, the resolution at which data is measured, which data are stored, or whether data back-ups are on-site or in the cloud, can all affect how KPIs are calculated, affect future modifications to contractual clauses, or the need for SCADA upgrades. Hence, this work aims to provide all stakeholders deeper insights and a shared understanding of the most important technical KPIs.

The work is divided into three parts, each addressing different aspects of KPIs, data management, and their mapping potential.

Table 1: Usage overview of technical KPIs.

KPI	Abbreviation	Private equity / Bank	Project Developer	Asset Owner / Asset Manager	EPC	O&M	Service provider / consultant
Pxx energy yield	**P50 Yield**	T/C	T/C	T/C	T		T/C
Performance ratio	**PR**			T/C	T/C	T/C	T/C
Availability				T/C	T/C	T/C	T/C
Soiling ratio	**SR**	T	T	T	T	T	T
Degradation rate	**Rd**	T	T	T	T	T	T/C
Performance loss rate	**PLR**	T	T	T	T	T	T/C
Energy performance index	**EPI**			T/C	T/C	T	
Capacity test	**CapTest**			T/C	T/C	T	
Capacity utilization factor	**CUF / PLF**				T	T	T
Maintenance response time	**MRT**			C	C	C	C

T – technical, C – contractual binding

A comprehensive overview of key performance indicators (KPIs) that are important across technical, economic, and sustainability domains, highlighting their common definitions and variations, are presented. In addition, this work delves into the typical advantages and challenges associated with each KPI, and which variations of each KPI exist. The focus of this report is centered on technical KPIs. It has been demonstrated that the application of all investigated KPIs poses challenges, either in terms of their formulation, interpretation, or due to inherent limitations. This work is based on an extensive literature review and feedback from stakeholder questionnaires across various markets and regions. The objective was to understand which KPIs are widely used within the industry, which have contractual binding, and which are primarily applied in a technical framework. This information is summarized in Table 1.

KPIs that are contractually binding carry direct financial implications, while those used in a technical context serve to support the performance assessment of PV plants, and the associated decision-making by stakeholders. The survey showed additionally that while there are certain KPI usage trends per region, a globalized world and market means that there are no strict differences to be seen. Despite the nominal standardization of contractual KPIs such as the performance ratio and temperature-corrected performance ratio, there are still considerable variations in the data quality routines employed, and consequently, in the calculation of the resulting KPIs.

The report focuses furthermore on the challenges and best practices in managing PV system and weather data, covering the entire data processing cycle from input data collection to KPI computation. The most important signals, such as power, current, and voltage values from the PV system, as well as climatic variables from weather stations, are discussed. Commonly recorded variables include irradiance, temperature, wind speed, and wind direction. Additionally, the common structure of a PV system, along with its data and command streams, is presented. The quality of the input data directly influences the certainty of the calculated PV system KPIs. Therefore, the data quality and data cleaning steps within the data processing cycle are of utmost importance. Key data quality criteria to consider include accuracy, completeness, consistency, timeliness, and reliability. The report also presents the latest findings on imputing, or filling, data gaps in PV system power, irradiance, and temperature time series. However, even the best imputation strategies will inevitably increase uncertainty. In this regard, high-quality data should be viewed as an investment that enables better evidence-based decision-making by ensuring reliable KPI calculations, rather than as a cost that diminishes system profitability.

Finally, a holistic view on the mapping potential of KPIs is presented, emphasizing their applications in various contexts. These include performance ratio mapping of individual PV components to assess system health, as well as global geographical mapping of climate stressors and their impact on PV system performance statistics. The discussion extends to exploitable data resources, including raw time series, aggregated KPIs, geospatial weather data from satellite, geospatial post-processed PV data, aerial images from drones, and static data from current-voltage tracers. Additionally, five distinct case studies are presented, illustrating how data can be utilized to analyze specific aspects of PV system health. Each case study is broken down into its fundamental concept, required input data, the calculations performed, the data being analyzed, and how the results should be interpreted. This study sets the tone for future work, especially as the increasing availability of PV system data offers greater opportunities for comprehensive mapping and analysis.

By providing a thorough and practical framework, this work aims to enhance the understanding and application of KPIs in the PV industry. It adheres to the three-part IEC 61724 standard, ensuring consistency and compatibility with existing industry standards, and contributes to the improvement and evolution of the current standard by identifying potential areas of enhancement and providing recommendations for more efficient and standardized KPI usage.

1 INTRODUCTION

Key Performance Indicators (KPIs) are essential tools for assessing the performance of photovoltaic (PV) systems. They provide a framework for evaluating how PV systems operate across different conditions and climates. KPIs are commonly used to assess the value and viability of PV systems from both technical and economic perspectives. The latest Solar PV Analysis by the IEA showed that the Net Zero Scenario requires 7,400 TWh of solar PV generation provided by a PV capacity of about 5 TWp by 2030 [1]. Such goals demand an unprecedented increase in PV deployment and efficiency. Effective usage of KPIs is crucial for optimizing PV system deployment while ensuring safe and efficient operation.

This report focuses mainly on the operational stage of a PV system lifecycle. The reliable and regular calculation of KPIs is thereby the first step to carry out insightful and smart operation and maintenance (O&M) activities within PV plants. Through analyzing reliably calculated KPIs, underperformance of PV systems can be detected, downtime reduced, and the lifetime of systems and components prolonged.

Not all stakeholders in the financing-to-operation chain are equally aware of the nuances and consequences of certain decisions, which are based on *how* technical KPIs are operationalized, i.e. translated from contracts to *how* and *where* raw data is stored, *which* data cleaning and imputation techniques are used, to *how* the technical KPIs are calculated and used for subsequent decision-making. In many cases, the decisions made in the development-to-construction phase, will affect the asset for a significant part of its lifetime. For example, the resolution at which data is measured, which data is stored, or whether data back-ups are on-site or in the cloud, can all affect how KPIs are calculated, affect future modifications to contractual clauses, or the need for Supervisory Control and Data Acquisition (SCADA) upgrades. Hence, this work aims to provide all stakeholders deeper insights and a shared understanding of the most important technical KPIs.

Figure 1: Data processing cycle.

Major challenges are the handling and processing of the available raw data and the selection of reliable tools to do so. Figure 1 shows the general data processing steps going from raw data over the calculation of KPIs to actionable insights. A first important, although not exhaustive, guide of how to process available PV performance data is the three-part standard IEC 61724 [2, 3, 4].

This report addresses the most critical questions surrounding the use of KPIs in the PV sector: What, Why, How, Where, and Who, particularly in relation to PV plant monitoring [5], plant performance comparisons, and indicators dependent on external factors. It is thereby a direct extension of the IEC 61724 by critically discussing relevant KPIs and providing best practices in data acquisition, handling, and usage. The effective formulation and usage of domain related KPIs necessitate high-resolution data and sophisticated mapping techniques, which will additionally be explored in this and future work.

The report lays the groundwork for advanced KPI mapping in the PV sector and is structured as follows:

Chapter 2 describes technical, economic, and sustainability KPIs, their usage, derivations, and the contexts in which they are valuable. This part has been developed carefully by reviewing current literature and incorporating expert knowledge through extensive interaction with PV specialists. This in turn was done through a number of stakeholder questionnaires and the execution of an interactive workshop.

Chapter 3 covers the difficulties and best practices in data management for KPI determination including data types and structure, used hardware and software, data quality, data preparation, and data aggregation. Thereby, data quality issues and tools are described, the impact of data quality attributes on KPI determination discussed, and a best practice data quality routine established.

Chapter 4 presents an overview of the different types of input data for KPI calculations and gives several field examples of the entire process, from data to actions. Thereby, KPIs are connected to their calculation pathways and exploitable results.

Chapter 5 closes with a summary and outlook. In future work, a great number of PV system KPIs will be computed and spatially mapped using extrapolation techniques to study KPI computation sensitivity and to visualize global performance trends.

2 OVERVIEW OF THE MAIN KEY PERFORMANCE INDICATORS IN THE PV SECTOR

In this chapter, the most important KPIs are summarized, focusing on technical, economic and sustainability KPIs. It describes how the KPIs are calculated, their intended purpose, and which stakeholders typically use them. Additionally, benefits and challenges associated with each KPI are discussed.

While health & safety KPIs are not within the scope of this report, it is worth noting that health and safety measures can, under certain circumstances, be linked to performance.

2.1 Technical Key Performance Indicators

2.1.1 Pxx Energy Yield

Description	The most basic technical KPI is the expected energy production of the plant. The Pxx Energy Yield describes the probability of exceedance that the PV system will produce in a given year. The P50 Energy Yield is the median scenario, where 50% of the years, the energy yield will exceed this value, and 50% of the years the yield will be below. The P50 is typically the base case used by investors and proponents, which uses a Typical Meteorological Year (TMY). TMY datasets represent an average year based on historical weather data. By contrast, the P90 and P99 energy yield values are used to stress-test financial models, as these represent 1-in-10 and 1-in-100 probability scenarios of lower energy yields, where a PV project must still remain profitable or capable of servicing its debt. The inter-annual irradiation variability is the main weather uncertainty contributor to the spread of P90 versus P50, whereas other sources of uncertainty (modelling, system operation, system degradation, effects of climate change etc.) are also important. Non-P50 values, such as P90 or P99, are determined by interpolating a cumulative distribution function (CDF) to identify to many standard deviations it corresponds to. This CDF can be derived from Monte Carlo simulations using TMY datasets, long-term historical data from a nearby meteorological station, or, in some cases, Typical Reference Year (TRY) datasets, which are regionally standardized representations of typical weather conditions.
Variations	Varying probabilities of exceedance are used, the most common being P50, P75, P90, P95 and P99.
Application	The Pxx energy yield is an important KPI for financing and offtake contracts. It is also a comparison reference for the expected energy yield (given measured weather data) and the measured energy yield during the first year of operation. Pxx energy yield should not be used for naïve comparisons of different PV systems as it depends on PV technology and available solar resource at each site.
Advantages	Commonly used and easily understandable KPI.

Challenges	Calculating the Pxx energy yield of a site requires an in-depth understanding of uncertainties related to weather data and PV plant performance as well as related to interannual weather variations. Climate change and its impact on weather predictability increases the uncertainty [6].

2.1.2 **Performance Ratio**

Descrip-tion	The performance ratio (PR) is a measure of the efficiency and performance of the entire PV plant and can be considered a measure of energy production (E_{out}), normalized by the nominal size of the array (P_{STC}) and the available plane-of array (POA) irradiation (H_{POA}). E is aggregated power (P) and H is aggregated irradiance (G). Here, the specific (final) yield, Y_f, of a system is set into relation with the reference yield Y_r. The yields are ratios of measured values of energy or plane-of array irradiation with values obtained under standard test conditions (STC). $$PR = \frac{Y_f}{Y_r} = \frac{\sum_k P_{out_k} / P_{Stc}}{\sum_k \frac{G_{POA_k}}{G_{STC}}} = \frac{E_{out}/P_{STC}}{H_{POA}/G_{STC}}$$
Variations	**Temperature-corrected PR** The PR can additionally be corrected for temperature using power temperature coefficient, γ, provided either by the PV module manufacturer or obtained from time series data, to better reflect the actual outdoor performance of the modules and to decrease temperature-related seasonal variations. The correction should be performed according to standard IEC 61724-1:2021 [2]. PR_{Tcorr} is calculated by estimating the module temperature $T_{mod,k}$ at each time interval k and factoring in its difference with reference temperature $T_{mod,ref}$ using the module temperature coefficient (γ) [5]: $$PR_{Tcorr} = \frac{\sum_k P_{out_k} / P_{STC}}{\sum_k \left[\frac{G_{POA_k}}{G_{STC}} * \left(1 + \gamma * \left(T_{mod,k} - T_{mod,ref}\right)\right)\right]}$$ The reference temperature, $T_{mod,ref}$, can either be the annual weighted module temperature, or the STC temperature of 25°C. It is crucial that PR_{Tcorr} is calculated as a correction on the energy at the highest available data resolution (preferably sub-daily resolution), as correcting the PR at lower data resolutions will lead to erroneous results due to averaging issues. **Bifacial PR** The IEC 61724-1:2021 [2] proposes to correct the PR for bifacial PV systems, using the bifaciality coefficient at maximum power ($\varphi_{P_{max}}$) and the in-plane rear-side irradiance ratio at each time interval k (ρ_k): $$PR_{BI} = \frac{\sum_k P_{out_k} / P_{STC}}{\sum_k \left[\frac{G_{POA_k}}{G_{STC}} * \left(1 + \varphi_{P_{max}} * \rho_k\right)\right]}$$

	The bifaciality coefficient, or more commonly bifaciality factor, is the ratio of rear efficiency to the front efficiency when subject to the same irradiation [7]. The rear-side irradiance ratio ρ_k is the ratio between in-plane rear-side and front-side irradiance. If in-plane rear-side irradiance sensors are available, they can be used along with the front-side sensors to calculate ρ_k. IEC 61724-1:2021 rightfully warns about non-uniformity of rear-side irradiance and the need to place multiple sensors to capture this variability. Another option is to measure horizontal albedo or use albedo values from an accurate table according to the type of environment and use an optical model to estimate rear-side irradiance. Both options are complex and error-prone.
Contractual exclusions	The following types of contractual exclusions may be applied: • curtailment • planned system/component shutdown • force majeure, including grid outage • tracker wind-stow • missing data Periods of curtailment may be excluded, even if the curtailment is partial and thereby not causing unavailability. While only targeted at performance indices as described in [4], the concepts “in-service” (excluding all unavailability) and “external-cause-excluded” (excluding only external sources of unavailability) may be applied as well.
Application	The PR is the most commonly used KPI in Europe and other parts of the world (except US). No significant differences within Europe have been observed. Temperature correction is usually only applied in countries with warmer climate (e.g. Spain). It is a bankable KPI, as it is straight-forward and easily understood by all project stakeholders.
Advantages	The main advantage is that it is simple, clearly defined, and easily accessible to all stakeholders.
Challenges	There are well-known limitations to the PR. Among the challenges that are growing in importance are data losses and exclusions (e.g. due to curtailment or grid events), as systems increasingly operate with mandatory curtailment. For heavily oversized PV plants with a high DC-to-AC ratio, higher irradiation reduces the PR due to clipping. The main limitation is the temperature dependency. The temperature-corrected PR is used to lower the temperature dependency of the PR. In absence of module temperature measurements, the estimation of this parameter poses additional complexity.

2.1.3 Availability

Description	Availability KPIs measure the extent to which the plant was generating electricity throughout the period of examination [5]. The basic formula for technical availability of any component is [8]: $$A_t = \frac{T_{useful} - T_{down}}{T_{useful}}$$ Here, T_{useful} is defined as time periods when the irradiance is above a predefined threshold (daytime – based on elevation or irradiance thresholds), and T_{down} is a subset of this period in which the component under investigation is not operating. This formula can be used at plant level, but also at component level such as inverter, junction box or also for trackers. In case of component availability, the availability across the plant can be calculated by the availability sum of all components, $A_{t,k}$, weighed by the installed DC power of the component, P_k, and compared to the DC power of the plant P_{STC} [8]: $$A_{t\ total} = \sum_k A_{t,k} * \frac{P_k}{P_0}$$ The inverter is the most commonly used availability level. The time resolution is also an important aspect and is typically 15 minutes or 1 hour.
Variations	**Energy-based Availability** The definition of energy-based availability varies depending on the source. Two commonly used equations are [8, 4]: $$A_{energy} = \frac{Y_{exp} - Y_{unavailable}}{Y_{exp}} = \frac{Y_{exp,available}}{Y_{exp,total}}$$ $$A_{energy} = \frac{Y_{meas}}{Y_{meas} + Y_{exp,loss}}$$ Y_{exp} is the expected yield and Y_{meas} the measured yield. The first definition has the advantage that it's independent of the performance of the plant. In turn, the latter definition yields a lower availability if the measured yield during times of availability is lower. There are known examples in Asia where energy-based availability takes into account tracker downtime. The estimated loss of energy between the tracker-based irradiance profile and the irradiance profile in stow position is added to $Y_{unavailable}$.
Contractual exclusions	Regardless of the type of availability calculated, certain periods may be excluded from the calculation. In general, this comprises external sources of unavailability, including: • curtailment • planned system/component shutdown

	• force majeure, including grid outage • tracker wind-stow (only applicable for tracker availability or energy-based Availability) • missing data According to [4], missing data should always be substituted, as long as the total period is less than 1 week per year. However, it is common practice to fully exclude periods with missing data. Chapter 3 elaborates further on data cleaning and filtering approaches.
Application	Availability is typically a contractually binding KPI in operation & maintenance (O&M) and engineering, procurement, and construction (EPC) contracts. Both time-based and energy-based availability are commonly used. Availability KPIs are relevant for most stakeholders including O&M parties, asset managers (AM), investors and asset owners (AO). Additionally, they are used during PR liability periods for EPCs. Power purchase agreements (PPAs) can also specify a minimum availability, which is typically lower than the O&M requirements. Lastly, with most new utility-scale systems being installed on trackers, tracker availability is becoming more widely used.
Advantages	Time-based availability is well understood and therefore easy to use for all stakeholders, independently of their technical background. Energy-based availability has the advantage of being irradiation- and therefore energy-dependent, which means that outage times where high losses are recorded have a higher weight.
Challenges	In deregulated markets, energy may be worth less during times of high irradiation. Also, energy-based availability is less transparent, as it relies on a simulation model to calculate the expected energy. For those reasons, time-based availability is still the most contractually used availability KPI. Tracker-based availability is even more challenging to calculate, partially because of the typically wireless communication causing frequent data gaps or stalled data. Furthermore, safety-related stow times, for example due to high winds or hail, should be distinguished from outages.

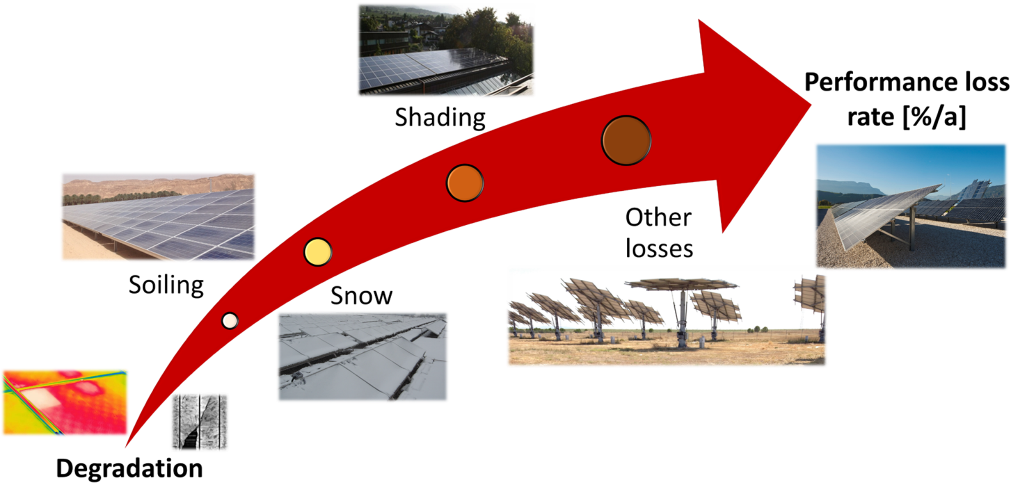
2.1.4 Soiling ratio

Description	The soiling ratio (SR) is calculated as the ratio of the measured power output of a soiled PV cell to the power output that would be expected if the array was clean. The soiling ratio calculation is far from being standardized. The IEC 61724-1:2021 [2] describes the soiling detection approach, but in very general terms. The soiling ratio can be calculated in various ways, including [9]: • optical soiling sensors • manual calculation of the ratio of power or short-circuit current of an uncleaned module versus a clean module • soiling extraction algorithms based on monitoring data

	• image analysis of aerial photos • calculation based on environmental factors
Variations	**Soiling level** The soiling level/loss (SL) is calculated as 1-SR. It describes the transmission loss of the soiled PV array/system. **Soiling rate** The soiling rate describes the daily value of the soiling ratio, if no cleaning occurs.
Application	The SR is used in several ways. O&M contracts may stipulate a maximum soiling loss. O&M operators may use detailed soiling data of their plants to do condition-based cleaning [10]. Asset owners may keep track of the soiling ratio per region to assess required actions and to improve projections for future yield. The relevance of soiling depends heavily on the climate. For example, in most of Europe, rainfall is frequent throughout the year. Due to simplicity, most asset owners define a fixed cleaning schedule or no cleaning at all. However, in desert areas, cleaning schedules should be optimized based on periods of heavy soiling to maintain efficiency.
Advantages	A reliably calculated SR allows for well-informed decisions on cleaning schedules and business plan projections.
Challenges	The main challenge is to get a reliable assessment of soiling losses. Each of the determination methods has advantages and disadvantages. For instance, soiling sensors require periodical maintenance, are an indirect way of measuring performance loss presenting limitations [11], and provide a low spatial resolution. Similarly, the frequency of cleaning for the reference clean module can affect the SR. Soiling extraction algorithms are susceptible to noise in the monitoring data, and the quality of the results depends heavily on the calibration and postprocessing steps of the specific algorithms.

2.1.5 Degradation and performance loss rate

Description	The degradation rate (Rd), often wrongly equated with the performance loss rate (PLR), describes irreversible (e.g. material degradation) losses that can occur in a PV plant and is an essential parameter for performance modelling, monitoring, and O&M [12]. Instead, the PLR represents all irreversible and reversible (e.g. soiling) losses a PV system can experience. Figure 2 depicts the relation between the two closely interrelated parameters. In general, the PLR is calculated by linearizing a selected performance metric and taking the slope of the linear model as the rate of performance loss. An accurate assessment of the PV module degradation rates requires standardized indoor tests.
Application	The Rd, and PLR, are relevant for PPAs and project finance (investor, AO, AM). The calculation requires several years of high resolution data and is generally used for project finance to feedback estimates and correct the finance model.

Ad-vantages	It is the primary KPI to predict and monitor the decreasing performance of a plant over time.
Chal-lenges	A reliable calculation of both Rd and PLR requires several years of data. While two years is the minimum to theoretically calculate this KPI; Jordan et al. [13] showed that PLR calculated even on five years of data with limited weather data accuracy are expected to have high uncertainties. Statistical approaches used to determine PLR and Rd are also highly influenced by climatic variability. To this day, there is no standardized way of calculating the Rd/PLR. Best practices have been developed by the IEA PVPS Task 13 [14]. It is very difficult to accurately divide Rd from PLR and can only be attempted if all reversible losses are determined and subtracted from the PV system performance.  **Figure 2: Relation between Rd and PLR; PLR includes all system-wide performance loss mechanisms while Rd describes only material degradation; taken from [12].**

2.1.6 Energy Performance Index

Descrip-tion	The Energy Performance Index (EPI) expresses the ratio of the measured yield and the expected yield: $$EPI = \frac{Y_{meas}}{Y_{exp}}$$ The calculation of the expected yield is a key aspect of calculating the EPI. While not specifying which performance model in particular, standard IEC TS 61724-3:2016 [4] stipulates the use of a performance model which, given the input weather data, is used to calculate the expected energy output at any given time. Ideally, the performance model is the same as the model being used during the design of the plant. This is not always practically possible: asset models may get lost as assets change hands, and performance modelling tools are generally not built for continuous re-computation based on measured weather data.

Contractual exclusions	The following types of EPI exist: • In-service EPI: excluding periods of unavailability • External-cause-excluded EPI: excluding only external sources of unavailability – the same exclusion sources listed under the availability indicator In [4], it is not clearly defined whether partial curtailment, which is not defined as unavailability, is to be considered an external cause.
Application	Technical asset managers, especially in Europe, are increasingly referring to the EPI as a complementary KPI to the PR. The contractual implementation of this KPI is slowly but surely increasing.
Advantages	For a well-performing system, the EPI typically is more stable (around 100%) compared to the PR, as it is less influenced by system design and weather-related variations, even when compared to the temperature-corrected PR. For instance, with a high DC-to-AC ratio, the PR may decrease due to inverter clipping, whereas the EPI remains unaffected.
Challenges	There are practical boundaries, e.g. the fact that design tools are not typically integrated in monitoring platforms, so that the EPI calculation is a semi-manual process. Nevertheless, the main reason for the slow adoption is the uncertainty. The EPI calculation requires more sophisticated and less straightforward equations compared to the PR, and as a result is also less transparent. This causes reluctance in using it. While the physical model to be used is not yet defined in the standard, machine learning models are emerging as alternative options for reliably calculating the expected yield. However, their inclusion may introduce more uncertainty when it comes to standardization, as these models can vary in methodology and outputs. Bifacial modules present the same challenges in terms of uncertainty to EPI as to PR.

2.1.7 Capacity test

Description	The Capacity test is described both in ASTM-E2848-13 [15] and in IEC 61724-2:2016 [3]. The two standards follow a different approach to arrive at similar results. Both evaluate the correlation between power (not energy) and weather conditions at reference weather conditions, which can differ from standard test conditions (STC). The estimated power at these reference conditions is then compared to a target power at those conditions. The ASTM approach uses multiple linear regression to compare measured power with expected power derived under reference conditions (irradiance G_{RC}, temperature T_{RC}, wind speed v_{RC}) according to: $$P_{RC} = G_{RC}(a_1 + a_2 \cdot G_{RC} + a_3 \cdot T_{RC} + a_4 \cdot v_{RC})$$ The coefficients a_1, a_2, a_3 and a_4 are determined based on measured data from the plant over a defined period. Either the specific reference conditions, or the method to calculate the reference conditions based on the data, are agreed upon contractually. On the other hand, the IEC method uses a non-

	regression comparison of measured to expected power. This approach utilizes plant design parameters to calculate a correction factor. This factor adjusts the measured performance to compare it against the targeted performance under reference conditions. IEC 61724-2:2016 states that the approaches can be used interchangeably. The IEC standard is currently under revision exploring the idea to account for bifaciality. In the implementation within the Python package pvcaptest [16], the ASTM method is used to calculate the $P_{RC,sim}$ based on simulated data, with the same reporting conditions. The capacity ratio is then calculated as $P_{RC}/P_{RC,sim}$. If the capacity ratio is within predefined bounds, the test is considered as passed.
Contractual exclusions	In the ASTM test, missing data, incorrect data and non-linear behavior, like clipping or curtailment, are excluded from the regression. The IEC test foresees an optional additional target under “constrained” operation for inverter clipping and curtailment.
Application	The capacity test is most commonly used in the US for acceptance testing of a PV plant.
Advantages	It could be considered a slightly more accessible alternative to the EPI. Instead of building a performance model and comparing real to theoretical performance, this method works the other way around. A performance model of the real plant is calculated and used to simulate the power output under reference conditions.
Challenges	The ASTM capacity test has similar challenges as the EPI in terms of lacking transparency. Additionally, the exclusion of all non-linear events, for example derating because of overheating, may give a less complete representation of the system performance compared to the EPI. Climatic variability and data quality may influence the regression resulting in a bias in the capacity ratio estimation. Furthermore, this test captures only a snapshot of performance over a year. As such, phenomena that average out on an annual basis (e.g., spectral effects) may introduce significant errors when not accounted in shorter timescales.

2.1.8 Capacity Utilization factor

Description	The Capacity Utilization Factor (CUF) is the ratio of the energy production to the maximum energy theoretically achievable given the nominal AC capacity of the plant [4]: $$CUF = \frac{E_{out}}{AC\ capacity\ \times 24\ \times days}$$ The CUF is usually in the range of 10% to 20%, with higher values found for systems installed in high-insolation regions using trackers and having high DC:AC ratios.
Variations	The term Plant Load Factor (PLF) is sometimes used instead of CUF in an operational context. Alternatively, the term Capacity Factor (CF) is used.

Applica-tion	It can be considered a different way to represent the P50 energy yield.
Ad-vantages	The CUF is technology-independent, so it can be used to intercompare PV plants with different energy production technologies.
Chal-lenges	CUF/PLF is highly dependent on the weather conditions (especially annual insolation), and much care must be applied to compare performance between different periods or geographical regions. For example, solar brightening (increase in solar irradiation, typically caused by a reduction in atmospheric aerosols, clouds, or other particles that scatter or absorb sunlight) can cause an upward trend in CUF, which may mask a decrease in plant performance.

2.1.9 Maintenance response time

Descrip-tion	The maintenance response time, described in detail in [5], is the time required by an O&M operator to have a technician dispatched to a PV plant after an alarm has been triggered (based on detection of a fault). The guaranteed response time depends on the severity of the fault. The resolution time, which describes the period between a technician's arrival and the resolution of the fault, is typically not guaranteed [8].
Variations	Other related KPIs are: • Mean time to repair [17] • Mean time before failure [17] • Failure rate [8]
Applica-tion	Maintenance response times are sometimes used in O&M contracts, though they are partially redundant with the availability KPIs.
Ad-vantages	It is a good internal KPI for O&M providers to track their performance and for asset managers/owners to verify the efficacy of the O&M team.
Chal-lenges	An advanced digital system is required to track this KPI reliably. In contrast to availability or other performance KPIs, the maintenance response time is not directly related to the revenues of the asset.

2.2 Economic KPIs

Although not really being KPIs, capital expenditures (CAPEX) and operational expenditures (OPEX) are important parameters to be aware of, as they drive most economic KPIs. CAPEX is the initial investment made by the asset owner. The main component is the Engineering, Procurement and Construction (EPC) cost. Other owner's costs include development costs, permitting costs, etc. [18]. OPEX represents the costs required to operate and maintain a plant over its lifetime. It includes fixed costs (including insurance and land lease), replacement of large components, maintenance and replacement costs [18].

2.2.1 Levelized Cost of Electricity

Descrip-tion	The Levelized Cost of Electricity (LCOE) is a broadly used way to compare the cost of different energy generation systems. The PV LCOE includes all the costs and profit margins of the whole value chain including manufacturing, installation, project development, O&M, inverter re-placement, etc. A parameter that will become even more important in future LCOE calculations is the system residual value, also called salvage value. In other words, the residual value corresponds to the possible earnings coming from the disposal of the power plant at the end of its life. It is like a revenue, which has the effect of reducing the overall costs of the plant because of the possible recycling and sales of the reused materials, thus decreasing the LCOE. PV LCOE also includes the cost of financing but excludes the profit margin of electricity sales and thus represents the generation cost, not the elec-tricity sales price which can vary depending on the market situation. A typical simplified formula is the following [19]: $$LCOE = \frac{CAPEX_0 + \sum_{t=1}^{N} \frac{OPEX_t}{(1+i)^t}}{\sum_{t=1}^{N} \frac{Production_t}{(1+i)^t}},$$ with $CAPEX$ and $OPEX$ as described in this section, $Production$ the energy yield per timepoint, and i the discount rate. Another, more detailed approach is [20]: $$LCOE = \frac{CAPEX + \frac{InvRepl}{(1+WACC_{nom})^{N/2}} + \frac{ResValue}{(1+WACC_{nom})^{N}} + \sum_{t=1}^{N} \frac{OPEX(t)}{(1+WACC_{nom})^{t}}}{\sum_{t=1}^{N} \frac{Yield_0 \cdot (1-PLR)^t}{(1+WACC_{real})^t}} \left[\frac{€}{kWh}\right]$$ Where *N* is economic lifetime of the system, *t* is year number ranging from 1 to *N*, $CAPEX$ is total capital expenditure of the system (at *t* = 0 in €/kWp), $OPEX(t)$ is operation and maintenance expenditure in year *t* in €/kWp, $InvRepl$ is the cost of inverter replacement (at *t* = *N*/2 in €/kWp), $ResValue$ is the residual value of the system at *t* = *N* in €/kWp (can be either positive or negative), $Yield_0$ is initial annual yield in year 0 in kWh/kWp (equivalent to P50 yield), $WACC_{nom}$ is nominal weighted average cost of capital per annum and $WACC_{real}$ is real weighted average cost of capital per annum.
Variations	ACOE (Actual cost of electricity) [21] is a metric where the actual electricity demand and curtailment are considered, aimed as an expansion or correction to the LCOE.

Applica-tion	LCOE is a metric that can compare various generation technologies (see challenges below for completeness). LCOE is also used as the benchmark for developers to bid in auctions and it is used to develop business models and calculate Net Present Value (NPV) and IRR. LCOE is also used as a floor value to define the profitability of investments.
Chal-lenges	Even though the LCOE is intended to compare generating technologies on a neutral basis, the assumptions used can affect results strongly. For example, front-loading O&M expenses, or distributing these over the project life, will affect the LCOE [22]. One of the most important factors that affect the LCOE is the (assumed) discount rate [22, 20]. There are many ways to calculate the components of the LCOE, with various degrees of sophistication and detail. For example, Lazard's LCOE is calculated by "creating a power plant model representing an illustrative project for each relevant technology and solving for the $/MWh value that results in a levered IRR equal to the assumed cost of equity" [23]. By contrast, NREL's System Advisor Model has different methods to calculate the LCOE, depending on the project type and financing structure [24]. The LCOE is a techno-economic parameter used to evaluate the cost of a kilowatt-hour of energy produced from a selected power plant. The most typical approach to calculate the LCOE does not account for the interaction of the new power plant with the existing energy system, assuming indirectly the power plant as stand-alone. Some research studies have been proposing methodologies to overcome these limits by estimating the impacts of high VRES penetration on the LCOE. They have determined the so-called integration costs that can be combined with the LCOE to include the impacts of adding new intermittent generation in the existing energy systems. In this way, a new parameter can be defined (system LCOE) which include the integration costs as the sum of balancing costs, grid costs, adequacy costs (or backup costs), full-load hours reduction, and overproduction costs [25]. Other more recent attempts to go beyond the classical LCOE formulation are represented by Levelized Avoided Cost of Electricity developed by the US Energy Information Administration and the value-adjusted LCOE (or VALCOE) developed by the IEA, as well as the ACOE. The LCOE as a metric can continue to be used, even in the context of very high levels of curtailment and/or negative prices. It does, however, require increasingly sophisticated (sub)models for the metric to be calculated correctly. For example, the denominator of the LCOE is essentially an energy sales forecast model which is brought to the present, which historically assumed 100% of possible energy generation to be sold. Today, very few systems worldwide have such a certainty over the technical or financial lifetime and must therefore include curtailment for the correct calculation of the LCOE.

2.2.2 Net Present Value

Descrip-tion	The Net Present Value (NPV) is the discounted difference between the present value of cash inflows and the present value of cash outflows over a period of time and it is used to analyze the profitability of an investment or project over its financial lifetime. The NPV gives an idea of how much revenue a solar project will bring, accounting for the time value of money. The NPV is displayed in currency and is thereby an absolute measure. It is calculated as: $$NPV = \sum_{t=1}^{N} \frac{cash\ flow_{PV}}{(1+i)^t} - C_0.$$ Here, $cash\ flow_{PV}$ represents the annual net of money earned through selling electricity vs. spent on the PV project. Expenditures can be, among other things, OPEX costs or loan payments. i represents the interest rate, t number of years the system is expected to operate, and C_0 the initial investment.
Variations	A simplified evaluation of a PV project is the return of investment (ROI). It describes the time for a project to pay for itself and is the ratio between net income and investment. The issue is that ROI does not account for factors such as inflation, depreciation, OPEX, project lifetime, and other relevant considerations, which are factored in when calculating metrics like NPV or IRR.
Applica-tion	The NPV is used to assess the profitability of investing in a solar project. If the NPV is positive over the lifetime of the PV system, the project can be considered profitable.
Chal-lenges	Inflation as well as discount rates must be assumed, which affects the final NPV value.

2.2.3 Internal Rate of Return

Descrip-tion	IRR, or internal rate of return, is a metric used in financial analysis to estimate the profitability of potential investments. IRR is a discount rate that makes the NPV of all cash flows equal to zero in a discounted cash flow analysis. The IRR is given in percentage and expresses the annual return over the lifetime of the PV project. The IRR can be calculated when setting the NPV formula to 0 and solving the equation for the discount rate i.
Applica-tion	It represents at what rate the project would have to make money in order to break even over the lifetime of the PV project.
Ad-vantages	The IRR is used as decision-making KPI, as the percentage value makes for easy comparison with other investments under consideration.

2.3 Sustainability KPIs

2.3.1 Global Warming Potential

Descrip-tion	Mitigating climate change is one of the main reasons for the large-scale deployment of solar PV. The impact of electricity generation on climate change is assessed using the Global Warming Potential (GWP). GWP is expressed in units of carbon dioxide equivalents, a standardized measure that quantifies the global warming impact of various greenhouse gases (mainly CO_2, CH_4, N_2O, and SF_6) based on their radiative effects in the atmosphere. This unit provides a consistent way to compare the climate impact of different gases on a common scale. For comparing electricity generation technologies, it is mostly expressed as carbon intensity per kWh of generated electricity, e.g. in gCO_2-eq/kWh. This metric is determined by calculating the total gCO_2-eq released during the lifecycle of the PV system and dividing it by the total lifetime electricity generation in a lifecycle assessment (LCA) analysis.
Applica-tion	The GWP of electricity generation is mostly used in research, to compare different forms of electricity generation. This knowledge of the GWP of electricity generation is driving the transition of fossil fuel-based electricity generation to lower GWP electricity sources like PV. Within the context of PV, the GWP is further applied to compare different types of existing and prospective PV technologies. PV module producers are increasingly publishing so-called environmental product declarations (EPDs), showing the environmental impact of the module production, determined based on LCA, including the GWP.
Ad-vantages	The advantage of the GWP is that it expresses the impact of electricity generation on climate change, and as such allows the comparison of different technologies of electricity generation with each other. Since climate change poses one of the main challenges for our energy system, this metric evaluates the suitability of an electricity generation source within this context and compared to other sources of electricity. By performing LCA analyses, it also allows us to evaluate the effects of improvements in PV technology on its GWP.
Chal-lenges	Like other environmental impacts, the GWP can only really be determined by means of an LCA analysis, which is a challenging and time-consuming activity. It requires the collection of data on manufacturing processes along the whole lifecycle of a PV system. There are frequent efforts within IEA PVPS Task 12 to collect up-to-date life cycle inventory data for different PV technologies, but due to the fast development of the PV market and PV technologies this remains a difficult endeavor. Another issue with the GWP is that it expresses the environmental impact of PV electricity in only one environmental impact category. Hence, the GWP is not a suitable indicator to assess the full environmental impact.

2.3.2 Energy payback time

Description	The energy payback time (EBPT) is defined as the time required for a renewable energy system to generate an amount of electricity that is equivalent to the amount of energy needed to manufacture and install the PV system. Since the manufacturing of a PV system requires different forms of energy (e.g. electricity and heat), the energy terms should be expressed in primary energy equivalents. The EPBT is calculated as [26]: $$EPBT = \frac{E_{mat} + E_{manuf} + E_{trans} + E_{inst} + E_{EOL}}{\frac{E_{annual}}{\eta_G} - E_{OM}}$$ Here, E_{mat}, E_{manuf}, E_{trans}, E_{inst}, E_{EOL} and E_{OM} are the primary energy demand for the used materials, the manufacturing process, transport of the system components, installation, end-of-life treatment and annual operation and maintenance, respectively. E_{annual} is the mean annual electricity generation. The parameter η_G describes the grid efficiency, from primary energy to electricity, for the grid as a whole. Determining the value of η_G is one of the key methodological challenges of the EPBT. It should be determined based on the scope of the study. Currently, it is typically fair to assume that the electricity generated with PV systems is replacing non-renewable electricity generation, and η_G should reflect the primary energy to electricity ratio of non-renewable electricity generation. However, as the penetration of PV in electricity grids increases, it will be more important to consider the η_G of marginal electricity mixes.
Variations	Two variations are used: the regular energy payback time (EPBT), and the non-renewable energy payback time (NREPBT). These should be applied depending on the assumption of whether the generated electricity of the PV system under study will replace all sources of electricity in the grid where it is to be installed (regular EPBT) or solely the non-renewable sources of electricity in the grid where the PV system is to be installed (NREPBT). The NREPBT goes beyond the methodological choice of η_G as discussed above, and additionally only considers non-renewable energy use in the demand side and operational phase (E_{mat}, E_{manuf}, E_{trans}, E_{inst}, E_{EOL} and E_{OM} from the equation above). For more information see a recent report developed in IEA PVPS Task 12 by [26].
Application	The EPBT is commonly applied, together with the GWP, to express the environmental impact of PV electricity in a simple and relevant manner. The EPBT was first used around the time of the oil crises, when it seemed that energy scarcity was one of the main challenges of the global energy system. Since the EPBT depends so heavily on the annual electricity generation, it is a method by which to show the lifecycle energy efficiency of installing a PV system in a certain location.
Advantages	One of the main advantages of the EPBT is that it is a clear metric that is relatively easy to understand, that expresses the energy efficiency of a PV system installed in a certain location. It is a reasonable proxy of environmental impact, as energy efficiency and environmental impact are reasonably well correlated. However, specific environmental impacts are not necessarily well represented

	by an energy-based indicator. Like the GWP, the EPBT can be tracked over time to show the effects of technological progress in PV on reducing the environmental impact.
Challenges	Like the GWP, the EPBT can be only properly determined by performing a full lifecycle assessment of the system under study. As mentioned before, performing an LCA is a complex and time intensive process. Additionally, the EPBT does not account for system lifetime, hence two PV systems with an equal EPBT could have very different lifetimes and as such a substantially different environmental impact. In this case, the EPBT is not a good metric to compare these systems. Finally, the choice of η_G strongly affects the final EPBT value calculated, but choosing and determining the right value can be challenging.

2.3.3 Energy Return on (Energy) Investment

Description	The Energy Return on (Energy) Investment (EROI) was developed to complement the EPBT and address some of its issues. It describes the ratio of the useful energy output of the PV system to the total energy invested over its lifetime. In general terms, the EROI is defined as [27]: $$EROI = \frac{Out}{Inv}$$ Here, *Out* refers to the lifetime energy output of the PV system, and *Inv* to the total of invested energy during the lifetime (e.g., the sum of E_{mat}, E_{manuf}, E_{trans}, E_{inst}, E_{EOL} and E_{OM} from the EPBT equation). A high EROI value indicates an energy-efficient PV system that generates substantially more energy over its lifetime compared to the total energy invested.
Variations	In the context of EROI, two definitions are presented in literature [27]: the $EROI_{el}$, where *Out* and *Inv* are the direct energy units of the delivered form of energy, or $EROI_{PE\text{-}eq}$, where *Out* and *Inv* are both converted to primary energy, as with the EPBT. The choice of which to use is not trivial, and the calculation methods and assumptions underpinning them are not either. For more information, see the report from IEA PVPS Task 12 by Raugei et al. [27].
Application	The main application of the EROI is to assess the energetic balance of a PV system over its entire lifecycle. It is a metric that allows us to evaluate PV in the context of energy security, and to compare PV electricity to other forms of electricity generation, both renewable and non-renewable.
Advantages	Like EPBT, the EROI is a single, clear, and relatively easy to understand metric to express the energy efficiency of a PV system. It is similarly also a reasonable proxy for environmental impact, but as mentioned for the EPBT, not a complete indicator for it.
Challenges	As with the other KPIs, the EROI can only be determined by performing a complex and time-consuming LCA and expresses environmental impact only in one impact category: energy.

2.3.4 Environmental Footprint 3.1 - Single Overall Score

Descrip-tion	The Environmental Footprint 3.1 (EF3.1) method is the latest version of the Environmental Footprint (EF) method developed by the European Commission [28]. The method aims to offer a comprehensive environmental impact assessment to be used in LCA studies, to perform Product Environmental Footprint (PEF) and Organization Environmental Footprint (OEF) evaluations. The EF method determines environmental impact in 16 impact categories in topics such as climate change, resource depletion, human health and ecosystem quality, and water and land use. The European Commission recommends that for a comprehensive PEF the EF3.1 method is applied, however, with 16 impact categories the interpretation and communication of results is complex. Hence, from version EF3.1 the method includes a first proposal of combining the impacts into a single score, by using weighting factors for each impact. The determination of the EF3.1 Single Overall Score is complex and is based on a full LCA study of the PV system lifecycle.
Applica-tion	The EF3.1 Single Overall Score is a KPI that is not commonly applied in the PV sector, especially compared to the many KPIs in this report. We present it here as a potential alternative to the single-impact sustainability KPIs discussed above.
Ad-vantages	Compared to the other sustainability KPIs, the EF3.1 single overall score includes the environmental impact in a broad set of impact categories, allowing to get a single score for the environmental impact of PV systems or PV electricity.
Chal-lenges	This KPI suffers from the same issue as the other sustainability KPIs: its calculation can only be performed by means of a complex and time-consuming LCA study. Additionally, the process of weighting and aggregation of the 16 EF3.1 impact indicators into a single score is a procedure for which there is no real consensus in the scientific community, as the authors of the first European Recommendation for the weighting factors also acknowledge, stating that the choice of the weighting factors is inherently not natural science-based, but rather a result of value choices [29].

2.4 Standard technical KPIs per stakeholder

In order to reliably map the different technical KPIs used per stakeholder and per region, PV industry experts have been approached using the following different channels:

- One-on-one conversations with contacts in the industry
- A workshop at Solar Quality Summit Europe 2024, with 60 participants
- A survey which was disseminated via LinkedIn, with 23 respondents.

Most of the information was obtained from experts of a large variety of countries in Europe, which made up about 90% of respondents. The US was represented by 4 respondents, whereas other regions had only one respondent, or no respondents at all in the case of South America and Africa. About 40% of respondents were Asset Owners or Asset Managers, with other roles well represented by at least 2 respondents.

Initially, a clear trend in regional variations across different markets and the use of certain KPI variations was expected. However, these assumptions were relativized when compiling the results:

- **Regional variations**: While trends do exist within regions, a globalized world and market mean that no strict differences are apparent.
- **Contractual KPIs**: There is no consistent pattern in which KPI variations are used contractually. For example, both PR and temperature-corrected PR, as well as both time-based and energy-based availability, are used depending on the specific contract.

Table 2: Usage overview of technical KPIs.

KPI	Abbreviation	Private equity / Bank	Project Developer	Asset Owner / Asset Manager	EPC	O&M	Service provider / consultant
Pxx energy yield	**P50 Yield**	T/C	T/C	T/C	T		T/C
Performance ratio	**PR**			T/C	T/C	T/C	T/C
Availability				T/C	T/C	T/C	T/C
Soiling ratio	**SR**	T	T	T	T	T	T
Degradation rate	**Rd**	T	T	T	T	T	T/C
Performance loss rate	**PLR**	T	T	T	T	T	T/C
Energy performance index	**EPI**			T/C	T/C	T	
Capacity test	**CapTest**			T/C	T/C	T	
Capacity utilization factor	**CUF / PLF**				T	T	T
Maintenance response time	**MRT**			C	C	C	C

T – technical, C – contractual binding

Hence, the overview below, where KPIs are mapped along the PV value chain, is combined for all regions and for all variations of a single KPI. Where applicable, different stakeholders were merged because the same KPIs apply. Insurance companies, product manufacturers and authorities were excluded because technical KPIs were not applicable and/or because no information was obtained. Whenever there were inconsistencies in the data, the authors used experience and best judgment to decide.

Table 2 summarizes the technical (T) and contractual (C) use of technical KPIs.

3 DATA PROCESSING

This chapter discusses the crucial implementation of data preparation or data processing, so that the KPIs discussed in the last section can be calculated reliably. Data processing for PV projects involves several steps to transform raw data into meaningful information that can be used for decision-making. One important, although not exhaustive, reference for PV data analysis including data terminology, equipment, and methods is the standard IEC 61724-1:2021 [2]. The standard is designed to apply to a wide range of PV systems, including both grid-connected and off-grid systems.

The standard covers several key areas:

1. Measurements: The standard provides guidelines on what measurements should be taken to monitor the performance of a PV system. This includes measurements of irradiance, temperature, and wind speed.
2. PV System Performance: The standard provides guidelines on how to measure the performance of the PV system itself. This includes measurements of DC power and energy (before the inverter), AC power and energy (after the inverter), and system efficiency.
3. Data Acquisition: The standard provides guidelines on how to collect and store the data from the PV system. This includes recommendations on data acquisition systems, data logging intervals, and data storage.
4. Performance Indices: The standard provides guidelines on how to calculate performance indices from the collected data, such as the PR. These indices provide a measure of the performance of the PV system.
5. Uncertainty: The standard provides guidelines on how to calculate the uncertainty of the measurements and performance indices. This helps to provide a measure of the reliability of the data.
6. Reporting: The standard provides guidelines on how to report the collected data and calculated performance indices. This includes recommendations on what information should be included in the report and how it should be presented.

In summary, IEC 61724-1:2021 provides a comprehensive set of guidelines for monitoring the performance of PV systems. It helps to ensure that performance data is collected and reported in a consistent and reliable manner, making it easier to compare the performance of different PV systems, and to identify any issues that may need to be addressed.

Figure 3 depicts the general data processing steps. The first step is data collection. This involves gathering data from various sources such as solar irradiance sensors, temperature sensors, inverters, power meters, and other system monitoring tools. The data collected includes parameters like solar irradiance, ambient and module temperature, wind speed, system output (DC and AC), and system downtime. Once the data is collected, it is logged and stored in a database. The data logging system should be reliable and secure to prevent data loss. The next step is data cleaning, which involves checking the data for errors or inconsistencies and correcting or removing them. This includes dealing with missing data, outliers, or data that is clearly incorrect. Data analysis follows data transformation. This involves analyzing the processed data to extract meaningful insights. This involves calculating KPIs, identifying trends, or comparing actual performance with expected performance. The results of the data analysis are then interpreted in the context of the PV project. This involves identifying issues that need to be addressed, assessing the performance of the PV system, or making decisions about future operations. The results are then reported in a clear and understandable format. This may involve creating a dashboard, generating a report, or presenting the results in a meeting. The report should include key findings, interpretations, and recommendations. Finally, based on the insights gained from the data, actions are taken to improve the performance of the PV system. This may involve adjusting the operation of the system, carrying out maintenance, or making changes to the design of the system.

As shown in Figure 3, data processing should be an ongoing process, with data being collected and analyzed regularly to monitor the performance of the PV system and make informed decisions. In the following section, important data processing concepts are discussed focusing on the first four points: data collection, data logging, data cleaning and data aggregation.

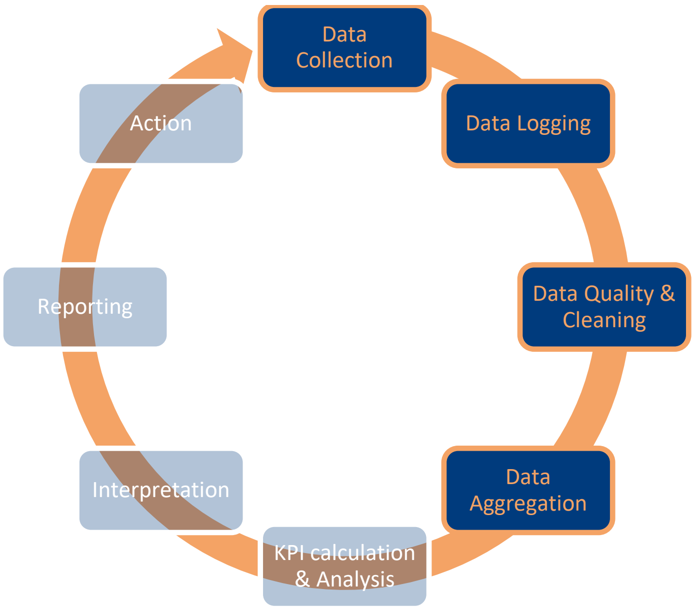


Figure 3: Preparational stages in data processing cycle.

3.1 Data collection

Apart from advanced measurements using imaging techniques or similar approaches, measured data types for PV system performance assessments can be generally divided into two groups: electrical data coming from PV system components and weather data from on-site installed sensors. These sensors are either attached to the PV system itself (modules, frames, structure) or are integrated into a dedicated weather station. Basic measurements include plane-of-array (POA) irradiance, ambient temperature, module temperature and wind speed and direction. Optional measurements are global horizontal irradiance (GHI), direct normal irradiance (DNI), diffuse horizontal irradiance (DHI), rainfall and relative humidity (RH). Complementary modelled data based on satellite data can be added to impute missing data or improve the overall data quality. It is important to obey certain taxonomy guidelines as best as possible. The names and abbreviations of some parameters are defined in standards such as the IEC 61724-1:2021 [2]. Users can also refer to specifically developed guidelines such as the Orange Button Solar Data Standard [30], the DuraMAT pv-terms project [31], or the TRUST-PV Risk Matrix [32], which is a standardized taxonomy to improve and standardize the categorization and the readability of O&M tickets.

The geographical location of a PV system and the complexity of the project play a significant role in determining the environmental measurement requirements. Different PV applications, such as agrivoltaics or floating PV systems, face distinct environmental challenges compared to traditional ground-mounted PV systems. It’s essential to identify and monitor these stressors as accurately as possible. For instance, in an agrivoltaic system, where crops grow beneath solar panels, specific sensors might be needed to monitor factors like soil moisture, humidity, and temperature to ensure both the farming activity and the PV system operate efficiently. In addition, depending on the climate, a PV system in a hot, dry environment may accumulate more soiling than one in a region with consistent rainfall, which could require the installation of dedicated soiling sensors.

3.2 Data logging

Standard IEC 61724-1:2021 [2] lays out guidelines concerning PV system performance monitoring. Depending on the PV plant size, weather monitoring systems are divided into three classes based on measured variables and performance assessment types. Class A signifies the most comprehensive data monitoring and quality, while Class C represents the least. The selection of the class is a trade-off between data accuracy and cost, and usually the monitoring system quality tends to increase together with the size of a PV project. A reliable and well-functioning monitoring system is an indispensable component to calculate most KPIs, which are used to evaluate PV system performance. This in turn enables personnel to detect/verify possible deviations from the norm to trigger and implement performance improving actions. One general issue when computing KPIs is that the measurements coming from the installed sensors must be trusted. Certain data tests can be executed to verify the reliability of the sensors, discussed in greater detail in chapter 3.3. Apart from following calibration and cleaning recommendations, the insertion of a certain redundancy by having more sensors than the bare minimum is another way of improving data coverage and quality.

Arguably the most important measurement for PV system assessment comes from reliable plane-of-array, or in-plane, irradiance measurements. These are recorded using either thermopile pyranometers or photovoltaic reference devices. Guidelines for device use, calibration intervals, and potential measurement adjustments are covered in IEC 60904-2:2015 [33] and IEC 61724-1:2021 [2]. The placement of the sensors in the PV system will contribute to the accuracy of the measurements in comparison to the PV system power measurements. A recent study by OTT HydroMet has shown that irradiance sensors should be placed away from PV plant edges, at least 5 rows away from the north/south end and not along the most eastern or western row, to minimize induced measurement errors [34]. This also means that sensor measurements installed at dedicated weather stations are subject to even higher uncertainties.

While both pyranometers and PV reference cells serve the purpose of measuring solar irradiance, PV reference cells offer certain unique advantages. One of the key benefits of PV reference cells is their possible similarity to the PV modules they monitor. If the reference cell material is the same as the material of the selected PV module technology, they respond to changes in irradiance, temperature, and other environmental conditions in a manner akin to one another, thereby providing a more accurate representation of the module's performance. Unlike pyranometers, which measure total solar irradiance, PV reference cells of the same material possess a spectral response like that of PV modules. This makes them more suitable for monitoring the performance of PV systems, as they provide a more precise measurement of the irradiance that the PV module can convert into electricity. However, pyranometers are generally of higher quality and offer more accurate and stable long-term measurements, making them the preferred choice in most cases, especially when precise irradiance data is needed for performance analysis and system comparisons. Regardless of the choice of sensors, periodic calibration and maintenance is essential to ensure reliable readings.

After irradiance, temperature is the most significant and common climatic factor affecting PV system performance. Measuring module temperatures accurately over long periods of time in outdoor conditions is a challenge, as the temperature sensors are exposed to the module temperature cycles and are required to remain in place over time. Temperature data are captured using sensors such as thermo-couples or resistance-based devices like Pt100 (where Pt denotes platinum). While ambient air temperature must be sufficiently ventilated and shielded from solar radiation, PV module temperature sensors are affixed at the module's back, requiring strong adhesion for accurate readings. Adhesives with thermal conductivity are used for extended outdoor use, aiming for uncertainties below 2°C. The number of sensors depends on the plant size as defined in IEC 61724-1:2021 [2]. Industry experience has identified that the center of the module, of a module in the center of the array, will best approximate the (average) module temperature of the array [35]. Temperature sensors are thus best avoided at module and array edges. IEC 61724-1 recommends temperature sensors to be replaced or recalibrated regularly, with temperature sensors being recalibrated every two years in Class A systems, while less clarity exists for Class B and Class C systems [2]. However, typical module temperature measurements are often permanently fixed to module backsheets with tape or glue (and possibly a high-conductivity thermal conductive paste). The mode of sensor placement is important, as the sensor itself may disturb the module, e.g. by creating a local insulation point, which then results in localized heating, as well as introducing different temperature measurement dynamics [36]. Permanent installation of temperature sensors can result in mixed class monitoring systems according to IEC 61724, where for example pyranometers and reference cells are recalibrated on a regular basis to Class A or B, while temperature sensors remain fixed in place (and thus are categorized as Class B or C).

In addition to irradiance and temperature, wind speed is an important parameter to monitor in PV systems, as it can significantly affect the cooling of PV modules and, therefore, their overall efficiency. Additionally, extreme winds may pose mechanical stress on the panels, mounting systems, and other infrastructure, making it essential to track wind data to assess the potential impact on system durability. Wind speed measurements are carried out using anemometers. Challenges with wind speed measurements include wind gusts occurring over short bursts (0.1 s – multiple seconds), while typical SCADA data storage resolution is at one minute average resolution or longer. Such wind gusts may thus cause trackers to move into a safe wind stow position, while the root cause is not easily detectable from stored data.

A more recent development is the usage of soiling sensors. The reliable detection of soiling losses, which can enable optimized cleaning schedules, is very difficult, as soiling losses are often masked within other PV system losses. Soiling sensors can help to achieve more reliable results. IEC 61724-1:2021 [2] defines the functionality and calibration of a soiling sensor. The described soiling detection approach is experimental and has several drawbacks which can influence the reliability of the results. For example, the measured soiling ratio will always depend on the cleanness of a reference device, or the physical properties of the devices used. That is why several soiling sensors have been developed in recent years based on different physical approaches. This in turn makes it harder to follow standardized maintenance schedules and intercompare results of different devices. Reliable soiling measurements are still subject to ongoing discussions.

Data loggers are used to record the data collected by the sensors. They can store data over time for later download or transmit data in real-time to a remote server. Power meters are used to measure the power output of the PV system, measuring both DC power before the inverter and AC power after the inverter. Modern inverters often come with built-in data logging capabilities. They can record data on the power output of the PV system and other parameters such as voltage and current. Weather stations can provide additional data that may be useful in analyzing the performance of the PV system, such as rainfall, humidity, and atmospheric pressure. Power plant controllers are used to regulate, and control connected inverter, devices, and equipment to meet certain setpoints and to also change grid parameters at the point of interconnect. SCADA systems, equipped with protocol adapters and device drivers required to communicate with all connected devices and equipment, are used to collect, order, store and visualize all data efficiently and in real-time. Servers, either local or cloud-based, are used to store the data collected by the data loggers, power meters, inverters, and other devices.

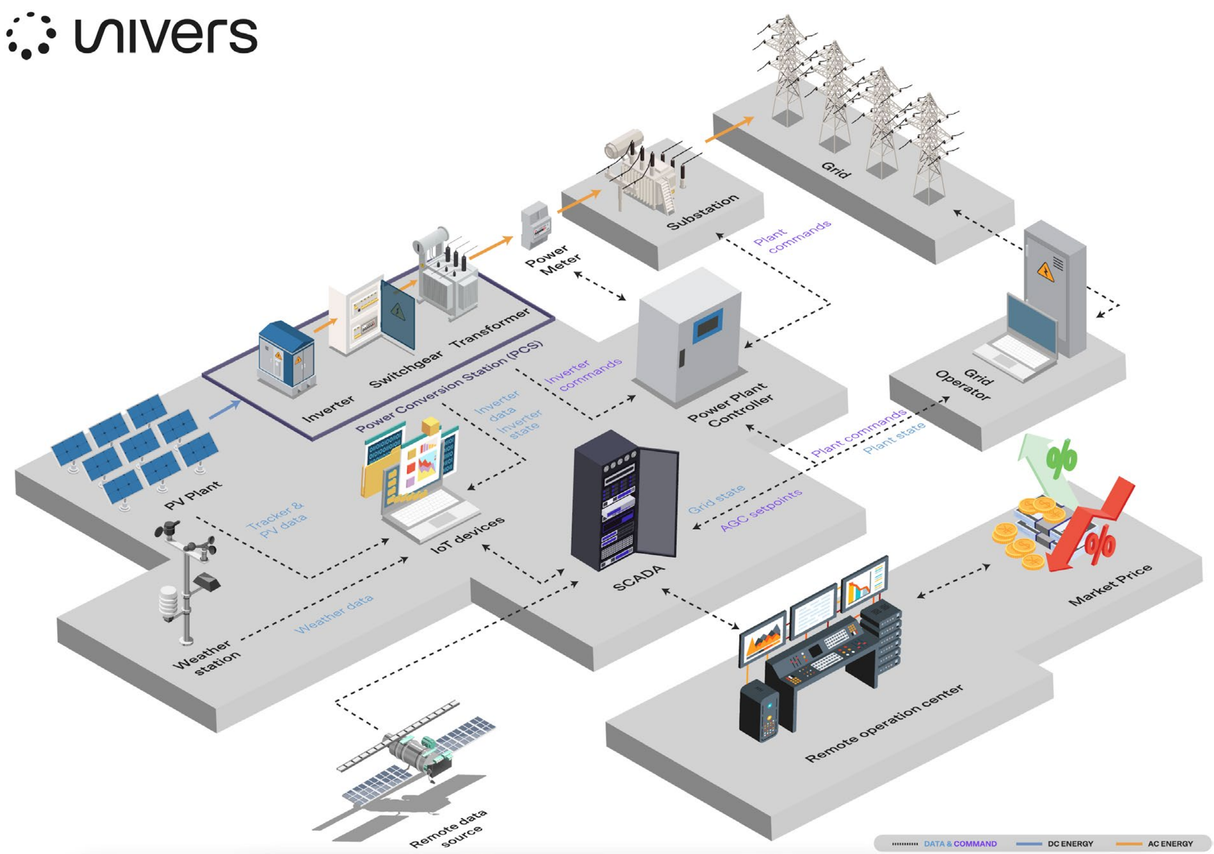


Figure 4: Components and connections in a PV plant [37].

Computers are used to analyze the data and generate reports. They can run software for data analysis, data visualization, and performance modeling. Communication devices like modems or routers are used for transmitting the data from the PV system to the server, especially in the case of remote monitoring. An uninterruptible power supply (UPS) can be used to ensure that data collection and transmission can continue even in the event of a power outage. The specific hardware required can vary depending on the size and complexity of the PV project, as well as the specific data collection and analysis needs.

3.3 Data quality & cleaning

While this section is aimed primarily at stakeholders (asset owners, EPCs, O&M providers) responsible for ensuring that data is of the highest possible quality, it may also be useful to other stakeholders in understanding the importance of these steps and philosophies to achieve better outcomes. This chapter discusses data quality concepts, the impact of data quality on result reliability, data filtering and imputation approaches, and introduces a unified data quality routine.

Data cleaning and data filtering are crucial steps to receive reliable results in any data driven evaluation. High-quality data is crucial for obtaining accurate and meaningful insights from PV system monitoring and analysis and the cleaning process plays a vital role. Data quality in the context of PV systems can be categorized into the following aspects:

- Accuracy
- Completeness
- Consistency
- Timeliness
- Reliability

Accuracy is directly linked to the quality of the measurements being performed. It expresses the trust we can have in a measurement. Measured input data coming of PV systems and associated weather stations are the basis for all downstream performance analyses. Therefore, highly accurate data are indispensable. Measuring the accuracy of measured data can be very challenging. Common approaches include logical thresholds depending on the measurements source, known physical relations between different parameters and comparison with other nearby measurements of the same kind, i.e. peer-to-peer assessments.

Completeness of data is essential to provide a seamless analysis of the studied subject. In the context of PV system performance analytics, long outages of measurements or missing measurements result either in incomplete calculated performance indicators or the usage of imputation techniques to recover missing data, which will have unquestionably an impact on the overall quality of calculated results. An example of the effect of missing data on the reliability of performance loss rate (PLR) calculation results is discussed in chapter 3.3.1. A minimum threshold of data completeness must always be provided to produce an accurate performance evaluation. The level of completeness will depend on the task at hand.

Consistency describes data integrity across different sources, systems, or calculation steps. A measured data point must be the same, regardless of where or when we query the measurement. Consistency secures that evaluations accurately capture and utilize the data value.

Timeliness of data is affected by the period between data collection and processing, the time resolution of measurement of each data point, and the communication speed of the SCADA system, as well as the sensors to the SCADA. Synchronization of data is an important aspect, as it affects performance evaluations, and it may affect control of a power plant, e.g. the time taken for a curtailment signal to travel through the plant, interpretation by inverters, and its measurable effect at the plant connection point.

Data reliability combines completeness and accuracy.

Usually, data quality is evaluated across different data quality aspects. PV power data quality criteria specifically for PLR analyses have been suggested by the IEA PVPS Task 13 [14]. Here, PV power input data are graded depending on the number of outliers, the amount of missing data and the longest gap in the time series. It was seen that these are the most critical data quality parameters for this specific KPI.

Data cleaning and filtering goes hand in hand with data quality, as it improves the quality through removing a part of the data, which would otherwise impact the reliability of the analysis carried out or the KPIs calculated. However, if the raw data quality is too low to begin with, cleaning and filtering may remove too much data so that the remaining dataset is no longer representing the inherent information the data was carrying. Cleaning is applied to ensure and improve the data quality categories. Filters are applied to remove unnecessary data, outliers, and, in the case of many PV-related KPIs, data which deviate from the known or expected relationship between the output of a PV system on one hand and weather inputs (irradiance, temperature, spectrum, …) on the other. This step is very complex, as there is ongoing debate of what are good vs bad data and each analysis/KPI might require a different approach. The impact of different filtering criteria on performance time series of a PV system was studied by Lindig et al. [38]. As laid out in this work, IEC standard 61724-3:2016 [4] provides a set of basic range filter for AC power and weather parameter where data have a 15-min resolution:

Table 3: Recommended filter IEC 61724-3:2016 [4].

Minimum		Parameter		Maximum
-0.01 * P_{STC}	<	AC power	<	1.02 * P_{STC}
-6 W/m^2	<	Irradiance	<	1500 W/m^2
-30 °C	<	Ambient temperature	<	50 °C
0 m/s	<	Wind speed	<	32 m/s

Furthermore, the standard specifies to detect and omit missing data or duplicates, stuck values and data with abrupt changes. Stuck, or dead, values as well as abrupt changes are detected using derivatives. The application of these filters should be the minimum requirement of a cleaning process but will often not be sufficient. According to the standard, erroneous data can be either filtered out or imputed. Once detected, missing data, outliers, and duplicates are essentially treated equally. As such, erroneous values can be derived from interpolation or other imputation techniques. The primary drawback of the IEC 61724 lies in its qualitative description, lacking a case-specific approach that could facilitate reproducible and unbiased outcomes. For example, the standard does not specify at what levels of erroneous data imputation or filtering should be conducted.

Figure 5 provides a detailed overview of input variables, common data issues and filtering approaches to choose from. The selection of individual cleaning/filtering strategies will depend on the target. Nevertheless, this table serves as a common guideline framework. The Python package pvanalytics [39] is an easy-to-use toolbox where many of the mentioned strategies are formalized. Other recommended tools which can help analysts in studying PV system data are pvlib [40], RdTools [41] and Solar Data Tools [42].

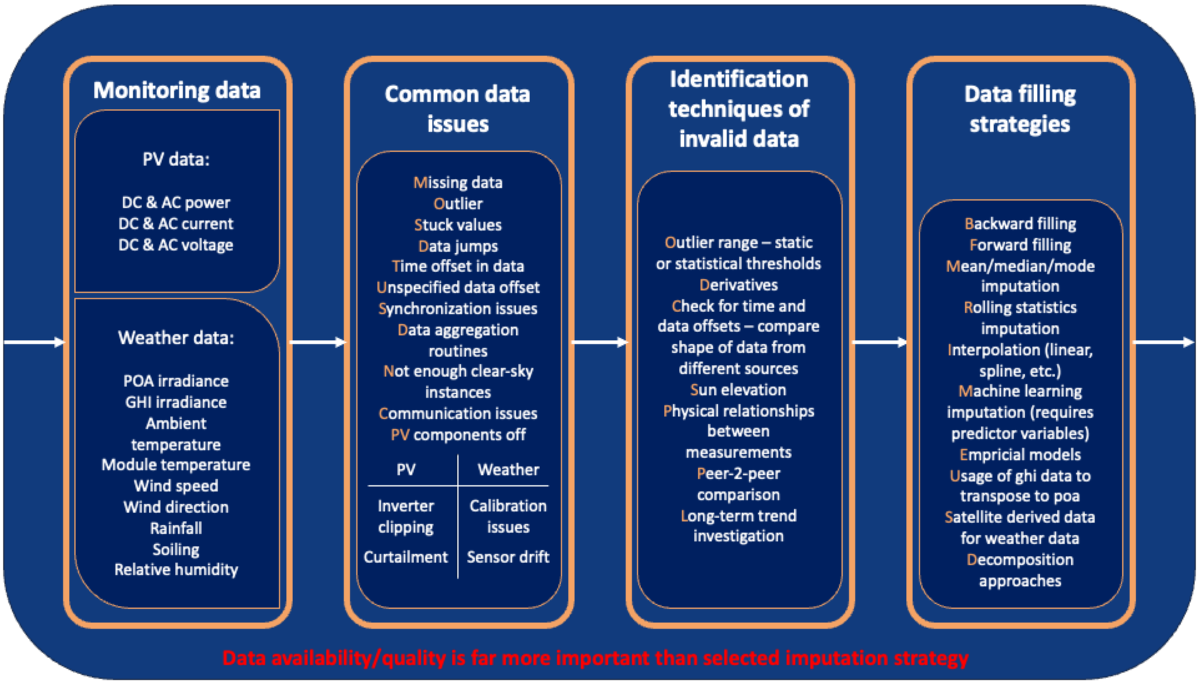


Figure 5: Data quality issues and tools.

3.3.1 Impact of data quality

Data quality is a driving factor for the reliability of any KPI calculation. As discussed above, data quality is a property inherent to raw input data, but, to a certain degree, can also be improved using filtering and cleaning approaches. Usually, KPIs represent an aggregated view on specific issues PV systems can be exposed to. They give indications of how a system is performing, either from a global performance perspective, for example through the energy performance index, or regarding a specific issue, such as the soiling rate. KPIs can be utilized only if the underlying data are reliable, so that the conclusions drawn from certain KPIs can be trusted. This is extremely important in operating PV systems, as KPIs are the primary driver to evaluate PV system performance, thereby detecting underperforming components. Based on these results, data-driven and informed decisions can be made to optimize PV system performance on one hand, but also to assign liquidated damages to the responsible parties. A data-driven O&M approach will improve the selection and timing of maintenance activities. Early-state, or even predictive, detection of PV system issues allows for a proactive approach to resolve issues before they have detrimental impacts.

Various research initiatives investigated the effects of data quality on PLR, including the works of Jordan et al. [43], Romero-Frances et al. [44], and Theristis et al. [45]. As described in chapter 2.1.5, this KPI encapsulates both reversible and non-reversible losses that can occur in PV installations. Nonetheless, the true PLR value is unknown when only processing operational data. The only way to derive the exact value would be to monitor each loss type individually. To assess the precision and uncertainty involved in distinct PR estimation techniques, Theristis et al. [45] developed a framework utilizing approximately 200 million synthetic datasets of known data quality and PLR behavior across the contiguous USA. Regarding data quality, the study focused on the effects of:

a) Application of erroneous power temperature coefficients (γ):

The common assumption that any γ would provide accurate PLR estimations was found to be invalid. The study (see Figure 6) showed that a substantial bias could be introduced when applying an erroneous γ, especially in shorter time series. Nevertheless, it is worth noting that employing a standard γ still enhances PLR precision and reduces uncertainty compared to no temperature coefficient at all.

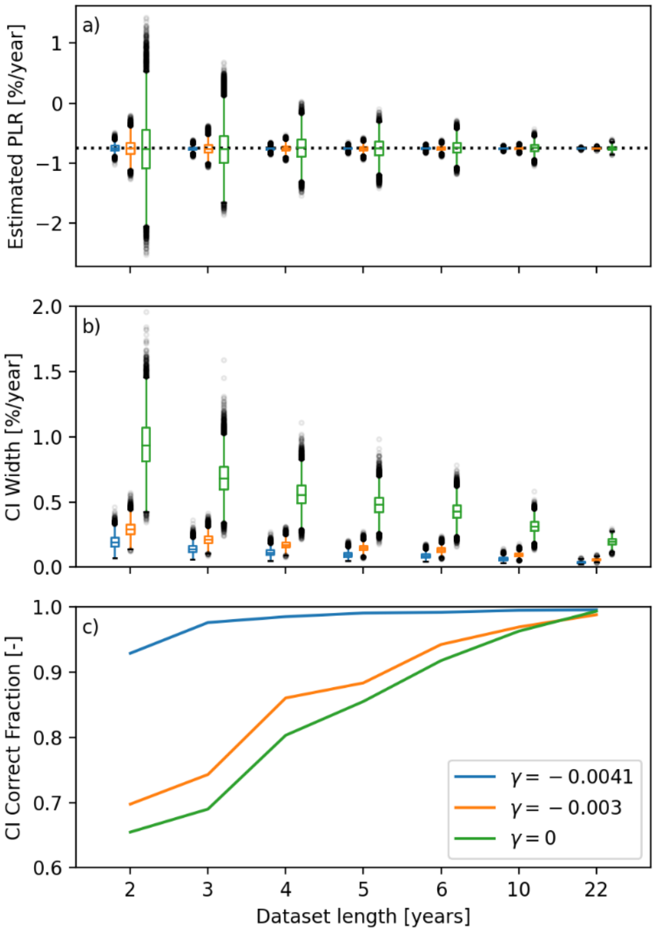


Figure 6: Impact of temperature correction on a) estimated PLR, b) width of confidence intervals & c) the fraction of correct confidence intervals as a function of dataset length in years. Three scenarios are investigated: 1) no temperature correction (γ = 0%/°C), 2) temperature correction with incorrect temperature coefficient (γ = -0.30%/°C), 3) temperature correction with correct temperature coefficient (γ = -0.41%/°C). The true PLR is linear at -0.75%/year (dotted horizontal line in a)). Figure obtained from Theristis et al. [45].

b) Erroneous data:

An increase in the prevalence of erroneous data, attributed to extended periods of system downtime, decreases accuracy of PLR assessments while increasing uncertainty (in relation to confidence intervals [CI]). While data sets without outages attained narrow CIs of 0.2%/year within a five-year span, those missing 30% of data needed a decade to reach similar CIs.

Moreover, the correlation between dataset length and the presence of errors is not linear. To quantify this effect, a new metric called "Effective dataset length", based on the CI width, was proposed to account for the influence of faulty data.

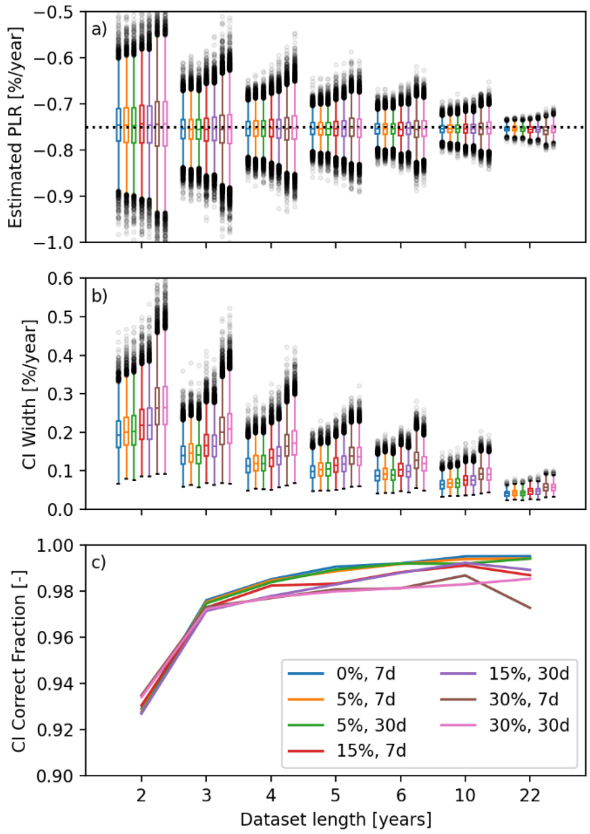


Figure 7: Impact of erroneous data on the a) estimated PLR, b) width of confidence intervals & c) the fraction of correct confidence intervals as a function of dataset length in years. The time series are temperature corrected, aggregated daily and their true PLR is linear at -0.75%/year (dotted horizontal line in a)). Figure obtained from Theristis et al. [45].

c) Irradiance sensor drift:

This phenomenon will introduce a bias when evaluating a power plant's PLR. Sensor drifting will have a compounding effect on the calculated PLR, meaning that it causes a positive bias in an almost linearly dependent manner. Furthermore, the study found the RdTools clear-sky normalization method [41] to be unreliable, especially at sites prone to overcast conditions. Hence, for time series impacted by sensor drift, the utilization of satellite data with clear-sky filtering could be a viable alternative, provided that the inherent uncertainties of such data are considered.

In conclusion, the investigation demonstrated that the impact of data quality on PLR evaluation is significantly dependent on climatic variables. For instance, the effect on PLR accuracy and uncertainty from a dataset with a specific percentage of missing data at a clear-sky location would differ from that of a dataset with an identical amount of missing data in a region with more dynamic weather conditions (example in Figure 8).

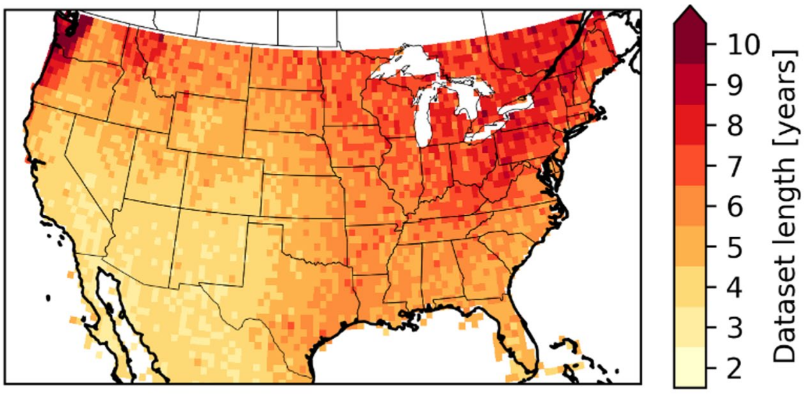


Figure 8: Example of PLR uncertainty depending on climatic variability and dataset length. The map shows the minimum number of years of data to achieve uncertainties within ±0.05%/year. When assessing PV systems across diverse climates, attention is required as universal assumptions are not applicable. Figure obtained from Theristis et al. [45].

3.3.2 Data imputation

As mentioned before, data are aggregated to calculate KPIs. The aggregation step will depend on the KPI itself and the intended purpose and can usually range from one hour to annual, with daily being the most common. Data cleaning and filtering is applied to the input data to remove non-relevant or false data points, which could otherwise impact the results. In case of missing data or removal of large parts of the data due to low data quality there will be a trade-off between accepting the non-existence of a certain amount of data or the application of data imputation techniques to recover that data. If the missing/removed data is only a small percentage of the overall data, and if these data holes are well spread across the dataset, it is likely that the data and information loss will not have a big impact on the final results. Once a certain threshold of missing data is crossed, and that threshold will also depend on the distribution across the dataset, data imputation could be a last resort to fill the missing data. In the PV sector, imputation is mostly applied to weather signals such as irradiance or temperature, as these are used to establish the intended relationships within KPIs, or power signals coming from the PV array. In a recent study by Deville at al. [46], it has been shown that input data availability and quality is far more important than the selected imputation model itself. If a model has generic input data, like spec sheet or a generic PAN file, then even the most complex model cannot provide adequate accuracy. In the following, different models depending on the target variable are discussed.

If only small portions and single datapoints are to be imputed, simple interpolation approaches such as linear, polynomial or cubic-spline interpolation or moving averages can be applied without significantly affecting the data quality. For larger parts of the datasets, regression machine learning techniques such as k-nearest neighbors, decision trees, random forests [47] or gradient boosting regressors [48] might be a valuable approach. The downside is that they require more expertise and reliable training data in the form of other dependent variables. An important part of building such an imputation model is to have thoroughly cleaned training data and to select useful training inputs, which highly correlate with the target signal. Apart from these data science driven approaches, signal-specific imputation approaches can be used.

For **power data**, empirical models may also be a viable alternative. Here, the PVWatts model [49, 50], the Sandia PV array performance model (SAPM) [51, 52] and the three parameter model [53] are commonly used candidates. Lindig et al. [54, 55] performed a comparative study to find optimal imputation approaches depending on the amount and type of missing data, the predictor availability and the ratio between training and test data. Here, the empirical models, different machine learning approaches and several univariate methods have been tested. For univariate methods, no additional signal is available as training data, and only power data before and after the outage instance are used to train the models and to perform the imputation. For smaller data gaps, empirical power models and machine learning imputation models perform well, and machine learning models seem to be preferable for bigger data gaps. The SAPM model, multivariate regression and ExtraTree regression [56] showed the highest accuracy.

Irradiance correlates strongly with PV power and is thereby the most valuable weather input parameter. In practice, irradiation data from different sources are prepared, rated, and selected to ensure the usage of the best possible data source for any operation. From the field, measured POA irradiance or decomposed and transposed GHI signals are generally used. A decomposed GHI signal is split into its direct and diffuse components which are the input for the transposition (together with ground albedo estimates), which presents the calculation of the incident irradiance on a tilted plane**.** Different models exist, with the Perez model being the most popular followed by Hay Davies, and their accuracy often depends on the geographical location [57]. If no, or only unreliable, weather data are available, data from nearby weather stations can be considered as distance-weighted averages, or single data points if the station is in very close proximity. The accepted distance will depend on the terrain topology and complexity. Another important resort are satellite derived irradiance data. Nowadays, there are multiple providers delivering data in different spatial and temporal resolutions. A way to improve the quality of satellite-derived data is to apply site-adaptation techniques. Site-adaptation refers to the process of optimizing and adjusting input satellite irradiance data using high-quality short-term ground measurements to reduce the bias and overall error in the data. Important contributions for benchmarking modelled solar irradiance data as well as regarding site-adaptation techniques have been and are published by IEA PVPS Task 16 [58, 59].

On-site ground measurements should be prioritized over the mentioned alternatives. This is expressed through the irradiance data categorization according to Ascencio-Vásquez et al. [60]:

1. *High accuracy*: POA irradiance is measured on-site.
2. *Medium accuracy*: GHI is measured on-site and POA is estimated using decomposition and transposition approaches.
3. *Low accuracy*: POA is estimated using decomposition and transposition approaches from extracted GHI, which is taken from one of the following sources: interpolated (weighted regression) using peered data of different weather stations in relatively close proximity to the test side, satellite or reanalysis-based datasets (atmospheric data generated by combining historical observations with numerical weather prediction models), clear-sky modeled datasets.

Louwen et al. [61, 38] compared different data imputation strategies for POA time series using measured GHI, DHI and relative humidity datasets. The dataset was taken from the publicly available DKA Solar Centre which operates PV plants in Alice Springs/Australia [62]. The aim was to impute four years of missing POA data for a ten-year dataset. Different traditional irradiance transposition models, available in the Python package pvlib [40], were compared with several machine learning-based models. The machine learning models consisted of random forest [47], extra trees [56], gradient boosting [48], and histogram-based gradient boosting [63], utilizing the scikit-learn library [64]. Solar position parameters (solar zenith, solar azimuth, and solar elevation) were incorporated into the input dataset, and measurements with solar elevation equal to or less than zero were excluded. Upon assessing all considered methods for estimating the plane-of-array irradiance, it was evident that the machine learning-based models outperformed the transposition-based models. The histogram-based gradient boosting regressor demonstrated the highest accuracy. However, this approach should be considered primarily for larger data gaps, as simpler methods, such as linear interpolation or transposition-based models, are sufficient for short data gaps.

Module temperature data are a key parameter for system performance analysis and (temperature) corrections, e.g. the temperature-corrected PR for performance loss rates [65], and weather-corrected PR [66] for PV system contractual evaluations during the EPC system acceptance testing period. Data imputation for missing data or as quality control is often based on explicit equation-based models, where the Ross [67], SAPM [68], Faiman [69], Herteleer-WM1 and Herteleer-WM2 models [70] can be used. At “lower” timescales of 15 min to 1 hour, equation-based thermal models can be used with few issues and reasonably high accuracy (RMSE 2-3°C for 1 year of data). However, at shorter timescales (1 s to 5 min), the thermal lag of PV modules versus irradiance and wind speeds increases the thermal model error of standard models as the timestep shortens, unless the model is made dynamic. The methodology of Prilliman et al. [71], implemented in pvlib-python [40], requires knowledge of module parameters (unit mass: kg/m²), and achieves better results with calibrated coefficients which are found through Finite Element Analysis (FEA). The approach of Herteleer et al. [70] can be either data-driven (finding coefficients from measured data) or using the average values to model the dynamic thermal behavior of PV modules. This method uses the thermal time constant to find the smoothing coefficient in python pandas and uses this to calculate the exponential weighted mean value of the irradiance and wind speed signals, which drive the module temperature. Both the methods of Prilliman et al. and of Herteleer et al. can be applied to different “root” or steady-state thermal models, to calculate the temperature of PV modules in dynamic (1 s to 5 min) time resolution. Machine learning approaches can also be employed to model module temperatures [72], even in cases where a system lacks module measurements [73]. Even though these approaches are promising, it is still recommended to have multiple temperature sensors installed and routinely checked or calibrated, to have a ground truth for models, while the models can then be used for quality control on the sensors and data imputation when needed.

3.3.3 A unified Data Quality Routine

Inconsistency in datasets might be caused when data quality controls are applied by different analysts. This may occur because analysts may implement these controls in different sequences. Furthermore, when standards like IEC 61724 or other PV data quality reports do not quantitatively define a data processing step, interpretation is left to the analyst. Consequently, the absence of a standardized approach to data quality control in PV performance and reliability analyses presents significant challenges such as inconsistent data interpretation or data integrity issues.

To address this issue, Livera et al. published a comprehensive and quantitative PV data processing methodology. The methodology, which relies on quantifiable criteria from IEC 61724 and other PV data quality reports, reduces the ambiguity and qualitative nature of current practices. It consists of a sequentially structured pipeline of Data Quality Rules (DQRs) involving initial statistics, consistency examination, filtering, invalid data detection and data rate quantification, treatment of invalid data, and aggregation at various granularities. The overall objective is to reduce uncertainty when datasets are excessively filtered or aggregated differently, and how data should be inferred.

To develop the methodology, artificially invalid datasets were created by introducing invalid data at various rates and sequences and DQRs were applied to detect and treat invalid datasets using different filtering and inference methods. Each step of the parametric analysis was then compared against reference values to optimize the DQR methodology (Figure 9). The analysis showed that when invalid data exceed 10%, data inference techniques should be applied to minimize uncertainty.

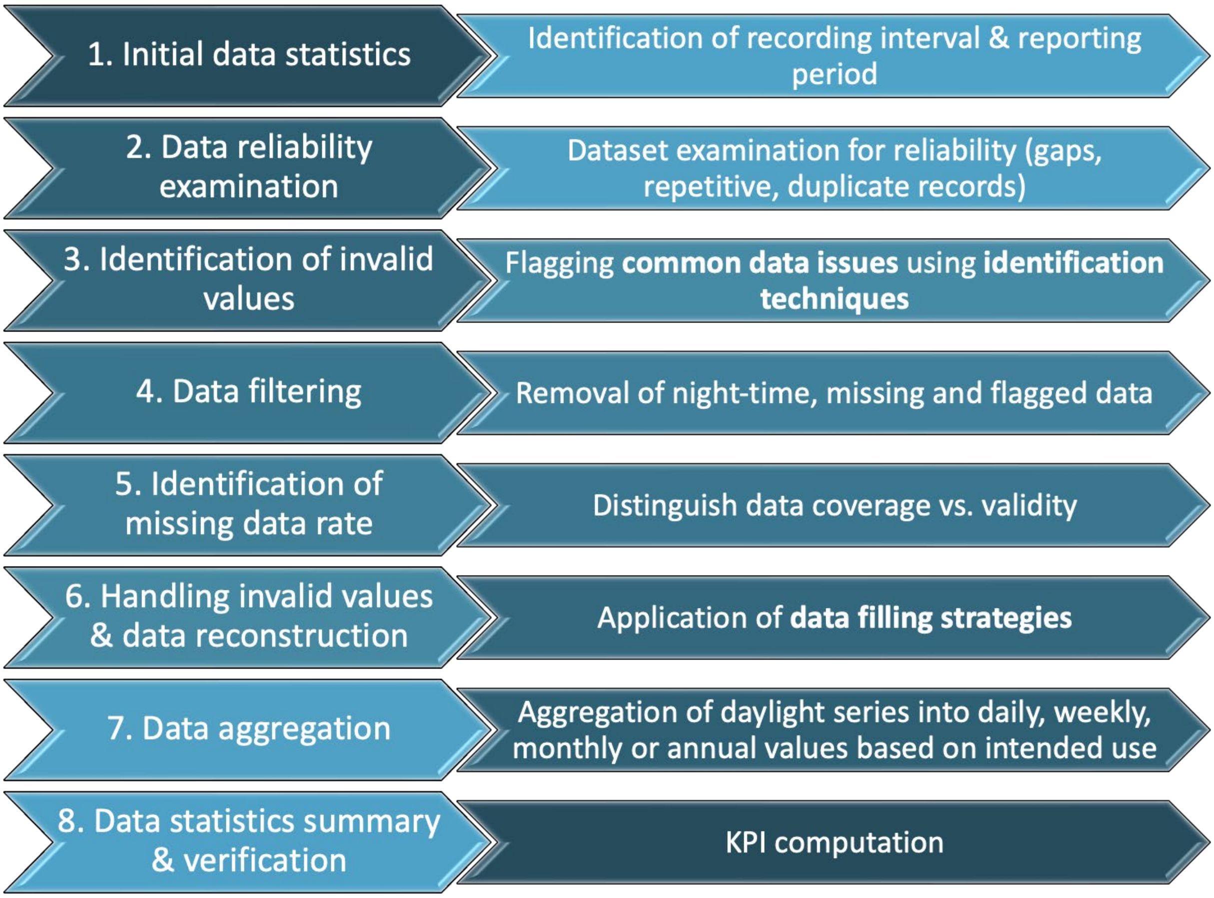


Figure 9: Flowchart of a unified data quality approach for PV performance and reliability analytics. Figure adapted from Livera et al. [74].

3.4 Data aggregation

When monitoring a PV plant, the processing of data from collection to actionable insights involves a structured process that starts with sampling and culminates in reporting.

At the most granular level, we have the sampling process, where individual data points are acquired from sensors or measuring devices. These samples are taken at regular intervals, known as the sampling interval, which can vary from seconds to minutes depending on the parameter being measured and the desired level of detail.

Once samples are collected, they are compiled into records. A record is an aggregation of samples that is stored at the end of a predefined time period, the recording interval. This interval is carefully chosen to be an integer multiple of the sampling interval, ensuring that a consistent number of samples is used to create each record and that these intervals align neatly within an hour.

The value of each record is not just a random or isolated sample; rather, it is a calculated representation—such as an average, maximum, minimum, or sum—of all the samples taken during the recording interval. This method provides a more accurate and meaningful picture of the data collected. Records can also include supplementary information, such as additional statistics, counts of missing data points, error codes, or notes on any transients, enriching the dataset with valuable context.

Finally, the reporting stage is where data is synthesized over multiple recording intervals into reports. These reports provide actionable insights by summarizing the aggregated data, highlighting trends, and identifying any anomalies or issues that need attention.

Example: Considering a PV plant where the sampling interval is set to 1 minute, and the recording interval is set to 30 minutes. Over the course of an hour, 60 samples are collected for each parameter (e.g., power output, temperature, irradiance). These samples are then aggregated into two 30-minute records. Each record might include the average power output, the maximum and minimum temperatures, and the sum of irradiance over the 30-minute period.

At the end of the day, these 30-minute records are further aggregated into daily reports. The daily report might include the total energy produced, the average daily temperature, and any periods of significant deviation from expected performance. For example, the report might highlight that the total energy produced was 12 MWh, the average temperature was 22°C, and there was a significant drop in power output between 2 PM and 3 PM due to a transient shading event.

These reports are then used by plant operators to make informed decisions, such as scheduling maintenance, optimizing performance, or investigating anomalies. By providing a clear and comprehensive overview of the plant's performance, the reporting stage ensures that the data collected is transformed into valuable insights that drive operational efficiency and reliability.

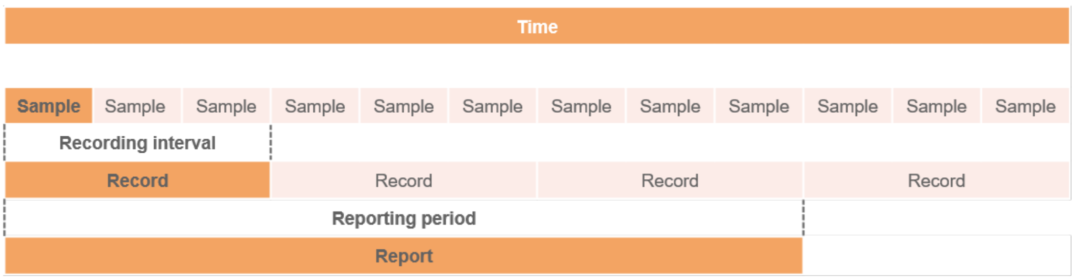


Figure 10: Temporal behavior between samples, records, and reports; adapted from [2].

4 EXPLOITATION AND ACTIONABLE INSIGHTS

As described in the previous chapters, the solar industry relies on different standardized and non-standardized KPIs during the lifecycle of PV assets. KPIs are an essential part of the data pipeline, where input data (in any form but mostly time series from SCADA systems) are processed to provide numerical outputs that, aggregated, would take the common name of a KPI (e.g., monthly performance ratio). What is the value of calculating a KPI if it does not translate into tangible outcomes, improved performance reporting, or meaningful enhancements? This section focuses on using KPIs during different phases along the value chain of a PV asset.

As indicated in Figure 11, this section focuses on the workflow steps of analysis, interpretation, reporting and actions.

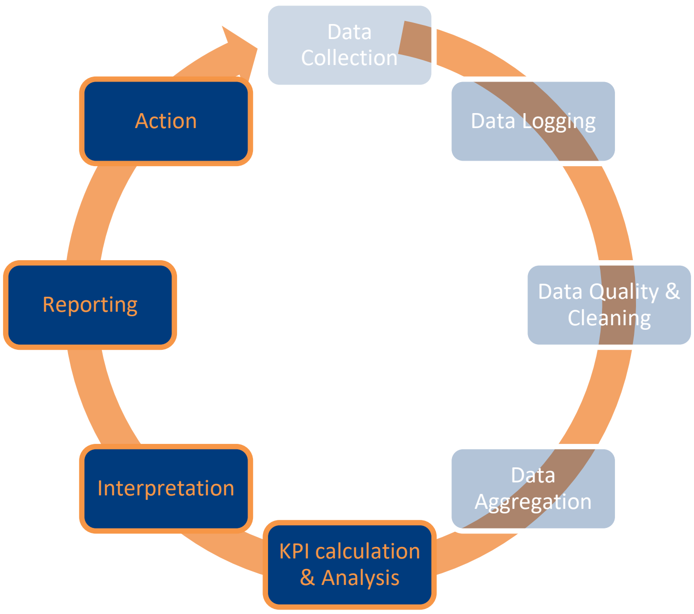


Figure 11: Stages of data processing cycle that can provide outcomes in a PV asset.

From Data Collection to Data Aggregation, as the names indicate, the data is collected, processed, cleansed and prepared for analysis. At the data analysis stage, the relevant KPIs will be computed, and a preliminary assessment will be made to make sure that any following stage has the correct resources (i.e., valid data) to make decisions. Having the correct data will serve the Interpretation phase with information to make decisions; those decisions can be made in the form of Reports, as industry conventions to share outcomes with other stakeholders, or as Actions mainly on the field where the asset performance, reliability, safety or economics can be hopefully optimized.

In the following sections, examples of typical data sources and tools for Data Analysis are given, as well as real-world case studies where KPIs result in tangible outcomes in different stages of the PV value chain.

4.1 Most common sources of data

4.1.1 Time series from monitoring systems

Solar farms need to be constantly monitored to support ongoing technical activities or sometimes energy trading operations, so the need for such data is primordial to secure sufficient knowledge for decision-makers. In order to provide reliable data and KPIs, the solutions need to include proper on-site hardware (i.e., sensors, data loggers, servers, gateway, etc.), and the software needs to follow standard taxonomies to ensure interoperability [30]. After well defining data structures and connectivity, algorithms can process, clean, and aggregate the data. Figure 12 shows a screenshot of a commercial monitoring software application aggregating PV production and irradiation per hour on the left, and it shows detailed data on the right side.

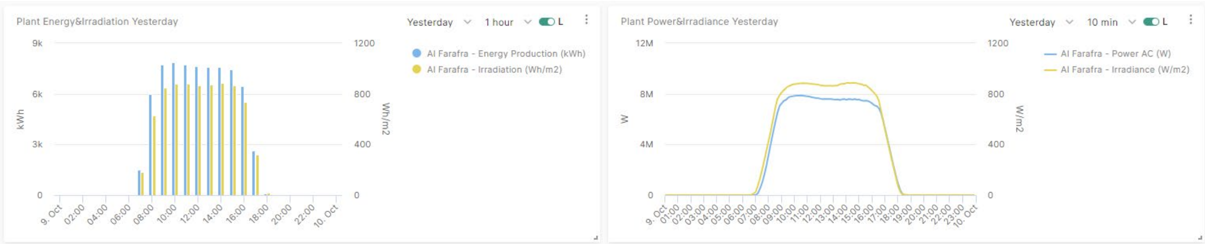


Figure 12: Screenshot of typical dashboards in real-time monitoring software tools [75].

4.1.2 Aggregated KPIs from time series

Simple aggregations, such as power to energy values or irradiance to irradiation in specific periods, are good enough for certain PV operations; however, as described in previous chapters, many other KPIs are needed to get a better picture of the health state and performance of a PV system, and they require a certain level of calculations and process. When analyzing large-scale PV systems, any form of manual data handling process becomes too expensive due to the time-consuming process, so automatic tools can support such tasks if they are appropriately set up.

In Figure 13, a screenshot of a typical dashboard in analytics software tools is presented. Here, three KPIs are displayed over time for a single PV plant. The following technical KPIs are computed and shown:

- $\mathrm{PR}_{\mathrm{Tcorr}}$: temperature-corrected performance ratio
- TBA: time-based availability
- EPI: energy performance index

To achieve a high level of accuracy, each of the previous data processing stages, from data collection to data aggregation, had to be carried out thoroughly.

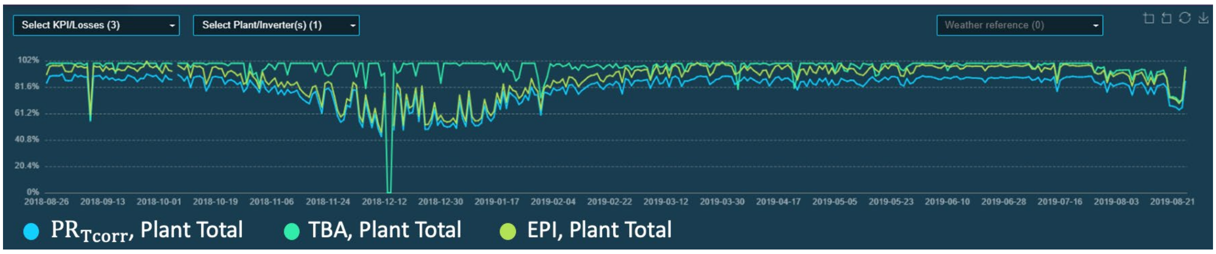


Figure 13: Screenshot of typical dashboards in analytics software tools [37].

Figure 14 displays another representation of KPIs computed in analytics software tools, where the temperature-corrected PR, PR_{Tcorr}, is obtained daily for each inverter in a PV system. This type of data calculation and visualization is intended to support data-driven decision-making tasks in operations or technical asset management. This heatmap representation allows for a fast and easy inverter-to-inverter performance comparison to detect underperforming components.

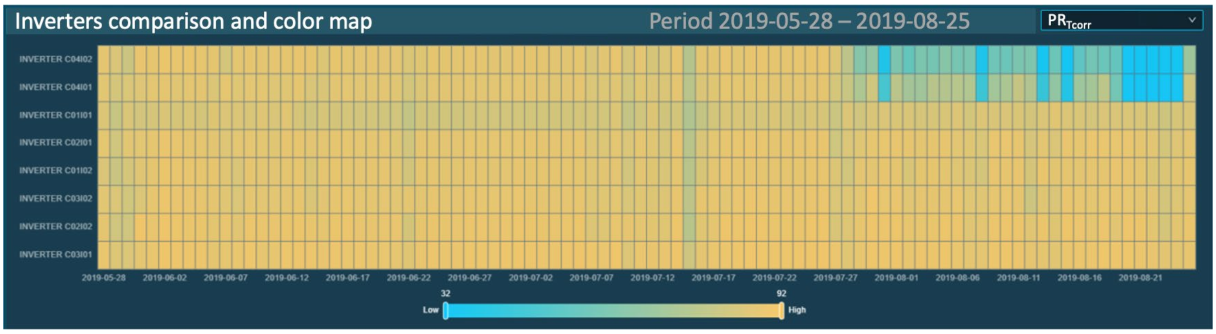


Figure 14: Screenshot of a heatmap representation with PR_{Tcorr} calculations for many inverters in analytics software tools [37].

4.1.3 Geospatial weather data from reanalysis- and satellite-based methods

Reanalysis-based and satellite-based information can support KPI calculations for geospatial analysis by providing historical weather data like ambient temperature, global horizontal irradiance, and wind speed for PV modelling in different geographical regions. Such data can also serve as a backup source when on-site sensors present issues like miscommunication, miscalibration, or simply wrong installation.

Figure 15: Referential illustration of a mesh over the Earth representing where reanalysis-based models run calculations [76].

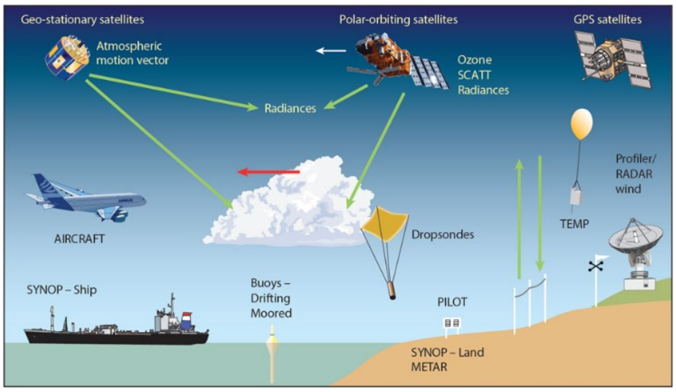


Figure 16: Practical illustration indicating various types of data input sources for reanalysis-based models enabling calculations with global coverage [77].

4.1.4 Geospatial post-processed PV data

Physics- or empirical-based models on top of geospatial weather data can be used to simulate PV-related information, such as expected energy for predefined PV systems and estimated soiling rates at the regional or country level, among others. This type of information also represents a source of data for KPIs and further actions from this. For example, in Figure 17, the usage of geospatial data resulted in estimations of soiling rates and further soiling risks at global and country levels. A potential use-case of such information can be region- or climate-tailored PV system designs or adapted O&M approaches and routines.

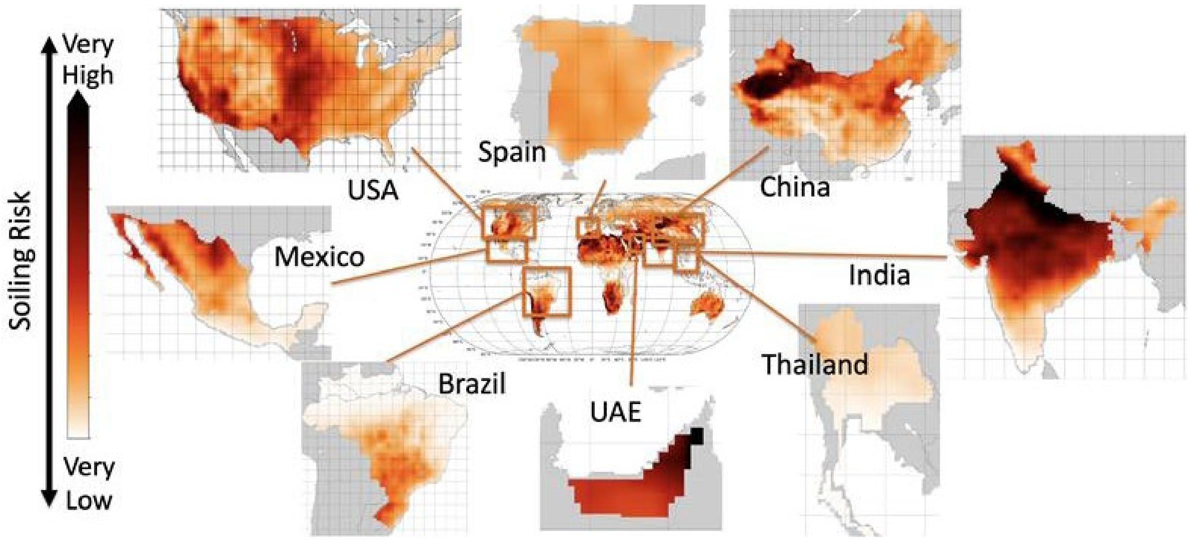


Figure 17: Geospatial post-processed PV data indicating Soiling Risk from AOD and physics-based soiling models [78].

4.2 Other forms of data

More sources of data are being used in PV operations with specific goals such as failure detection, warranty claims, yield optimization or reliability analysis. All those data streams are part of the digitalization era that PV plants are currently experiencing.

4.2.1 **Aerial Imaging from drones**

Essential KPIs can be retrieved by processing thermographic images using drones (see Figure 18). The thermal signatures of PV cells and PV modules can be indicators of hotspots, potential-induced degradation, diode, or interconnection failures, etc. Such indicators can be represented by additional KPIs, such as the number of PV modules impacted or the ratio of PV power plant damage.

Figure 18: A stylized image showing aerial IR imaging conducted by a drone [79].

4.2.2 Static data from IV tracers

Typical O&M procedures to analyze PV module or string health include the measurement of current-voltage (I-V) curves in the field. Besides manual measurements, some inverter manufacturers have improved inverter capabilities by including the option to trigger measurements regularly and remotely.

I-V curves provide a set of pair values that characterize the PV behavior. Thereby, a trained eye can evaluate PV performance based on the shape. As shown in Figure 19, the evolution of series resistance (R_S) or shunt resistance (R_{SH}) can be understood, or certain defects such as mismatches or shading detected. After data processing and analysis, KPIs, such as actual STC power, can support triggering different actions like PV module layout reorganization or PV module warranty claims in exceptional circumstances.

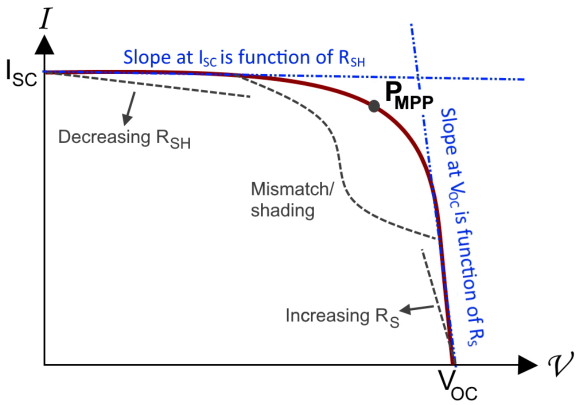


Figure 19: Illustration of the I-V curve from a PV string [80].

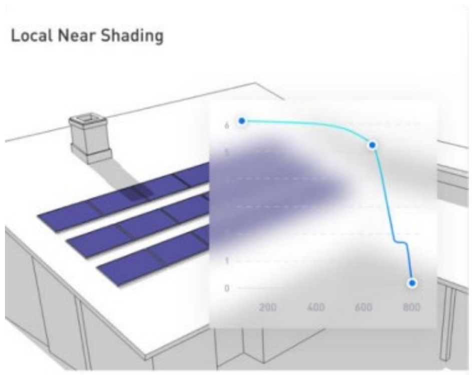


Figure 20: Illustration of I-V curve patterns under local near shading - common underperformance scenario [79].

4.3 Case studies

Depending on the source of data and the KPIs, numerous applications and decision-making processes can be generated. In this section, various case studies are presented, closing the loop from data collection to reporting or actions in the field. To summarize each case study, a table is included, going from input data to action.

4.3.1 From soiling KPI to cleaning planning

Concept	field data to cleaning annual planning
Input data	monitoring/sensor data
Calculation	physics-based or statistical-based
Data Analysis	soiling ratio, KPI
Interpretation/ Reporting/Action	optimized cleaning planning

Dust accumulation on PV modules can have a high impact on yield in specific climatic zones. Development of O&M management including plans for cost-effective mitigation require good estimates of soiling rates and ratios. Diverse methods and solutions have been proposed to offer proper O&M management in terms of cleaning procedures and schedules, and some of the advanced methods use field data for this task.

Soiling sensors are frequently used devices in dusty or polluted zones. These devices offer local field estimates of soiling rates and soiling levels; some examples include DustIQ [81], Fracsun [82], and Atonometrics [83]. Such sensors provide raw time series data that require appropriate processing and filtering to offer high accuracy. The resulting data analysis involves calculating average soiling rates or soiling ratios over time. Those KPIs can then be used as input to data-driven cleaning planning optimization, enabling O&M teams to better allocate resources ahead of time.

Soiling ratios can also be calculated from production data time series available from the inverters [84]. Below, data regarding soiling and cleaning observed in a utility-scale PV power plant with a nominal capacity exceeding 150 MW_p in a tropical wet and dry climate with seasonal rainfall is shown. Production data is available from more than 100 inverters installed in the PV power plant. Moreover, 12 Atonometrics RDE300-series soiling stations are installed across the site. Inverter data enables calculations of both the daily, median performance index ($PI_{soiling}$) and Soiling Ratios (SR). In this case, the calculation is performed using the Combined Degradation and Soiling (CODS) algorithm [85]. The calculated values based on inverter data are compared to soiling station measurements on site and are in good agreement. Also included in the graph are measured precipitation events and logged cleaning events. The shaded grey data points of the soiling station measurements indicate measurements where output from the soiling station is deemed faulty.

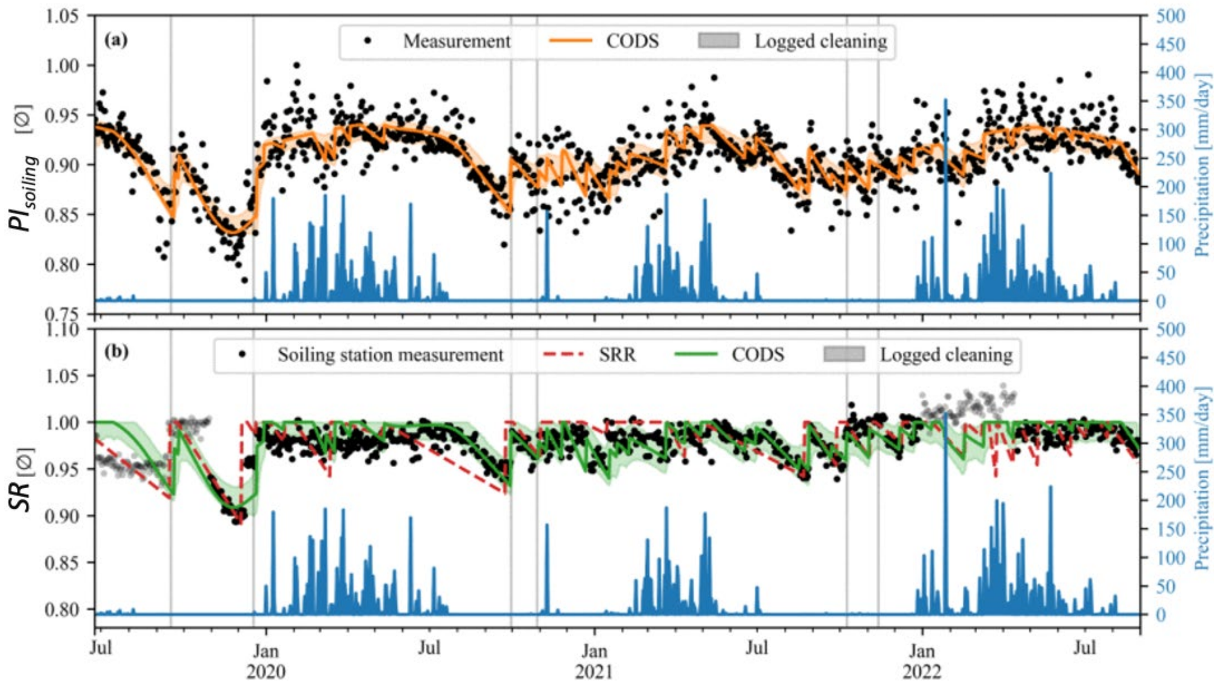


Figure 21. (a) Normalized $PI_{soiling}$ with CODS model fit for an inverter, and (b) CODS and stochastic rate and recovery (SRR) estimates of the soiling ratio SR obtained from the performance time series data compared with independent measurements from the closest soiling station to the selected inverter [85].

4.3.2 From KPIs to corrective actions and O&M planning

Concept	time series to economic KPIs and corrective actions
Input data	monitoring data
Calculation	loss detection and breakdown
Data Analysis	from energy losses to monetary losses
Interpretation/ Reporting/Action	prioritized corrective actions based on cost-benefit analysis for O&M planning

Nowadays, the primary devices in the PV plants are being monitored, and data are collected through data loggers or similar, ending up on a data lake or SCADA solution. The vast amount of data gathered by inverters, combiner boxes, weather stations, trackers, and more, together with advanced monitoring and analytics tools, are helping to digitalize the daily operations of many stakeholders.

Considering all the operational monitoring data of PV plants, diverse algorithms can be used to detect losses and disentangle such losses into several categories, such as curtailment, clipping, DC issues, or tracker malfunction. The estimated energy losses can be multiplied by the energy prices or PPA values to obtain the monetary losses per category.

The data analysis of the energy and monetary losses can be further extended to the automatic triggering of corrective actions whenever an event happens, and its consequences are worse than the actual O&M cost to fix the issue. An approach to quantify this breakpoint and to relate monetary losses due to lost energy with fixing costs is the cost priority number (CPN) methodology [55].

4.3.3 **From geospatial data to PV design**

Concept	satellite data to single values
Input data	satellite and reanalysis information
Calculation	physics-based PV model to compute irradiance POA and energy production for different scenarios
Data Analysis	from energy production to payback and ROI
Interpretation/ Reporting/Action	engineers can rapidly get preliminary assumptions for basic and detailed design

Another important type of input data is all the information given and resulting from satellite images and geospatial data. Satellite- and reanalysis-based datasets are becoming increasingly essential in small and large PV power plants, mainly triggered by significant consistency over time and reliable data coverage in contrast with weather stations and potential communication issues. Such information can help to geospatially map large regions worldwide and give quick estimations of energy production, as well as even more potential payback and return-of-investment (ROI) of photovoltaic projects.

A good example of this use case is a tool offered by the Joint Research Centre (JRC), called PVGIS [86], where users can freely access and select any location on the map and quickly get preliminary assumptions for the basic design of PV plants. KPIs such as energy yield or annual effective irradiation can be enough information to kick-off projects at an early stage.

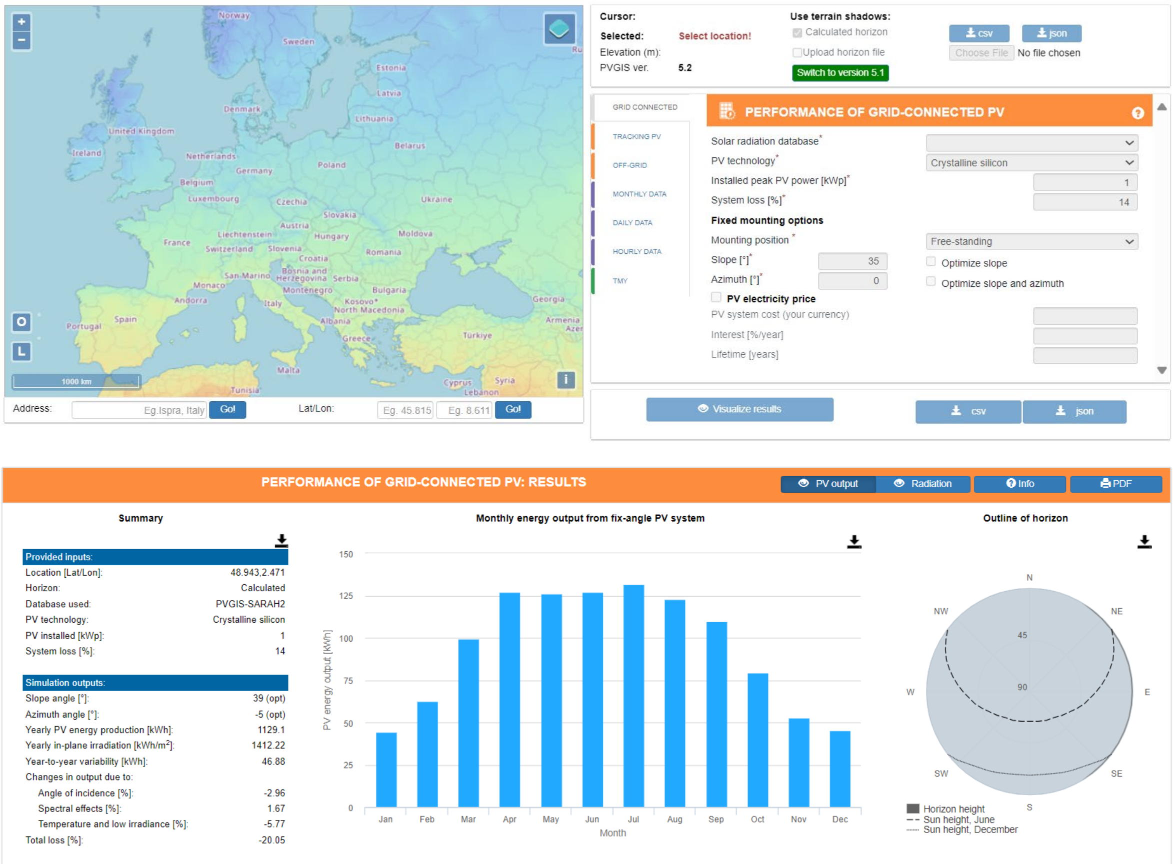


Figure 22: Dashboards available on the application PVGIS from the Joint Research Centre of the European Commission [86].

4.3.4 From geospatial and physics-models to risk maps

Concept	satellite data to technical KPIs
Input data	satellite data and reanalysis data
Calculation	physics-based models
Data Analysis	degradation rates
Interpretation/ Reporting/Action	geographical zones expect less production due to environmental factors and support climate-specific design

Following the usage of satellite- and reanalysis-based data, geospatial modelling, including physics-based models, can be an interesting approach to produce maps that can help cluster geographical regions into specific PV KPIs. As an example of this, PV degradation maps can be constructed to indicate risky areas for long-term PV operations based on climatic risks. Such outcome can be considered for interpretation for geographical zones with less expected production due to environmental factors and support climate-specific PV module designs.

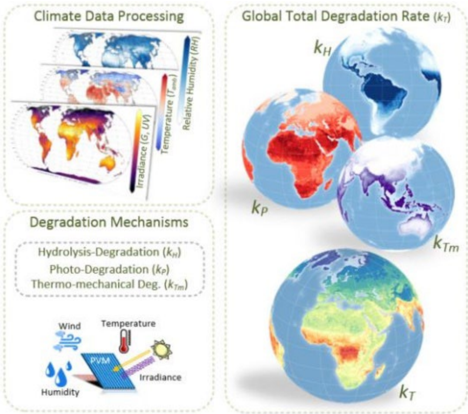


Figure 23. Graphical abstract with main calculation steps to construct degradation rates global maps [87].

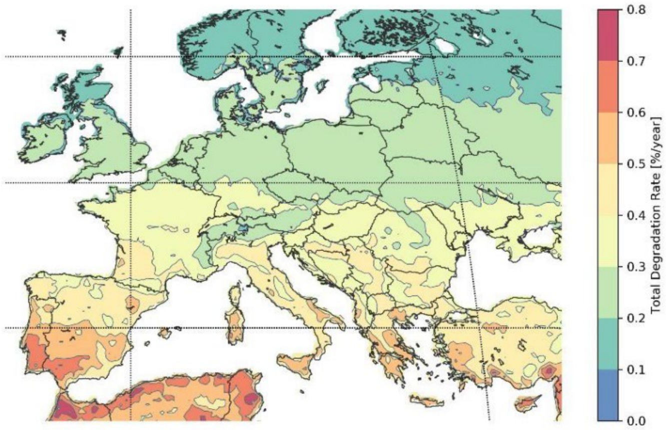


Figure 24. Total degradation rates [%/year] modelled using reanalysis data and physical models for the European Continent [87].

4.3.5 Detection of PV module faults and PV module replacement actions

Concept	combined analysis of time series data and IR imaging to detect faulty PV modules for subsequent replacement
Input data	monitoring data, UAV-based IR imaging
Calculation	loss detection
Data Analysis	from energy losses to actionable information
Interpretation/ Reporting/Action	prioritized corrective actions when combined with cost-benefit analysis for O&M planning

Although the impact of single, faulty PV modules in a utility-scale PV power plant are usually miniscule, the accumulated impact of all PV module faults can become substantial over time.

Identification of faulty PV modules combined with an impact analysis and local O&M costs can enable cost-effective replacement strategies.

Time series analysis performed on string level can allow for quantification of losses associated with faulty PV modules. However, the granularity is limited, as several tens of PV modules are connected into each individual inverter channel. Unmanned aerial vehicle (UAV)-based infrared (IR) imaging, on the other side, can localize hot spots in individual PV modules. However, this method does not directly give access to the impact of the observed thermal signatures [88]. This complementarity makes the combination of these two methods quite powerful. Below, the result of combining string-level time series analysis with IR imaging for identifying faulty PV modules for subsequent replacement is shown. In this case, the focus is on activated bypass diodes, which can give clear signatures in the production time series. The data was obtained from central inverters installed in a 75 MWp PV power plant located in a cold, arid steppe climate with a mean of 160 strings in parallel per maximum power point tracker, each consisting of 24 PV modules. The shown IR imaging was obtained by a commercial third party after approximately 5 years of operation.

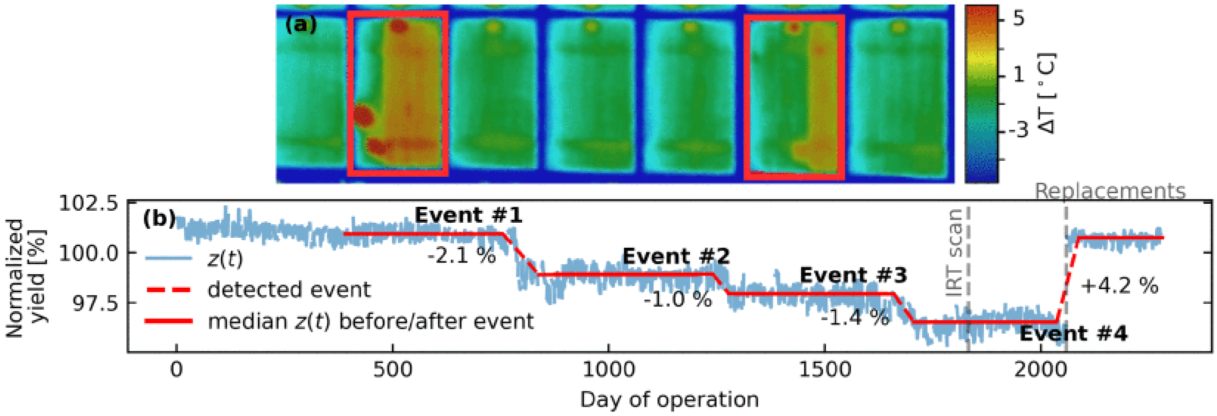


Figure 25. (a) IR images of PV modules in a string with a total of 3 detected bypass diode signatures, and (b) the calculated, normalized yield based on the time series data clearly showing the timing of each bypass diode activation, as well as their individual impact on normalized yield. Figure taken from [89] - Reprinted with permission from the IEEE JPV.

The case study presented in Figure 25 was extended beyond the time of the IR scan, after which a replacement campaign was performed, resulting in close to complete recovery of the normalized yield (Event #4).

5 CONCLUSIONS & OUTLOOK

This report focuses on the calculation and application of the main key performance indicators (KPIs) for operating PV systems, used for technical and contractual purposes.

Contractual agreements and associated KPIs vary by project size, market, locations. KPIs even vary by their definition and/or method of calculation, with attendant financial consequences. For example, in the hypothetical scenario where a KPI (e.g., performance ratio) calculation is based on a highly filtered raw dataset, a positive or negative bias might be introduced: in this case, an operation and maintenance (O&M) provider will falsely meet (or fail to meet) their agreement while the asset owner will falsely lose/earn revenues. Transparency and standardization of KPI definitions (where it does not exist) to eliminate bias and promote reproducibility is therefore of utmost importance.

By collecting expert knowledge from stakeholders and through a comprehensive literature review, a list of the most used KPIs and their usage within day-to-day operation was collected. Furthermore, each individual KPI is clearly defined, and their advantages and challenges are being discussed to have a critical view on their application and limitations. This concludes the first part of this report.

The second part is walking through the most important data processing applications. Here, existing standards and guidelines are the starting point. This part of the report is intended to guide the reader towards best practices in data handling and to help avoid common pitfalls when handling PV system data. Commonly used devices and weather sensors are discussed, going from data collection and logging to recommended cleaning and filtering techniques towards data aggregation. This part closes with a graphical representation of a unified data quality approach for PV performance and reliability analytics.

In the last section, KPI mapping possibilities beyond contractual agreements and minimum performance thresholds are discussed. Many exciting advanced evaluation possibilities arose in recent years and will be more and more common practice in order to provide the best possible insights into PV system data, so that PV system performance can be tracked better and, through pro-active data driven actions, kept at high and desired levels. Thereby, special attention is paid to the geo-spatial mapping of performance related KPIs. This type of mapping exercise is extremely valuable for early-stage PV system design. Based on local requirements, equipment, installation design and climatic load resistance, O&M approaches could be tailored, something which is today not fully utilized.

For future work, PV system performance data from large PV system analytics platforms are being collected to calculate technical system KPIs and interpret this data in the form of geo-spatial maps. This approach will serve to expand the knowledge from specific sites where data are available to locations where PV systems might not even be installed yet. This exercise will be very valuable for the design of new PV plants and tailoring PV system contracts utilizing higher certainty in performance estimates.

BIBLIOGRAPHY

[1] International Energy Agency, "Solar PV," September 2022. [Online]. Available: https://www.iea.org/reports/solar-pv.

[2] International Electrotechnical Commission, «IEC 61724-1:2021 Photovoltaic system performance, Part 1: Monitoring Standard,» Geneva, 2021.

[3] International Electrotechnical Commission, «IEC 61724-2:2016: Photovoltaic system performance - Part 2: Capacity evaluation method,» Geneva, 2016.

[4] International Electrotechnical Commission, «IEC 61724-3:2016: Photovoltaic system performance, Part 3: Energy evaluation method,» Geneva, 2016.

[5] U. Jahn, B. Herteleer, C. Tjendgdrawira, I. Tsanakas, M. Richter, G. Dickeson, A. Astigarraga, T. Tanahashi, F. Valencia, M. Green, A. Anderson, B. Stridh, A. Lagnuas et Y. Sangpongsanont, «Guidelines for Operation and Maintenance of Photovoltaic Power Plants in Different Climates, Report IEA-PVPS T13-25,» ISBN 978-3-907281-13-0, 2022.

[6] D. Moser, S. Lindig, M. Richter, J. Ascencio-Vásquez, I. Horvath, B. Müller, M. Green, J. Vedde, M. Herz, B. Herteleer, K.-A. Weiss et B. Stridh, «Uncertainty in Yield Assessments and PV LCOE, Report IEA-PVPS T13-18,» ISBN 978-3-907281-06-2, 2022.

[7] R. Guerrero-Lemus, R. Vega, T. Kim, A. Kimm et L. Shephard, «Bifacial solar photovoltaics – A technology review,» *Renewable and Sustainable Energy Reviews,* vol. 60, p. 1533–1549, 2016.

[8] SolarPowerEurope, «Operation & Maintenance Best Practice Guidelines Version 5.0,» SolarPowerEurope, 2021.

[9] J. G. Bessa, L. Micheli, F. Almonacid et E. F. Fernández, «Monitoring photovoltaic soiling:assessment, challenges, and perspectives of current and potential strategies,» *iScience,* vol. 24, p. 102165, March 2021.

[10] J. Gifford, M. Lynas, J. Deckx and P. Agrawal, "Unlocking the power of advanced analytics: maximizing value in multi-GW utility-scale solar portfolios," 4 October 2023. [Online]. Available: https://www.pv-magazine.com/wp-content/uploads/2023/09/All-presentations_final.pdf. [Accessed 22 February 2024].

[11] C. Schill, A. Andersson, C. Baldus-Jeursen, L. Burnham, L. Micheli, D. Parlevliet, E. Pilat, B. Stridh, E. Urrejola, E. Whitney et U. Jahn, «Soiling Losses - Impact on the Performance of Photovoltaic Power Plants, Report IEA-PVPS T13-21,» ISBN 978-3-907281-09-3, 2022.

[12] S. Lindig, M. Theristis and D. Moser, "Best practices for photovoltaic performance loss rate calculations," *Progress in Energy,* vol. 4, 2022.

[13] D. C. Jordan, B. Marion, C. Deline, T. Barnes et M. Bolinger, «PV field reliability status—Analysis of 100 000 solar systems,» *Prog Photovolt Res Appl.,* vol. 28, p. 739–754, 2020.

[14] R. French et al., «Assessment of Performance Loss Rate of PV Power Systems, Report IEA-PVPS T13-22,» ISBN 978-3-907281-10-9, 2021.

[15] American Society for Testing and Materials, «ASTM E2848-13: Standard Test Method for Reporting Photovoltaic Non-Concentrator System Performance,» ASTM International, West Conshohocken, PA, United States, 2018.

[16] B. Taylor, "Contents — pvcaptest v0.12.1-2-gefa5e22 documentation," 2023. [Online]. Available: https://pvcaptest.readthedocs.io/. [Accessed 22 February 2024].

[17] D. Moser, M. Del Buono, U. Jahn, M. Herz, M. Richter et K. De Brabandere, «Identification of technical risks in the photovoltaic value chain and quantification of the economic impact,» *Prog Photovolt Res Appl.,* vol. 25, pp. 592-604, 2017.

[18] NREL, "Utility-scale PV," 2023. [Online]. Available: https://atb.nrel.gov/electricity/2023/utility-scale_pv. [Accessed 1 March 2024].

[19] R. Musi, B. Grang, S. Sgouridis, R. Guedez, P. Armstrong, A. Slocum and N. Calvet, "Techno-economic analysis of concentrated solar power plants in terms of levelized cost of electricity," in *SolarPACES*, Abu Dhabi, UAE, 2017.

[20] E. Vartiainen, G. Masson, C. Breyer, D. Moser et E. Román Medina, «Impact of weighted average cost of capital, capital expenditure, and other parameters on future utility-scale PV levelised cost of electricity,» *Prog Photovolt Res Appl.,* vol. 28, pp. 439-453, 2020.

[21] G. Manzolini, M. Binotti, G. Gentile, G. Picotti, L. Pilotti and M. E. Cholette, "Actual cost of electricity: An economic index to overcome levelized cost of electricity limits," *iScience,,* vol. 27, no. 6, p. 109897, 2024.

[22] Ekistica, «ARENA's Large-Scale Solar Program: A look at Levelised Cost of Energy,» 2018.

[23] Lazard, «Levelized Cost of Energy: Version 16.0,» 2023.

[24] P. Gilman, «LCOE in SAM using FCR and Single Owner Financial Models,» NREL, Golden, 2020.

[25] E. Veronese, G. Manzolini et D. Moser, «Improving the traditional levelized cost of electricity approach by including the integration costs in the techno-economic evaluation of future photovoltaic plants,» *Int. J. Energy Res.,* vol. 45, pp. 9252-9269, 2021.

[26] R. Frischknecht, P. Stolz, G. Heath, M. Raugei, P. Sinha et M. de Wild-Scholten, «Methodology Guidelines on Life Cycle Assessment of Photovoltaic, Report IEA-PVPS T12-18,» ISBN 978-3-906042-99-2, 2020.

[27] M. Raugei, R. Frischknecht, C. Olson, P. Sinha et G. Heath, «Methodological Guidelines on Net Energy Analysis of Photovoltaic Electricity 2nd Edition, Report IEA-PVPS T12-20,» ISBN 978-3-907281-18-5, 2021.

[28] S. Andreasi Bassi, F. Biganzoli, N. Ferrara, A. Amadei, A. Valente, S. Sala et F. Ardente, «Updated characterisation and normalisation factors for the Environmental Footprint 3.1 method, Publications Office of the European Union,» Luxembourg, doi:10.2760/798894, JRC130796, 2023.

[29] S. Sala, A. Cerutti et R. Pant, «Development of a weighting approach for the Environmental Footprint, EUR 28562 EN,» Publications Office of the European Union, Luxembourg, ISBN 978-92-79-68041-0 (print),978-92-79-68042-7 (pdf), doi:10.2760/945290 (online),10.2760/446145 (print), JRC106545., 2017.

[30] C. Hansen, J. Rippingale, T. Transue, P. Court and J. Gorman, "Orange Button: Accelerating the Digital Transformation of Distributed Energy," in *50th IEEE PVSC*, San Juan, PR, 2023.

[31] DuraMat, "PV Modeling Glossary," 2020. [Online]. Available: https://duramat.github.io/pv-terms/index.html. [Accessed 8 March 2024].

[32] TRUST-PV Project, "TRUST-PV Risk Matrix," 4 October 2022. [Online]. Available: https://trust-pv.eu/reports/risk-matrix/. [Accessed 5 January 2024].

[33] International Electrotechnical Commission, «IEC 60904-2:2015: Photovoltaic Devices, Part 2—Requirements for Photovoltaic Reference Devices,» Geneva, 2015.

[34] M. Korevaar and D. Nitzel, "Simulation of POA front irradiance sensor mounting position," in *European PVPMC*, Mendrisio, 2023.

[35] M. Maeda, A. Nagaoka, K. Nishioka et K. Araki, «Practical and simplified measurements for representative photovoltaic array temperatures robust to climate variations,» *Solar Energy,* vol. 231, pp. 243-251, 2022.

[36] B. Herteleer, B. Morel, B. Huyck, J. Cappelle, R. Appels, B. Lefevre, F. Catthoor and J. Driesen, "High Frequency Outdoor Measurements of Photovoltaic Modules Using an Innovative Measurement Set-Up," in *EU PVSEC 2014*, Amsterdam, 2014.

[37] Univers, "EnOS Monitoring and Control for Solar application," 2024. [Online]. Available: https://univers.com/. [Accessed 8 July 2024].

[38] 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, 2020.

[39] K. Perry, W. Vining, K. Anderson, M. Muller and C. Hansen, "PVAnalytics: A Python Package for Automated Processing of Solar Time Series Data," in *PV Performance Modeling and Monitoring Workshop*, Salt Lake City, 2022.

[40] W. F. Holmgren, C. W. Hansen and M. A. Mikofski, "pvlib python: a python package for modeling solar energy systems," *Journal of Open Source Software,* vol. 3, no. 29, 2018.

[41] M. G. Deceglie, D. Jordan, A. Nag, C. A. Deline and A. Shinn, "RdTools: An Open Source Python Library for PV Degradation Analysis," in *PV Systems Symposium*, Albuquerque, 2018.

[42] B. Meyers, E. Apostolaki-Iosifidou and L. Schelhas, "Solar Data Tools: Automatic Solar Data Processing Pipeline," in *47th IEEE PVSC*, 2020.

[43] 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, 2014.

[44] I. Romero-Fiances, A. Livera, M. Theristis, G. Makrides, J. Stein, G. Nofuentes, J. Casa et G. Georghiou, «Impact of duration and missing data on the long-term photovoltaic degradation rate estimation,» *Renewable Energy,* vol. 181, pp. 738-748, 2021.

[45] M. Theristis, K. Anderson, J. Ascencio-Vasquez and J. S. Stein, "How Climate and Data Quality Impact Photovoltaic Performance Loss Rate Estimations," *Solar RRL,* vol. 8, no. 2, 2023.

[46] D. Lelia, T. Marios, B. King, C. Terrence et J. S. Stein, «Open-source photovoltaic model pipeline validation against well-characterized system data,» *Prog Photovolt Res Appl.,* pp. 1-13, 2023.

[47] L. Breiman, «Random Forests,» *Machine Learning,* vol. 45, pp. 5-32, 2001.

[48] J. Friedman, «Greedy Function Approximation: A Gradient Boosting Machine,» *Ann. Stat.,* vol. 29, pp. 1189-1232, 2001.

[49] NREL, "PVWatts Version 5 Manual," National Renewable Energy Laboratory, 2014.

[50] NREL, "PVWatts Calculator,," [Online]. Available: https://pvwatts.nrel.gov/index.php. [Accessed 19 January 2024].

[51] D. L. King, W. E. Boyson et J. A. Kratochvill, «Photovoltaic Array Performance Model,”,» Sandia National Laboratories, Alberquerque, 2004.

[52] T. Huld, G. Friesen, A. Skoczek, R. P. Kenny, T. Sample, M. Field and E. D. Dunlop, "A power-rating model for crystalline silicon PV modules," *Solar Energy Materials and Solar Cells,* vol. 95, no. 12, pp. 3359-3369, 2011.

[53] K. Ding, Z. Ye et T. Reindl, «Comparison of Parameterisation Models for the Estimation of the Maximum Power Output of PV Modules,» *Energy Procedia,* vol. 25, pp. 101-107, 2012.

[54] S. Lindig, A. Louwen, M. Herz, J. Ascencio-Vasquez, D. Moser and M. Topic, "Performance imputation techniques for assessing costs fo technical failures," in *38th EUPVSEC*, 2021.

[55] S. Lindig, S. Gallmetzer, M. Herz, A. Louwen, E. Koubli, P. S. Enriquez Paes and D. Moser, "Towards the development of an optimized Decision Support System for the PV industry: A comprehensive statistical and economical assessment of over 35,000 O&M tickets," *Prog Photovolt Res Appl.,* vol. 31, no. 12, pp. 1215-1234, 2022.

[56] P. Geurts, D. Ernst et L. Wehenkel, «Extremely Randomized Trees,» *Machine Learning,* vol. 63, pp. 3-42, 2006.

[57] M. Theristis, N. Riedel-Lyngskær, J. S. Stein, L. Deville, L. Micheli, A. Driesse, W. B. Hobbs, S. Ovaitt, R. Daxini, D. Barrie, M. Campanelli, H. Hodges, J. R. Ledesma and I. Lokhat, "Blind photovoltaic modeling intercomparison: A multidimensional data analysis and lessons learned," *Prog Photovolt Res Appl.,* vol. 31, no. 11, pp. 1144-1157, 2023.

[58] A. Forstinger, S. Wilbert, A. Jensen, B. Kraas, C. M. Fernández-Peruchena, C. Gueymard, D. Ronzio, D. Yang, E. Collino, J. Polo, J. Ruiz-Arias, P. Blanc et Y.-M. Saint-Drenan, «Worldwide Benchmark of Modelled Solar Irradiance Data, Report IEA-PVPS T16-05,» ISBN 978-3-907281-44-44, 2023.

[59] C. M. Fernández-Peruchena, J. Polo, L. Martín Pomares and L. Aguiar, "Site-Adaptation of Modeled Solar Radiation Data: The SiteAdapt Procedure," *Remote Sensing,* vol. 12, no. 13, 2020.

[60] J. Ascencio-Vásquez, J. Bevc, K. Reba, K. Brecl, M. Jankovec and M. Topič, "Advanced PV Performance Modelling Based on Different Levels of Irradiance Data Accuracy," *Energies,* vol. 13, no. 9, 2020.

[61] A. Louwen, S. Lindig and D. Moser, "Imputation of Missing Values in Irradiance Datasets," in *37th EUPVSEC*, 2020.

[62] Desert Knowledge Australia Solar Centre, "Download Data - Alice Springs," 2024. [Online]. Available: https://dkasolarcentre.com.au/download?location=alice-springs. [Accessed 15 February 2024].

[63] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye et T. Liu, «LightGBM: A Highly Efficient Gradient Boosting Decision Tree,» *Adv. Neural Inf. Process. Syst.,* vol. 30, p. 3149–3157, 2017.

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss et V. Dubourg, «Scikit-learn: Machine Learning in Python,» *J. Mach. Learn. Res.,* vol. 12, pp. 2825-2830, 2011.

[65] S. Lindig, D. Moser, A. J. R. K. Curran, A. Khalilnejad, R. H. French, M. Herz, B. Müller, G. Makrides, G. Georghiou, A. Livera, M. Richter, J. Ascencio-Vásquez, M. van Iseghem and M. Meftah, "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.

[66] T. Dierauf, A. Growitz, S. Kurtz, J. L. B. Cruz, E. Riley et C. Hansen, «Weather-Corrected Performance Ratio,» NREL, Golden, 2013.

[67] R. J. Ross, "Interface design considerations for terrestrial solar cell modules," in *12th PVSC*, 1976.

[68] J. A. Kratochvil, W. E. Boyson et D. L. King, «Photovoltaic array performance model,» Sandia, 2004.

[69] D. Faiman, "Assessing the outdoor operating temperature of photovoltaic modules," *Prog Photovolt Res Appl.,* vol. 16, no. 4, pp. 307-315, 2008.

[70] B. Herteleer, A. Kladas, G. Chowdhury, F. Catthoor and J. Cappelle, "Investigating methods to improve photovoltaic thermal models at second-to-minute timescales," *Solar Energy,* vol. 263, p. 111889, 2023.

[71] M. Prilliman, J. S. Stein, D. Riley et G. Tamizhmani, «Transient Weighted Moving-Average Model of Photovoltaic Module Back-Surface Temperature,» *IEEE Journal of Photovoltaics,* vol. 10, pp. 1053-1060, 2020.

[72] A. Kladas, B. Herteleer et J. Cappelle, «Deep Learning Temperature Estimation Model for PV Modules,» chez *WCPEC-8*, Milan, 2022.

[73] A. Kladas, B. Herteleer et J. Cappelle, «Blind PV temperature model calibration,» *EPJ Photovoltaics,* vol. 14, p. 28, 2023.

[74] A. Livera, M. Theristis, E. Koumpli, S. Theocharides, G. Makrides, J. Sutterlueti, J. S. Stein and G. E. Georghiou, "Data processing and quality verification for improved photovoltaic performance and reliability analytics," *Prog Photovolt Res Appl.,* vol. 29, no. 2, pp. 143-158, 2021.

[75] 3E, "3E SynaptiQ," [Online]. Available: https://synaptiq.3esynaptiq.com/sq2020. [Accessed 01 October 2024].

[76] European Centre for Medium-Range Weather Forecasts, "Homepage: ECMWF," [Online]. Available: https://www.ecmwf.int/en/research. [Accessed 8 July 2024].

[77] European Centre for Medium-Range Weather Forecasts, "Homepage: ECMWF," [Online]. Available: https://www.ecmwf.int/en/research/data-assimilation/observations. [Accessed 8 July 2024].

[78] J. Ascencio-Vásquez, "Doctoral dissertation: Modelling and Worldwide Assessment of Performance and Aging of Photovoltaic Modules and Systems," 2021.

[79] L. Haohui, Z. Defreitas, T. Congyi, G. Binnard et Z. Lu, «A Holistic Approach to Technical Asset Management,» Univers, 2021.

[80] S. Lindig, "Doctoral dissertation: Impact of outdoor effects, degradation and technical failures on uncertainty in energy performance of photovoltaic systems," 2021.

[81] "Kipp & Zonen, DustIQ," [Online]. Available: https://www.kippzonen.com. [Accessed 10 October 2024].

[82] "Fracsun," [Online]. Available: https://www.fracsun.com. [Accessed 10 October 2024].

[83] "Atonometrics," [Online]. Available: https://www.atonometrics.com. [Accessed 10 October 2024].

[84] Å. Skomedal, H. Haug and E. Stensrud Marstein, "Endogenous Soiling Rate Determination and Detection of Cleaning Events in Utility-Scale PV Plants," *IEEE Journal of Photovoltaics,* vol. 9, no. 8, pp. 858-863, 2019.

[85] M. M. Nygard, Å. Skomedal, M. S. Wiig and E. Stensrud Marstein, "Combined Degradation and Soiling With Validation Against Independent Soiling Station Measurements," *IEEE Journal of Photovoltaics,* vol. 13, no. 2, pp. 296-304, 2023.

[86] Joint Research Centre, "PV-GIS application," [Online]. Available: https://re.jrc.ec.europa.eu/pvg_tools/en/. [Accessed 13 August 2024].

[87] 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. 4749, 2019.

[88] Å. Skomedal, B. L. Aarseth, H. Haug, J. Selj and E. Stensrud Marstein, "How much power is lost in a hot-spot? A case study quantifying the effect of thermal anomalies in two utility scale PV power plants," *Solar Energy,* vol. 211, pp. 1255-1262, 2020.

[89] B. L. Aarseth, Å. Skomedal, M. B. Ogaard and E. Stensrud Marstein, "Detecting Permanently Activated Bypass Diodes in Utility-Scale PV Plant Monitoring Data," *IEEE Journal of Photovoltaics,* vol. 12, no. 5, pp. 1230-1236, 2022.

Technology Collaboration Programme

by iea

ISBN 978-3-907281-69-7

9 783907 281697