A Combined Approach for Model-Based PV Power Plant Failure Detection and Diagnostic

Christopher Gradwohl 1,*, Vesna Dimitrievska 2,*, Federico Pittino 2, Federico Pittino 2, Wolfgang Muehleisen 2, András Montvay 3, Franz Langmayr 4 and Thomas Kienberger 1

1 Energy Network Technology, Montanuniversitaet Leoben, Franz-Josef Str18, 8700 Leoben, Austria; thomas.kienberger@unileoben.ac.at
2 SAL Silicon Austria Labs GmbH, Europastr.12, 9524 Villach, Austria; Federico.Pittino@silicon-austria.com (F.P.); Wolfgang.Muehleisen@silicon-austria.com (W.M.)
3 SAL Silicon Austria Labs GmbH, Infeldgasse 33, 8010 Graz, Austria; András.Montvay@silicon-austria.com
4 Uptime Engineering GmbH, Schoenaugasse 7/2, 8010 Graz, Austria; f.langmayr@uptime-engineering.com
* Correspondence: christopher.gradwohl@unileoben.ac.at (C.G.); vesna.dimitrievska@silicon-austria.com (V.D.); Tel.: +43-3842-402-418 (C.G.); +43-664-88200120 (V.D.)

Abstract: Photovoltaic (PV) technology allows large-scale investments in a renewable power-generating system at a competitive levelized cost of electricity (LCOE) and with a low environmental impact. Large-scale PV installations operate in a highly competitive market environment where even small performance losses have a high impact on profit margins. Therefore, operation at maximum performance is the key for long-term profitability. This can be achieved by advanced performance monitoring and instant or gradual failure detection methodologies. We present in this paper a combined approach on model-based fault detection by means of physical and statistical models and failure diagnosis based on physics of failure. Both approaches contribute to optimized PV plant operation and maintenance based on typically available supervisory control and data acquisition (SCADA) data. The failure detection and diagnosis capabilities were demonstrated in a case study based on six years of SCADA data from a PV plant in Slovenia. In this case study, underperforming values of the inverters of the PV plant were reliably detected and possible root causes were identified. Our work has led us to conclude that the combined approach can contribute to an efficient and long-term operation of photovoltaic power plants with a maximum energy yield and can be applied to the monitoring of photovoltaic plants.

Keywords: PV system; failure detection; failure diagnostic; operation and maintenance; predictive-and reliability-based maintenance; model-based state detection; physical model; one-diode model; statistical model; virtual sensors

1. Introduction

Driven by global energy policies, solar power has become one of the key sources of renewable energy and is the most attractive source of renewable power generation today. In 2019, solar power accounted for 48% of newly added net power capacity, which is more than all new capacities of fossil power combined and almost twice as much as new wind power capacities. “Solar Power Europe” reported that in 2019, the global capacity of installed solar power reached a total of 630 GW [1]. This development is enabled and driven by the dramatically decreased levelized cost of electricity (LCOE) for solar power plants. Lazard [2] found that the average LCOE for solar power plants reached 40 USD/MWh, which is significantly lower than the LCOE of unsubsidized fossil and nuclear power plants and at the same level as wind power. The leading role of solar power in global electricity production is expected to be strengthened in the coming years. “Solar Power Europe” estimates that the global installed solar capacity will grow to a total of 1.2 TW until 2022. While this is a huge potential for solar power, it also comes with increased cost pressure and
a highly competitive market environment. Therefore, ensuring maximum performance is a requisite for securing profit margins and the bankability of photovoltaic (PV) projects. In this environment, optimized operation and maintenance (O&M) activities are one measure to maximize plant performance. The importance of early detection of performance losses and initiation of targeted countermeasures becomes particularly evident. Considering the cost-effectiveness of the different techniques for failure detection, an efficient procedure for identifying failures should focus first on performance loss identification via the monitoring system, followed by, if needed, other on-site inspections [3].

Today, modern PV power plants are typically equipped with a monitoring system aiming at collecting data from the main components such as inverters, energy meters, and meteorological sensors. Standards such as IEC 61724 [4] recommend data acquisition with a granularity of 15 min. These recorded data are typically used for the calculation of standardized performance indicators such as the performance ratio or technical availability. In recent years, computational methods for fault diagnostics include statistical methods, simulations, and intelligent techniques [5]. For example, intelligent methods for detection of a subset of common failures have been developed by Pei et al. [6] and Basnet et al. [7]. Unfortunately, these methods require a set of data, labeled with the known failure type, which is typically not available in a monitoring system. Other advanced sensing and monitoring methodologies, such as temperature and humidity measurements within PV modules [8] or visual monitoring, such as monitoring from areal systems [9], are typically not implemented due to cost restrictions. However, the ability of an accurate detection, diagnosis, and prognosis of failures and degradation could significantly improve the efficiency of O&M activities, maximize performance, and thus further reduce the LCOE [10].

Typically, maintenance strategies can be classified as corrective and preventive strategies [10]. In a corrective maintenance strategy, action is taken after a failure occurs. In contrast, preventive maintenance strategies aim at preventing the occurrence of failures by taking action before a failure occurs and prolonging the residual lifetime. Preventive maintenance can be further classified as time-based, condition-based, and predictive approaches [11]. Time-based preventive maintenance is a simple strategy where maintenance actions are scheduled periodically based on predefined intervals. Condition-based maintenance can be applied in cases where the component condition can either be measured directly (level of wear) or linked to measurable factors such as vibration which indicate the component condition. This approach often requires signal processing and automated algorithms, which makes it a more cost-intensive strategy. Predictive maintenance strategies employ prognostic models to estimate a residual lifetime and schedule maintenance activities accordingly [10]. Further steps towards more elaborate interventions based on what-if analysis, referred to as prescriptive maintenance, are being investigated, especially in the context of Industry 4.0 [12]. The applicability of maintenance strategies for individual components highly depends on the availability of relevant information for decision making.

The main aim of this paper is to demonstrate how appropriate failure detection and diagnostics can deliver valuable information for corrective or predictive maintenance strategies based on typically available supervisory control and data acquisition (SCADA) data. A key aspect in delivering valuable information for maintenance actions is to identify abrupt or gradual performance degradations and the cause of these failures in a quick and reliable manner within PV plant monitoring. On the other hand, diagnostic methods need to be implemented to identify the most probable causes of the detected failures. This information is valuable for O&M companies to plan the most appropriate actions and to correct the failures that have occurred.

Overall, the presented work was conducted within the framework of the OptPV4.0 research project (FFG number 871684) which is funded by the Austrian Climate and Energy Fund and carried out as part of the Energy Research Program 2018. The OptPV4.0 project aims to optimize the operation of solar power plants in order to increase and guarantee the
energy yield and profitability of these systems. Therefore, a loss of performance caused by abrupt failures or gradual degradations must be prevented by identifying the causes of failure in a quick and reliable manner.

Accordingly, in this paper, we propose a combined approach, in order to implement, in the first stage, a reference model of a PV plant in a healthy state to detect abnormal behavior. In the second stage, failure diagnosis is performed by using failure modeling of relevant degradation mechanisms. Previous work by Ali et al. [13] showed that physical models can be used to predict and compare electrical parameters of PV modules by means of specialized IV measurements. To face the challenge of advanced sensing, we propose a novel approach to model reference values and the actual behavior of PV strings suitable for typical available SCADA data, based on previous work by Villalva et al. [14] about the so-called “diode models” of PV cells. On the other hand, statistical models have been successfully used to model and predict daily power output [15], by means of weather data. We propose using supervised predictive models to predict the expected values of the electrical parameters in a PV system. Examples of such models are artificial neural networks, support vector regression, decision trees, and linear regression models. The available SCADA data of the electrical parameters during a period in which error-free operation of the PV system is assumed could be used in the learning phase.

Consequently, this paper attempts to create a reference model of a PV plant that allows predicting the expected value of the electrical parameters for every measured time step. The prediction should be performed in a fast and efficient manner, in terms of time and use of computational power, that is appropriate for online monitoring. Furthermore, this study intends to determine the similarities, advantages, and disadvantages of using statistical and physical models for this approach. Finally, several metrics to quantify the deviation of the measured values with respect to the reference values of a physical or statistical model and the time of occurrence are introduced. A common issue with current failure detection methods is that they are not able to accurately diagnose the cause of unusual behavior. This is necessary to remedy the abnormal performance by initiating an efficient maintenance activity. Previous research by Escobar et al. [16] established that the physics of failure approach identifies the most likely failure modes, but one of the greatest challenges is the limitation in inaccurately predicting the time of occurrence of failures. Consequently, the novel objective of this work is to combine the advantages of failure detection and diagnosis to identify suspicious behavior of PV power plants and the underlying root cause.

In this work, Section 2 presents the methods used for the combined modeling approach. A description of the monitoring data of the PV plant considered as a case study, together with the results and their discussion obtained with the proposed approach, is given in Section 3. Finally, Section 4 presents the conclusions.

2. Methodology

This section provides an overview of the applied methodologies to implement a digital representation of photovoltaic (PV) power plants as a type of “digital twin”. The overall goal is to realize a model-based concept of failure detection and diagnosis to optimize PV power plant operation and maintenance. Therefore, we explain the background of physical and statistical modeling of PV power plant behavior and the concept of physics of failure in Sections 2.1 and 2.2.

The basis of this approach is the acquisition of environmental and operating conditions (EOC) by means of available SCADA data. These data are used as an input for the physical and statistical models and the physics of failure approach. While the physical model uses the real irradiance and module temperature, the statistical model, in addition, uses sensor data such as ambient temperature and time-dependent data. Following that, the physics of failure models use individual inputs depending on the represented failure mode. By continuously comparing the modeled performance with the actual performance, performance losses can be detected immediately. Once a performance loss and thus a possible failure are detected, we use the quantification of damage accumulation from the physical damage
models to rank possible failure modes according to the risk of failure occurrence. With this combined concept, we can reliably detect performance issues, diagnose corresponding failure modes, and recommend targeted maintenance action.

2.1. Failure Detection Models

The aim of the failure detection models is to provide a fault-free reference condition by simulating the state and operation of a PV system in a holistic manner. For this purpose, the models incorporate the relationship between sensor data, such as irradiance and module temperature, and the electrical parameters. In particular, we use physical models to simulate the expected behavior of a PV system by means of the so-called “diode models” of PV cells. Moreover, statistical models are used to predict the expected electrical parameters by means of statistical prediction models. By using these models in combination with measured sensor data, the electrical parameters such as the DC current, DC voltage, and DC power can be predicted for every timestamp. However, this work provides an important opportunity to advance the understanding of the use of physical and statistical models for online monitoring, the advantages and disadvantages of both approaches, and the accuracy of predicting reference conditions.

To achieve this goal, two types of models were created: physical and statistical models, which differ by the data used to define the models. A detailed description of these models is given in Sections 2.1.1 and 2.1.2, respectively.

2.1.1. Physical Models

A digital representation of the expected PV system behavior as a function of real irradiance and temperature conditions can be modeled by the so-called “diode models” of PV cells. PV cells have a nonlinear current–voltage (IV) characteristic curve, where the current and power of the PV array depends on the terminal operating voltage. Furthermore, the maximum power point (MPP) varies with the environmental conditions, such as the solar irradiation and ambient and module temperatures. Different approaches of photovoltaic cell models are used to describe the real electrical behavior of PV cells [17]. These can be classified as one- and two-diode models. Based on the one-diode model, the output current of a PV cell can be modeled by considering the generated photocurrent $I_{PH}$ and Shockley’s equation to describe the exponential electrical current–voltage $I(U)$ characteristic of a p–n junction. In addition, the real electrical behavior of the cells caused by its losses are described by serial ($R_S$) and parallel ($R_P$) resistances. The equivalent circuit of the one-diode model is shown in Figure 1.

![One-diode model of a photovoltaic (PV) cell.](image)

The two-diode model is seen as more accurate in representing the behavior of photovoltaic cells, especially for cells with a higher bandgap energy, since the recombination rate at higher temperatures within the depletion region is also considered [17]. Due to more complex model parameters and the minimal performance advantage of the two-diode model, the modeling approach of this work was based on the one-diode model.
The output current of the photovoltaic cell of the one-diode model can be described by Equation (1) [14]. As depicted in the equation, the accurate calculation of the output current requires iterative computing with a feedback loop, which was implemented as a block diagram model in Simulink as proposed by [14].

\[
I(U) = I_{PH} - I_D - I_P = I_{PH} - I_0 \cdot \left( e^{\frac{U + I \cdot R_S}{U_T}} - 1 \right) - \frac{U + I \cdot R_S}{R_P}
\]  

(1)

In Equation (1), \(I_{PH}\) is the photocurrent, \(I_D\) is the diode current, \(I_P\) is the current through the parallel resistance, \(I_0\) is the diode reverse saturation current, \(U_T\) is the thermal voltage, \(m\) is the ideality factor, \(R_S\) is the series resistance, and \(R_P\) is the parallel resistance.

Most of the model parameters described above are based on datasheet values from the manufacturer and can be used directly for modeling. Only the series resistance \(R_S\) and the parallel resistance \(R_P\) need to be determined by a Newton–Raphson iterative calculation, as introduced in [14]. By performing this approach, Equation (1) needs to be resolved by \(R_P\) and multiplied by the voltage at the maximum power point (\(U_{MPP}\)) at standard test conditions (STC). By performing the iteration, the IV curve will be emulated by varying the values \(I_{MPP}\) (current at the maximum power point) and \(U_{MPP}\) until the experimental maximum power equals the datasheet power. At this point in the IV curve, only one pair of values of \(R_P\) and \(R_S\) exists, which represents the electrical behavior of the photovoltaic cell at the MPP [14]. By using the determined \(R_P\) and \(R_S\) values and the one-diode model, one can calculate an accurate photovoltaic module current of a PV module or the whole PV string by using Equation (1). Our work allowed a novel solution to model characteristic diagrams of the maximum power point power (\(P_{MPP}\)), current (\(I_{MPP}\)), and voltage (\(U_{MPP}\)) at a wide range of load situations, as depicted in Figures 2 and 3, by varying the input parameters, module temperature (\(T\)), and global tilted irradiance (\(G\)). Characteristic diagrams of this kind will later be used for calculating reference values for PV plant monitoring and failure detection.

![Figure 2. Characteristic diagram of P_{MPP} f(T,G).](image)
2.1.2. Statistical Models

Statistical models use the measured data to build models that learn the relationship between the input and output data. In the given task to model the expected PV system behavior, the output data are one of the electrical parameters: current (I_{MPP}), voltage (U_{MPP}), or power (P_{MPP}) at the maximum power point. Implementing statistical or intelligent models to predict the expected electrical parameters has also been applied in [18,19] as a step in failure detection. For example, to predict the power, a recursive linear model was used in [19], having the measured irradiance and the previously predicted power as the input, while a neural network was used in [18] to predict the maximum output power given the open-circuit voltage, short-circuit current, and other weather data.

In our scenario, the input data include the measured and valid sensor parameters available in a monitoring system. Examples of sensor parameters used as input are global tilted irradiance (G), ambient temperature (T_{AMB}), module temperature (T), and humidity (H). Taking all measured sensor parameters, a selection should be performed to choose the best set to be used as input to the models. To grasp the temporal and seasonal effects, the parameters for the hour in the day and day in the year calculated for each timestamp are added to the input data.

All measured data are divided into training data, used as input to build the statistical models, test data, used as a control set to estimate the performance of the model, and evaluation data that are inspected for failures. The assumption is that the PV plant is in fault-free conditions at the start of its operation. Hence, the training data and test data present the measured data before a given timestamp t^{split} and the evaluation data present the data after t^{split}. The selection of the time t^{split} depends on the available data. The training and test data are obtained by a random selection with a proportion of 70% and 30%, respectively. To be able to grasp the seasonal effects through the year, it is desired that data from all seasons are present, while to prevent overfitting of the data in a particular year, data from multiple years are needed.

Before building the models, the data cleaning process is performed. For this purpose, custom filters are defined for different types of data. For the electrical data and irradiance data, filters are defined that discard the data outside the given valid range. Based on our experience, the relationship between the irradiance and power is only linear in the range between 50 and 1200 W/m², which is set as a valid range for irradiance. Another filter checks the valid ranges for a set of electrical parameters (Table 1). It is used to select the timestamps where a significant production of power is observed. This filter is used only for the training and test data and enables only valid data to be used for the modeling. For the temperature values, global and local outliers are detected and discarded. The global filter finds all values in the time series that are outside the interval (q_1 - 1.5IQR, q_3 + 1.5IQR), where q_1 and q_3 are the 25th and 75th percentiles of the data, respectively, and IQR is the

![Figure 3. Characteristic diagram of I_{MPP} f(T,G) (a) and characteristic diagram of U_{MPP} f(T,G) (b).](image-url)
interquartile range that is calculated by $IQR = q_3 - q_1$. The local filter investigates daily values and discards all values that are outside the interval $(m - 3\sigma, m + 3\sigma)$, where $m$ is the mean and $\sigma$ is the standard deviation of all values in the specific day. At the end of the cleaning process, the data from all days, for which more than 90% of the data are missing or have been filtered out in previous steps, are discarded entirely.

Table 1. Valid ranges defined for different types of data.

| Data Type   | Valid Ranges           |
|-------------|------------------------|
| Irradiance  | 50 W/m²–1200 W/m²      |
| AC Current  | >0.05 A                |
| DC Current  | >0.05 A                |
| AC Voltage  | 207–253 V              |
| DC Voltage  | ≠0 V                   |

A variety of machine learning methods can be used to build statistical models. After an evaluation of several available machine learning methods, support vector regression (SVR) was chosen [20]. SVR models are capable of learning the nonlinear dependency between the electrical parameters and the temporal parameters in the input data, which would not be possible with linear models. The model implementation and evaluation are conducted in Python using the Scikit-Learn library [21]. The default parameters are used to build the SVR model, using the radial basis function as a kernel function. The input data to the SVR model are explained in detail in Section 3.1. For a better performance of SVR, the data used for the modeling and evaluation are standardized using the mean and standard deviation of the training and test dataset. After the training and validation of the statistical model, it can be used to predict electrical parameters such as $P_{\text{MPP}}$, $U_{\text{MPP}}$, and $I_{\text{MPP}}$, for given input data at each timestamp of the evaluation dataset. A reverse process to standardization is conducted to calculate the outputs in the initial units (W, V, A) from the standardized value predicted with the model.

2.2. Physics of Failure

The concept of physics of failure assumes that damage of components is accumulated due to an irreversible change in the microstructure of any component subjected to specific load conditions [16]. The accumulation of damage takes place incrementally over the operational lifetime of the component, although the effect might not be visible until a failure occurs. Load conditions such as thermal, electrical, or mechanical loads are directly induced by the environmental and operating conditions of the system. By quantifying the incremental damage as a function of the environmental and operating conditions (EOC), it is possible to derive statements about the condition of a component with respect to specific failure modes. The term failure mode is generally understood to mean how a device, equipment, or a machine can fail [16].

The first step towards physical damage models is an analysis of a component in terms of possible failure modes. In this context, a detailed analysis of failure mechanisms in photovoltaic systems and their subsystems was carried out in [22]. Based on this analysis, it can be said that inverters, rectifiers, bypass diodes, photovoltaic modules, cabling, and AC circuit breakers are the most critical components in the system. In most cases, there are several failure mechanisms for each component. An investigation of photovoltaic systems showed that wire bonding liftoff and solder fatigue of power electronic devices, as well as thermal aging of capacitors, are seen as the most probable failure mechanisms in inverters and rectifiers [23,24]. Furthermore, failure mechanisms such as delamination, backsheet cracking, cell cracks, potential-induced degradation (PID), burn marks, disconnections of ribbons, and defective bypass diodes are widely reported in photovoltaic modules [25]. The goal of this analysis is to find the physical mechanism which causes a specific failure and to focus on components and failure modes that have a significant impact on the reliability and thus the performance of a PV power plant. Once the possible failure modes and physical
root causes are identified, the analysis focuses on load and environmental boundary conditions which promote the incremental accumulation of damage for each identified failure mode. In this context, it has been shown that temperature and thermal cycling are seen as the most critical operation conditions for power electronic devices [26] and for photovoltaic modules, which, in addition, see humidity and solar radiation, especially UV radiation, as critical damage drivers [27]. Based on the understanding of failure modes and corresponding damage drivers, the physical damage models will be derived. These models describe the relation between the EOC as an input and the accumulation of damage kinetics as an output. To describe the physical damage modeling approach exemplarily, hydrolysis-driven failure mechanisms such as delamination of PV module junctions will be described in more detail. These damage mechanisms see temperature and humidity as the most critical load situations, and it has been shown in [27] that these mechanisms can be accurately described and calculated by an extended Arrhenius approach with Equation (2).

\[
k_H = A_H \cdot RH^n \cdot e^{-\frac{E_H}{kT_M}}
\]

(2)

\(k_H\) describes the damage kinetics, \(A_H\) is a curve fitting pre-exponential factor, \(RH\) is the humidity, \(n\) is a model parameter, \(E_H\) is the activation energy for hydrolysis-driven reactions, \(k\) is Boltzmann’s constant, and \(T_M\) is the mean PV module temperature.

Further, damage evaluation is performed by applying damage models to time series of load data over the operational lifetime of the PV power plant to deliver the rate of degradation. Subsequently, accumulating the degradation rate delivers the damage sum, indicating the amount of endurance, consumed for each failure mode due to exposure to the load history. Furthermore, critical load conditions are identified and understood in detail which, in some cases, allows avoiding damage-promoting operating conditions.

3. Results

To demonstrate the failure detection and diagnosis ability of our approach, a case study based on six years of SCADA data from a PV power plant in Slovenia was conducted within the scope of the OptPV4.0 R&D project. The PV power plant has an installed capacity of about 315 kWp and consists of 19 inverters, 57 strings, and 1313 pieces of 240 Wp modules. In addition, each inverter features two maximum power point trackers (MPPTs), whereby tracker one operates two parallel strings and tracker two operates only one string, where every string has the power of 5520 watts. Furthermore, for structural reasons, two different orientations of the PV strings are applied to the power plant: southwest (200°) and southeast (135°). A detailed explanation of the measured SCADA data, which were recorded with a granularity of 15 min by the PV plant operator, will be given in the following Section 3.1.

3.1. Data

First, available weather sensor data, such as global tilted irradiance (G) and temperature (T), will be investigated in more detail. As seen in Figure 4a, the irradiance measurement at the PV plant site shows a great lack of data. To overcome this issue, missing data were filled in with data from a neighboring plant into two additional datasets. The only difference in these data lies in the orientation of the PV strings. Consequently, the corrected irradiance values for the southwest orientation of 200° (Figure 4b) were recalculated to a southeast orientation of 135° (Figure 4c) on the basis of Quaschning [28]. By investigating the data of the measured module temperature (T) (Figure 5a), a gradual increase in temperature can be observed in the time series. This issue can most likely be explained by temperature sensor problems, which is in good agreement with the previously documented sensor issues by the plant operator. By inserting the data from a neighboring plant, an unusual change in the data in 2018 appears (Figure 5b). The last weather parameter represents the ambient temperature (T_{AMB}), shown in Figure 5c, where missing data can also be seen, but the values seem to be consistent through the years.
An important step in implementing the statistical models is the selection of input parameters from all available irradiance and temperature parameters. For this purpose, the test accuracy of the SVR models, using different sets of input parameters, was calculated. Experimental results showed insignificant differences between the performance of the models using the different irradiance parameters; hence, the selected parameter is the one shown in Figure 4b. Due to the issues in the available module temperature parameters that could lead to erroneous predictions, these parameters are not used as input in the models. Instead, even though there are lots of missing data, the measured ambient temperature ($T_{AMB}$) is used (Figure 5c). One disadvantage of this approach is that in cases of missing data, no predictions can be made. As a result, an evaluation of the performance of the system is not possible. On the other hand, the selected input leads to more accurate results of the statistical model. The selected parameters, together with the added parameters “Time in day” and “Day in year”, comprise the input for the statistical models (Table 2).

![Figure 4](image1.png) ![Figure 5](image2.png)

**Figure 4.** Visualization of the measured irradiance data (a) and measured irradiance with filled data gaps from a neighboring plant with orientations of 200° (b) and 135° (c).

**Figure 5.** Visualization of the temperature data: (a) measured module temperature, (b) module temperature with filled in gaps, and (c) measured ambient temperature.

| Physical model | Input | Output |
|----------------|-------|--------|
| Global tilted irradiance (G), Module temperature (T) | Global tilted irradiance (G), Ambient temperature ($T_{AMB}$), “Time in day”, “Day in year” | DC current ($I_{MPP}$), DC voltage ($U_{MPP}$), DC power ($P_{MPP}$) |

**Table 2.** Input and output summary for the physical and statistical models.

For the physical models, the corrected irradiance values (Figure 4b,c) and the corrected module temperature (Figure 5b) are used as inputs for calculating electrical reference parameters such as $P_{MPP}$, $U_{MPP}$, and $I_{MPP}$. A more detailed explanation will be given in Section 3.2.

Another point that depends on the data availability is the selection of the parameter $t_{split}$. Considering that the data are available from 2014, the parameter $t_{split}$ was set to 1.1.2015. Although a time interval of one year was taken for training, because of missing data, data from only 141 days were used for building the models. Such a choice for the
parameter $t_{\text{split}}$ is empirically validated by the satisfactory accuracy of the models. As a result, the evaluation of the PV plant performance, using the statistical models, can only be conducted from 1.1.2015.

3.2. Model-Based Failure Detection

3.2.1. Application of Reference Models

For continuous performance monitoring purposes, efficient calculations in terms of time and use of computational power are important. Therefore, in the first step, characteristic diagrams for predicting the expected reference values (as depicted in Figures 2 and 3) were derived from the one-diode model. Furthermore, a regression model as a function of the module temperature ($T$) and irradiance ($G$) was calculated from the characteristic diagram, which allows processing large amounts of data within a short period of time and allocates reference values of “$P_{\text{MPP}}$”, “$U_{\text{MPP}}$”, and “$I_{\text{MPP}}$”.

In addition, a separate SVR-based model was created for each MPP tracker of the PV plant, which uses the training and test data from the targeted tracker as input. The custom models for the MPPT should capture the features specific to the corresponding MPPT, but also to the PV modules connected to it. To demonstrate the performance of the applied model, Figure 6a depicts the predicted and measured $P_{\text{MPP}}$ values of PV modules with different orientations (southwest and southeast) that are connected to two different MPP trackers. The different orientations can be observed in the data by the shifts in the peaks with respect to the time in the day. In this case, the time-dependent variables are exploited to build models that capture this dependency. Similarly, the different capacities of the two trackers of inverter 2U01 are captured in the separate models (Figure 6b).

![Figure 6](image_url)

**Figure 6.** Measured and predicted values for 16 April 2019: (a) MPPT 1 for inverters 2U01 and 2U04; (b) MPPTs 1 and 2 for inverter 2U01.

To further illustrate the performance of the SVR models, let us take the model created for the prediction of $P_{\text{MPP}}$ for MPPT 1 of inverter 2U01. To validate the performance of the model, the residual error, which is the difference between the predicted and the measured $P_{\text{MPP}}$ values, was calculated for the test data set. The average residual error is 8.7 W or only 4.7%. Another metric for the performance of the model is the square of the simple correlation coefficient ($r^2$) between the measured and predicted outputs. The calculated value of $r^2$ (0.98) for the test data shows a strong linear relationship between the measured and predicted outputs. Therefore, the model is sufficiently accurate to be used for $P_{\text{MPP}}$ prediction within online monitoring.

The real-time application of the calculated reference value “$P_{\text{MPP}}$” to online monitoring for an exemplary string of inverter 1U01 MPPT 2 (1U01-T2) is depicted in Figure 7. Data gaps for the modeled power of the statistical model in Figure 7 are caused by the data cleaning procedures, where the data outside the valid ranges are discarded for the analysis and evaluation. As seen in Figure 7, for most timestamps, there is an insignificant difference between the modeled and the measured power. On the other hand, near the daily peak levels, the statistical model predictions are closer to the measured values, compared to
the physical model predictions (Figure 7). To be able to understand this phenomenon, an in-depth analysis was conducted to evaluate the overall performance of the statistical and physical models.

![Figure 7. Application of the physical and statistical models for inverter 1U01-T2: measured vs. modeled power.](image)

The analysis was performed by means of two sample MPP trackers “1U01-T2” and “1U02-T1”. Hence, the relative absolute error was calculated as the ratio of the absolute error and the measured power. Note that only timestamps where both the physical and statistical model predictions are available were considered. The empirical cumulative distribution of such errors is shown in Figure 8a. It can be observed that both models have high accuracy. In fact, for the inspected inverters, the error is less than 1% in more than 90% of the cases (Figure 8a). Furthermore, Figure 8b shows the cumulative distribution of the errors (difference between modeled and measured values) expressed in watts (W). Here, a similarity between the error distributions of the physical and statistical models can be clearly seen. This is further confirmed by a high correlation ($r^2$) of around 0.97 between the predictions of the physical and statistical models of the inspected inverters. By analyzing the shape of the cumulative distribution for the physical models and statistical models, two observations are made. First, in the highest range of positive errors, more errors are seen for the physical models, where the modeled power is overestimated. Examples of these errors are seen near the maximum value in the daily operation in Figure 7. The high positive errors can most likely be explained by the fact that physical models predict the rated power of an MPPT in a healthy state, which might have never been achieved by the MPPT since the start of its operation. The statistical model learns from the measured power in the first year and therefore shows lower deviations at the highest power (as can be seen in Figure 7). Second, there are more negative errors in the physical models. The effect of this observation can be observed near the start and end of the daily operation in Figure 7. An explanation can be that physical models produce higher errors for lower ranges of irradiation, due to the fact that actual control strategies of inverters in these load situations are not considered by the physical model.

The high accuracy and similarity between the performances of the physical and statistical models suggest that both types of models can be used to implement the reference models used for online monitoring. Since both types of models can use different input data (Table 2), a selection of the model can be made based on data availability. In the case where both models can be implemented, we believe that the redundancy can contribute to the reliability of the failure analysis. Although high accuracy of the models is explored in the given example (Figure 8), we expect that a lower accuracy of the models can be caused by erroneous input data. Additionally, a limitation of the statistical models is that their performance depends on the validity and coverage of the training data. For example,
training data with an introduced failure or training data that do not cover all seasons can lead to a poor estimation.

A comparison between the distribution for the trackers “1U01-T2” and “1U02-T1” leads to a conclusion that both models present higher accuracy for “1U01-T2” (Figure 8). The deviation can be explained by performance issues of the MPP tracker “1U02-T1”. Therefore, we propose applying statistical analysis to the measured and modeled time series of the PV plants, which can be applied to all MPPTs.

3.2.2. Energy Performance Index

In [29], the authors proposed their failure detection methods on the residuals used to quantify the difference between the measured and modeled values. In our approach, the so-called “energy performance index” (EPI) [30] and reference values (P_{\text{MPP}}, U_{\text{MPP}}, P_{\text{MPP}}) were used to determine the photovoltaic power plant performance and hence possible long-term degradations. The EPI (Equation (3)) is defined as the ratio of the measured energy and the expected energy as a function of the actual load condition.

\[
EPI = \frac{E_{\text{Measured}}}{E_{\text{Model}}} \tag{3}
\]

The EPI is calculated in regular assessment periods such as days, months, or years. We determined an EPI value for every available time step (in our case, 15-min intervals, based on the data acquisition of the PV plant operator) for all MPPTs of the photovoltaic park described above. In addition, a median value of the EPI values was calculated for each month (EPI(m)). The monthly median values for every MPPT were calculated using the modeled energy from both the physical and statistical models. The results are depicted in Figures 9 and 10, respectively. Furthermore, an EPI median value of all investigated strings (EPI_{\text{median}}(m)) was determined to detect suspicious outliers and gradual degradations of the MPPT performance.

However, the model-based failure detection by means of the physical model revealed a significant deviation at MPPTs “1U03-T1”, “1U09-T1”, “2U01-T1”, and “2U04-T1” for a limited time span, as well as a gradual performance decrease of MPPTs “1U02-T1”, “1U04-T1”, “1U09-T2”, and “2U02-T1”. Similar results were shown by using the statistical models (as depicted in Figure 10).

Besides the similarities in the results provided by both physical and statistical models, there are some differences. In particular, the use of physical models revealed EPI values

Figure 8. Empirical cumulative distribution of the relative absolute errors (a) and error (b) obtained with the modes for the maximum power point (MPP) trackers “1U01-T2” and “1U02-T1”.

(a) (b)
greater than one, indicating an impossible efficiency above the expected nominal value. It is very likely that this effect can be explained by a lack of measured temperature and irradiance data values, which led to fewer modeled values for an accurate EPI calculation in certain assessment periods. Consequently, these findings allow detecting sensor problems in PV systems in an easy manner. When using statistical models, EPI values greater than one are seen mostly for “1U09-T1”. This can most likely be explained by problems in the training data, but a more detailed explanation is given later.

In contrast, when using statistical models, a greater number of monthly indexes are missing, caused by missing data in the input parameter $T_{AMB}$, as well as by the data cleaning process. More precisely, in the evaluation dataset, the data are complete for only 52% of all 1925 days. Consequently, there is a lack of information about the performance of the PV plant in the periods of missing data. Furthermore, an occurrence of EPI values

\begin{align}
R_{EPI}(m) &= EPI(m) - EPI_{median}(m) \\
EPI_{median}(m) \times 100 \\
R_{CPI}(m) &= CPI(m) - CPI_{median}(m) \\
CPI_{median}(m) \times 100 \\
R_{VPI}(m) &= VPI(m) - VPI_{median}(m) \\
VPI_{median}(m) \times 100
\end{align}

Figure 9. Energy performance index (EPI)—string (MPP trackers (MPPTs)) comparison (physical model).

Figure 10. EPI—string (MPPTs) comparison (statistical model).
of 0.6 in some of the autumn months for the inverter 1U07 can most likely be stated as outliers, since the other values are in a higher range above 0.8. This pattern can most likely be explained by a lower accuracy of the statistical model for the autumn months caused by missing or faulty data in the training set. When analyzing the availability of the training data, it was found that only for the inverter 1U07, no data are available for the months November and December.

To further investigate the above detected performance issues by EPI, we introduce key performance indicators “voltage performance index” (VPI) and “current performance index” (CPI) to add information on whether the deviations are related to issues in the voltage or current. These indicators are determined similarly to Equation (3).

### 3.2.3. Relative Performance Index

Although the performance indexes give a valuable indication of performance deviations, these indexes are influenced by the weather conditions and erroneous input data. As can be seen in Figures 9 and 10, a shift in the EPI for most of the MPPTs can be observed between different months. Furthermore, the modeled values of both the statistical and physical models do not incorporate the well-known accelerated aging degradation. That value depends on the module technology and plant location [31] and has been estimated to a median degradation in the power of 0.5% to 0.6% per year [32]. A linear line fitted to the median values \( EPI_{\text{median}} \) from Figure 10 showed a 1.8% degradation per year that can be correlated to power issues, such as aging or other material problems. A degradation trend of the \( EPI_{\text{median}} \) can be also observed in Figure 9, which is a bit disturbed by the high values caused by sensor issues. Analysis of the EPI allows estimating long-term degradations. Since power degradation is seen in the data of all MPP trackers, we would also like to examine the trackers that have lower performance compared to the others.

Consequently, relative indexes are defined. Equations (4)–(6) provide an indicator of the monthly \( (m) \) median values relative to the monthly median of every MPPT of EPI, VPI, and CPI. These indexes represent the percentage deviation from the monthly median for all MPPTs that is taken as a reference.

\[
REPI(m) = \frac{EPI(m) - EPI_{\text{median}}(m)}{EPI_{\text{median}}(m)} \cdot 100 \tag{4}
\]

\[
RCPI(m) = \frac{CPI(m) - CPI_{\text{median}}(m)}{CPI_{\text{median}}(m)} \cdot 100 \tag{5}
\]

\[
RVPI(m) = \frac{VPI(m) - VPI_{\text{median}}(m)}{VPI_{\text{median}}(m)} \cdot 100 \tag{6}
\]

By applying the relative energy performance index (REPI) indicators to both the physical and statistical models, the results for every MPPT of the PV park described above are depicted as boxplots that show the 25th percentile \( (q1) \) and 75th percentile \( (q3) \), together with the median value of the relative performance index. In addition, the outliers outside the interval \( (q1 - 1.5\text{IQR}, q3 + 1.5\text{IQR}) \) are shown. Under ideal circumstances, the indicators should be distributed around 0. Under real-life conditions, the relative indicators will drop below zero when MPPTs show improper performances, worse than the reference median. The results of the distribution of REPI (as depicted in Figure 11) show great deviations from zero at MPPTs “1U02-T1”, “1U04-T1”, “1U09-T2”, and “2U02-T1”. This statement is in good agreement with the detected gradual degradations of the mentioned MPPTs in Figures 9 and 10. Furthermore, statistical outliers for MPPTs “1U03-T1”, “2U01-T1”, and “2U04-T1” were revealed (not shown in Figure 11 for better clarity), which correlates with EPI deviations depicted in Figures 9 and 10. Finally, for the MPPTs of inverter 1U09, a high spread in one of the indexes is seen, which indicates a significant change in the behavior of the inverters for the evaluated period. Consequently, both the revealed evidence of power degradation and the wide REPI distribution are indicators of abnormal PV plant behavior.
By comparing the boxplot distribution of the REPI of the statistical and physical models, a good correlation between the different approaches could be demonstrated for most of the MPPTs (Figure 11), whereas a huge difference can be seen in the distribution of REPI for the physical and statistical models at MPPTs “1U08-T2” and “2U10-T2”. This difference is caused by an investigated initial voltage deviation of these two MPPTs, as depicted exemplarily for MPPT “1U08-T2” in Figure 12a. Due to the fact that the statistical model uses this erroneous voltage data in the training dataset, the modeled MPPT voltage will incorporate the deviation that has already occurred (as can be seen in Figure 12b) and therefore the voltage indicator shows a lower deviation from 0 compared to the physical model. Furthermore, an unusual difference at MPPT “1U09-T1” is observed, which is again caused by erroneous training data used for the statistical models. In this case, the MPPT reaches the expected voltage at the beginning of 2014, but there is a voltage drop in the middle of 2014. Having bad training data for a part of the year caused the statistical model to underestimate the expected voltage in that part of the year, which resulted in performance indexes higher than 1. Considering this effect, a failure is detected at MPPT “1U09-T1”.

![Figure 11](image1.png)

**Figure 11.** Boxplot of REPI for all MPPTs using physical models and statistical models shown in the range (−30%, 30%).

To further restrict our findings to certain failure mechanisms and hence recommendable maintenance actions, an investigation of the operating parameters “U_{MPP}” and “I_{MPP}” was carried out. Hence, the distribution of the relative voltage performance index (RVPI) and relative current performance index (RCPI) for a subset of the most interesting MPPTs is depicted in Figure 13.

![Figure 12](image2.png)

**Figure 12.** Voltage issue MPPT “1U08-T2”: (a) initial voltage deviation; (b) comparison of reference values.
and relative current performance index (RCPI) for a subset of the most interesting MPPTs is depicted in Figure 13.

Figure 13. Boxplots of RVPI (a) and RCPI (b) for a selected subset of MPP trackers calculated using physical models and statistical models.

The results reveal a correlation between the detected power decline of MPPTs “1U02-T1”, “1U04-T1”, “1U09-T1”, “1U09-T2”, and “2U02-T1” and a voltage deviation, as can be seen in the boxplot distribution of RVPI in Figure 13a. In addition, a smaller deviation from zero is seen in the RVPI—distribution of MPPTs “1U01-T1”, “1U06-T1”, and “1U08-T1” (as seen in Figure 13a). Compared to the RVPI values, the median values of the RCPI for all MPPTs are closer to zero, but more outliers are seen here. The outliers most likely indicate short-term, large-scale failures. The degradation observed for “1U03-T1”, “2U01-T1”, and “2U04-T1” can be correlated to a significant loss in current seen in the outliers in Figure 13b. A smaller number of outliers, with values of around −20%, are seen in the REPI and RCPI distribution calculated from the statistical models for many inverters such as “2U06”, “2U07”, “2U08”, “2U09”, and “2U10”. It was discovered that they resulted from a short-term drop in $I_{MPP}$ to near-zero values for about 15 days in mid-2017. Having a loss in many of the inverters for the same month caused an erroneous shift in the median value $CPI_{\text{median}}(m)$. The effect of a biased median value is that some trackers are given a higher relative index of around 10%, seen as outliers in Figure 13b.

Overall, we have demonstrated that by combining physical and statistical reference models for predicting the healthy state of a PV plant, several issues in MPPTs could have been revealed efficiently. A summary of identified issues and a comparison of the sensitivity in detection are given in Table 3.
Table 3. Comparison of model-based monitoring results.

| Suspicous MPPT | Observations | Detectability |
|----------------|--------------|---------------|
|                | EPI          | REPI          | RVPI          | RCPI          | Phys. Model | Stat. Model |
| 1U01-T1        | Deviating    |               |               |               | Voltage deviation confirmed by RVPI |
| 1U02-T1        | Gradual decrease | Deviating    |               |               | Gradual EPI decrease confirmed by REPI and RVPI |
| 1U03-T1        | Temporary decrease | Stat. Outliers |               | Stat. Outliers | Temporary EPI decrease confirmed by REPI and RCPI |
| 1U04-T1        | Gradual decrease | Deviating    |               |               | Gradual EPI decrease confirmed by REPI and RVPI |
| 1U06-T1        | Deviating    |               |               |               | Voltage deviation confirmed by RVPI |
| 1U08-T1        | Deviating    |               |               |               | Voltage deviation confirmed by RVPI |
| 1U08-T2        | Deviating    |               |               |               | Voltage drop detected | Issue due to training data |
| 1U09-T1        | Temporary decrease | Deviating    |               |               | Voltage drop detected | Issue due to training data |
| 1U09-T2        | Gradual decrease | Deviating    |               |               | Gradual EPI decrease confirmed by REPI and RVPI |
| 2U01-T1        | Temporary decrease | Stat. Outliers |               | Stat. Outliers | Temporary EPI decrease confirmed by REPI and RCPI |
| 2U02-T1        | Gradual decrease | Deviating    |               |               | Gradual EPI decrease confirmed by REPI and RVPI |
| 2U04-T1        | Temporary decrease | Stat. Outliers |               | Stat. Outliers | Temporary EPI decrease confirmed by REPI and RCPI |
| 2U10-T2        | Deviating    |               |               |               | Voltage drop detected | Issue due to training data |

3.3. Failure Diagnostic

Our model-based monitoring approach has led us to conclude that by applying modeled reference values and performance indicators, MPPTs “1U03-T1” (Figure 14a), “2U01-T1”, and “2U04-T1” showed abnormal behavior related to outliers of \( I_{MPP} \) values. The MPPT “1U09-T1” (Figure 14b), on the other hand, showed an abnormal drop in performance, caused by a drop in the voltage. Furthermore, MPPTs “1U02-T1”, “1U04-T1”, “1U09-T2” (Figure 14c), and “2U02-T1” showed a gradual power decline correlated to voltage degradation. Finally, a slight deviation in the voltage was detected at MPPTs “1U01-T1”, “1U06-T1” (Figure 14d), and “1U08-T1”.

The additional investigation of relative performance indicators allowed us to determine possible failure mechanisms (e.g., thermal aging or thermomechanical fatigue of PV modules) which can be indicated by a voltage and/or current deviation. The findings of degrading operating parameter \( P_{MPP} \) in combination with \( U_{MPP} \) are also reported in [33] as a possible failure indicator for thermal aging, delamination, or connection issues of PV modules. On the other hand, degrading operating parameter \( P_{MPP} \) in combination with \( I_{MPP} \) is reported in [33] as a possible failure indicator for isolation problems of PV modules.

To further restrict the possible failure mechanisms, failure diagnosis based on the physics of failure was carried out, in order to identify critical components by quantifying the underlying load situations. In order to identify the root cause and critical load situations of critical components, a comprehensive failure potential analysis was performed within the scope of the OptPV4.0 project. As a result, a failure catalogue covering 50 failure mechanisms in PV systems was established and served as an input for the physical damage models. It turned out that failure modes such as delamination, backsheet cracking, cell
cracks, potential-induced degradation (PID), burn marks, disconnections of ribbons, and defect bypass diodes are mostly reported in photovoltaic modules. This confirms the previous findings in the literature cited above.

![Figure 14. Measured P_{MPP} Values together with the EPI calculated using physical and statistical models for MPPTs: (a) “1U03-T1”, (b) “1U09-T1”, (c) “1U09-T2”, and (d) “1U06-T1”.

We address the modeling process by using a specialized software, UPTIME HARVEST™, developed by Uptime-Engineering. This software allows automated periodic calculations of all developed damage models based on available SCADA time series data of the PV plant described above, such as inverter DC and AC electrical parameters. As a result, the software delivers accumulated damage values throughout the entire assessment period of the available SCADA data. In order to rank possible failure modes according to the risk of failure occurrence, the accumulated damage is summarized in an assessment table, as depicted in Figures A1 and A2, which shows the entire calculation result for the whole PV park. For a better understanding, a subset of investigated instances (MPPTs) and damage models (depicted as “Analysis Package”) describing failure modes of PV modules is depicted in Figure 15.

![Figure 15. Damage calculation results for suspicious identified MPPTs.
The assessment table allows comparing damage increments of one failure mode/instance combination per row of all PV system instances. High damage kinetics (highlighted in yellow) for a particular failure mode/instance combination indicate a higher failure risk in comparison to other failure mode/instance combinations with a lower failure risk (highlighted in green). In this way, likely root causes for detected performance issues can be identified. Consequently, this information is a valuable input for field inspection tasks.

By observing the entire assessment table in Figures A1 and A2, it could have been revealed that the PV park was exposed to a high load inhomogeneity, which is depicted by a high variation in the damage values. Several MPPTs of individual inverters were exposed to different loadings, especially at their power electronic devices. Furthermore, we observed a great variation in the load situations between the MPPTs in the PV park with different orientations. In particular, those trackers with a southwest orientation (“1U01-T1”–“2U03-T2”) were exposed to much higher loads.

Further analysis of the above as suspicious indicated gradual performance decreases of MPPTs “1U02-T1”, “1U04-T1”, “1U09-T2”, and “2U02-T1”, by means of the assessment table (Figure 15), was used to observe possible connections between detected failures and the quantified load situation.

The assessment revealed that MPPTs “1U02-T1” and “1U09-T2” had seen higher thermal overload on their DC fuse compared to other instance/failure mode combinations. Further than that, MPPT “1U03-T1” had seen the highest thermal load at the DC fuse. MPPT “1U09-T2” had seen higher thermal aging at electrical contacts and photodegradation and hydrolysis-driven aging at PV modules, which is in good agreement with the reported degrading operating parameter “PMPP” in combination with “UMPP” in [33] as a possible failure indicator for thermal aging, delamination, or connection issues of PV modules. In addition, “1U09-T2” had seen high loads at nearly every power electronic device (as depicted in Figures A1 and A2), such as IGBTs, capacitors, and electrical contacts. These findings are in good agreement with the previously reported usual failure probabilities of PV system components in the literature [23–25].

In contrast to the indicated gradual performance decrease, a correlation between the detected partial decreased performance in MPPTs “1U03-T1”, “1U09-T1”, “2U01-T1”, and “2U04-T1” was revealed. We found that MPPTs “1U03-T1” and “2U01-T1” had seen high damage accumulation by diffusion-driven failure mechanisms. There is satisfactory agreement in [33] between the degrading operating parameter “PMPP” in combination with “IMPP”, which could be a possible failure indicator for isolation problems of PV modules. However, this conclusion needs to be read with caution since we do not know why the temporary power decrease disappeared after the detection. This evidence needs to be further investigated by field analysis.

3.4. Discussion of the Results

The results presented here suggest that most of the power deviations could be detected by both approaches. However, physical models have some limitations in predicting reference values in partial load situations (as revealed by an error analysis in Figure 8) since actual control strategies of inverters in these load situations are not considered. One advantage of the physical models lies in predicting the reference as rated values, which is not dependent on the quality of the training dataset. This allowed, for example, the identification of performance losses and their relation to voltage drops. Furthermore, detected EPI values greater than one by means of physical models allowed us to identify sensor problems in PV systems in an easy manner. In contrast, the statistical models allow the prediction of the reference state in a more detailed way, which enabled the observation of gradual degradation of EPI values. One limitation of the statistical models is that their performance depends on the validity and coverage of the training data. For example, training data with introduced failures, or that do not cover all seasons in a year, can lead to a poor estimation. Nevertheless, the high accuracy and similarity between the performance of the physical and statistical models suggest that both types of models can be used as reference models for online monitoring and revealing major anomalies in PV plants.
The additional investigation of relative performance indicators allowed us to determine the cause of suspicious performance behavior by a voltage and/or current deviation. These observations are valuable inputs for failure diagnostics as failure indicators and limit possible failure root causes. However, since the presented work has only focused on modeling and analyzing failures in the DC system of a PV power plant, the accuracy of prediction monitoring can be further improved by also analyzing the performance of AC components, for instance, the inverter AC outputs, or the degradation of sensor measurements over time.

Finally, to further restrict the possible failure mechanisms, failure diagnosis based on physics of failure was carried out, which allowed us to identify critical components and likely root causes for detected performance issues. Consequently, this information is a valuable input for field inspection and maintenance tasks. Opposite to this, some of the MPPTs identified as suspicious did not show any considerable higher damage sums compared to other strings. This apparent lack of correlation can be explained by unknown failure mechanisms that have not been reported in the field yet and hence not modeled by our approach. Another possible reason for this can be explained by quality issues within the PV power plant, such as cabling or connection errors. Furthermore, shading issues and dirt could also have caused the investigated performance losses. Accordingly, these types of “failures” do not lead to an irreversible change in the microstructure of PV modules subjected to specific load conditions. Hence, they are not seen as failure root causes and were not modeled by the physics of failure approach. This type of failure needs to be validated by field analysis.

4. Conclusions

In this paper, we presented a combined approach of a model-based concept of failure detection and diagnosis to optimize PV power plant operation and maintenance. Therefore, physical and statistical models of a PV system were created, and the physics of failure models of well-known PV failures were implemented. Consequently, a digital representation of a PV power plant as a type of “digital twin” was established.

The calculated reference values by means of physical and statistical models in combination with applied performance indicators allowed simulating the state and operation of the PV system in a holistic manner. Furthermore, the time-related developments of the energy yield were predicted.

The following conclusions can be drawn:

- Strong correlations were observed between the obtained results of statistical and physical models. In contrast, a major source of uncertainty was discovered due to issues in the input data for these models. In particular, physical models showed issues when the sensor data (irradiance or module temperature) were erroneous. The statistical models are unreliable when the training set is not explanatory enough to grasp all of the physical characteristics of the system, or when erroneous output data in the training set are learned within the model. Therefore, using both types of models allows identifying initial PV plant problems, detecting instantaneous failures and gradual power degradation, and, finally, establishing a holistic interpretation of identified errors.

- In order to diagnose the root cause of these failures, the damage model calculation provided possible candidates which have seen a high failure-relevant load and hence a high damage sum. Due to the missing correlation of some detected abnormal behaviors of damage accumulations and the detected performance issues, it can be stated that the system deviation is correlated to quality issues of the PV system or shading or dirt issues, or due to unknown failure mechanisms which have not been reported in the field yet. This statement needs to be further investigated by field analysis and continuous use of model-based analysis. Finally, an actionable maintenance recommendation and hence a corrective action can be generated by limiting possible root
cause candidates and performing field analysis of service technicians on possible symptoms correlated to the failure event.

A promising application of our combined approach of model-based monitoring is the combination with adaptive expert systems. By continuously using model-based monitoring, obtaining feedback from service technicians about the degree of accuracy of the diagnosis and adaption of the domain knowledge, the precision of automated failure diagnosis will increase. An increased knowledge of the failure statistics of components in combination with linearized damage sums will allow estimating the remaining useful lifetimes for application of predictive maintenance strategies. Furthermore, failure mechanisms which have not been reported and modeled yet will be added to the domain knowledge for further failure detection and diagnosis.

In addition, the approaches presented in this work can, in principle, be applied beyond corrective and predictive maintenance strategies towards predictive control strategies. The advantage in such a scenario is that, by predicting the day-ahead performance values with reference models, a contribution to grid stability regulation such as frequency regulation can be evaluated.

Overall, the combined approach on failure detection and diagnosis contributes to faster response times of corrective actions, less downtime, and an efficient long-term operation at maximum performance of photovoltaic power plants.

**Author Contributions:** Conceptualization, V.D., C.G., F.P., W.M., F.L., and T.K.; methodology, validation, investigation, visualization, V.D. and C.G.; writing—original draft preparation: V.D. and C.G.; writing—review and editing, V.D., C.G., F.P., W.M., A.M., F.L., and T.K.; funding acquisition, F.L. and T.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by Austrian Climate and Energy Funds and this study was carried out as part of the Energy Research Program 2018 within the framework of the “OptPV4.0” project (FFG number 871684, Energieforschung (eMISSION), 5. Ausschreibung Energieforschung 2018).

**Institutional Review Board Statement:** Not applicable

**Informed Consent Statement:** Not applicable

**Data Availability Statement:** The data are not publicly available due to data protection reasons.

**Conflicts of Interest:** The authors declare no conflict of interest.

### Appendix A

| Analysis Package | 1U01 - T1 | 1U02 - T1 | 1U03 - T1 | 1U04 - T1 | 1U05 - T1 | 1U06 - T1 | 1U07 - T1 | 1U08 - T1 | 1U09 - T2 | 1U02 - T1 | 1U03 - T1 | 1U04 - T1 | 1U05 - T1 | 1U06 - T1 | 1U07 - T1 | 1U08 - T1 | 1U09 - T2 |
|------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| **PV-Cabling**   | 1.76E+03  | 0.0E+00   | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  | 1.40E+01  | 1.42E+02  |
| **PV-Modules**   | 1.35E+03  | 1.48E+00  | 8.0E+01   | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  | 3.16E+01  |
| **PV-Inverter**  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  | 3.15E+03  |

**Figure A1.** Damage calculation results for MPPTs “1U01-T1”–“1U09-T2”.
### References

1. Schmela, M. Solar Power Europe EU Market Outlook (Report); SolarPower Europe: Brussels, Belgium, 2020.

2. Lazard. Lazard’s Levelized Cost of Energy Analysis (Report); Lazard: Hamilton, UK, 2019.

3. Mühleisen, W.; Hirsch, C.; Brantegger, G.; Neumaier, L.; Spielberger, M.; Sonnleitner, H.; Kubicek, B.; Ujvari, G.; Ebner, R.; Schwark, M.; et al. Scientific and economic comparison of outdoor characterisation methods for photovoltaic power plants. Renew. Energy 2019, 134, 321–329. [CrossRef]

4. OVE EN 61724-1 (2018) Betriebsverhältnisse von Photovoltaik-Systemen, Teil 1: Überwachung; OVE Österreichischer Verband für Elektrotechnik: Wien, Austria, 2018.

5. Zinger, D.S. Review on Methods of Fault Diagnosis in Photovoltaic System Applications. J. Eng. Sci. Technol. Rev. 2019, 12, 53–66.

6. Pei, T.; Hao, X. A Fault Detection Method for Photovoltaic Systems Based on Voltage and Current Observation and Evaluation. Energies 2019, 12, 1712. [CrossRef]

7. Basnet, B.; Chun, H.B.J. An Intelligent Fault Detection Model for Fault Detection in Photovoltaic Systems. J. Sens. 2020, 2020, 6960328. [CrossRef]

8. Jankovec, M.; Galliano, F.; Annigoni, E.; Yu Li, H.; Sculati-Meillaud, F.; Perret-Aebi, L.-E.; Ballif, C.; Topić, M. In-Situ Monitoring of Moisture Ingress in PV Modules Using Digital Humidity Sensors. IEEE J. Photovolt. 2016, 6, 1152–1159. [CrossRef]

9. Grimalcchia, F.; Aghaei, M.; Mussetta, M.; Leva, S.; Quarter, P.B. Planning for PV plant performance monitoring by means of unmanned aerial systems (UAS). Int. J. Energy Environ. Eng. 2015, 6, 47–54. [CrossRef]

10. Pérez Alvarez, G. Real-Time Fault Detection and Diagnosis Using Intelligent Monitoring and Supervision Systems. In Fault Detection, Diagnosis and Prognosis; Márquez, F.P.G., Ed.; IntechOpen: London, UK, 2020.

11. Fischer, K.; Coronado, D. Condition monitoring of wind turbines: State of the art, user experience and recommendations. VGB PowerTech 2015, 2015, 51–56.

12. Ansari, F.; Glawar, R.; Nemeth, T. PriMa: A prescriptive maintenance model for cyber-physical production systems. Int. J. Comput. Integr. Manuf. 2019, 32, 482–503. [CrossRef]

13. Ali, M.H.; Rabhi, A.; El Hajjaji, A.; Tina, G.M. Real Time Fault Detection in Photovoltaic Systems. In Proceedings of the 8th International Conference on Sustainability in Energy and Buildings, Turin, Italy, 11–13 September 2017; pp. 914–923.

14. Villalva, M.G.; Gazoli, J.R.; Filho, E.R. Comprehensive Approach to Modeling and Simulation of Photovoltaic Arrays. IEEE Trans. Power Electron. 2009, 24, 1198–1208. [CrossRef]

15. Saberian, A.; Hizam, H.; Radzi, M.A.M.; Kadir, M.Z.A.A.; Mirzaei, M. Modelling and Prediction of Photovoltaic Power Output Using Artificial Neural Networks. Int. J. Photoenergy 2014, 2014, 469701. [CrossRef]

16. Escobar, L.A.; Meeker, W.Q. A Review of Accelerated Test Models. Stat. Sci. 2006, 21, 552–577. [CrossRef]

17. Alfrāhīm Shannan, N.M.A.; Yahaya, N.Z.; Singh, B. Single-diode model and two-diode model of PV modules: A comparison. In Proceedings of the 2013 IEEE International Conference on Control System, Computing and Engineering, Penang, Malaysia, 29 November–1 December 2013; pp. 210–214.

18. Samara, S.; Natsheh, E. Intelligent PV Panels Fault Diagnosis Method Based on NARX Network and Linguistic Fuzzy Rule-Based Systems. Sustainability 2020, 12, 2011. [CrossRef]
19. Lazzaretti, A.E.; Costa, C.H.D.; Rodrigues, M.P.; Yamada, G.D.; Lexinoski, G.; Moritz, G.L.; Oroski, E.; Goes, R.E.D.; Linhares, R.R.; Stadzisz, P.C.; et al. A Monitoring System for Online Fault Detection and Classification in Photovoltaic Plants. *Sensors* 2020, 20, 4688. [CrossRef]

20. Awad, M.; Khanna, R. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, 1st ed.; Apress: New York, NY, USA, 2015; pp. 1–18.

21. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* 2011, 12, 2825–2830.

22. Sayed, A.; El-Shimy, M.; El-Metwally, M.; Elshahed, M. Reliability, Availability and Maintainability Analysis for Grid-Connected Solar Photovoltaic Systems. *Energies* 2019, 12, 1213. [CrossRef]

23. Fischer, K.; Stalin, T.; Ramberg, H.; Thiringer, T.; Wenske, J.; Karlsson, R. *Investigation of Converter Failure in Wind Turbines (Technical Report)*; Elforsk: Stockholm, Sweden, 2012.

24. Sangwongwanich, A.; Shen, Y.; Chub, A.; Liivik, E.; Vinnikov, D. Mission Profile-based Accelerated Testing of DC-link Capacitors in Photovoltaic Inverters. In Proceedings of the 2019 IEEE Applied Power Electronics Conference and Exposition (APEC), Anaheim, CA, USA, 17–22 March 2019; pp. 2833–2840.

25. Köntges, M.; Oreski, G.; Jahn, U.; Herz, M.; Hacke, P.; Weiß, K.-A. *Assessment of Photovoltaic Module Failures in the Field, IEA PVPS Task 13, Subtask 3*, International Energy Agency: Paris, France, 2017.

26. Wang, H.; Ma, K.; Blaabjerg, F. Design for reliability of power electronic systems. In Proceedings of the IECON 2012—38th Annual Conference on IEEE Industrial Electronics Society, Montreal, QC, Canada, 25–28 October 2012; pp. 33–44.

27. Kaaya, I.; Koehl, M.; Mehilli, A.P.; Cardona Mariano, S.D.; Weiss, K.A. Modeling Outdoor Service Lifetime Prediction of PV Modules: Effects of Combined Climatic Stressors on PV Module Power Degradation. *IEEE J. Photovolt.* 2019, 9, 1105–1112. [CrossRef]

28. Quaschning, V. *Regenerative Energiesysteme*, 7th ed.; Carl Hanser Verlag: Munich, Germany, 2011.

29. Harrou, F.; Sun, Y.; Taghezouti, B.; Ahmed, S.; Hamlati, M.-E. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renew. Energy* 2017, 116, 22–37. [CrossRef]

30. SolarPower Europe. *Operation & Maintenance: Best Practice Guidelines*; Version 4.0; SolarPower Europe: Brussels, Belgium, 2019.

31. Frick, A.; Makrides, G.; Schubert, M.; Schlecht, M.; Georghiou, G.E. Degradation Rate Location Dependency of Photovoltaic Systems. *Energies* 2020, 13, 6751. [CrossRef]

32. Jordan, D.C.; Kurtz, S.R.; VanSant, K.; Newmiller, J. Compendium of photovoltaic degradation rates. *Prog. Photovolt. Res. Appl.* 2016, 24, 978–998. [CrossRef]

33. Köntges, M.; Kurtz, S.; Packard, C.; Jahn, U.; Berger, K.A.; Kato, K. *Performance and Reliability of Photovoltaic Systems (Report)*; International Energy Agency Photovoltaic Power Systems Program; International Energy Agency: Paris, France, 2014.