Submitted to **EPJ Photovoltaics**

Article Type OPEN ACCESS

Analysis of Fault Detection and Defect Categorization in Photovoltaic Inverters for Enhanced Reliability and Efficiency in Large-Scale Solar Energy Systems

Stephanie Malik[1,*], David Daßler[1], Dharm Patel[1], Carola Klute[1], Robert Klengel[1], Andreas Dietrich[2], Kai Kaufmann[3], Carsten Hennig[4], Danny Wehnert[5], and Matthias Ebert[1]

[1] Fraunhofer IMWS, Fraunhofer Institute for Microstructure of Materials and Systems, Walter-Hülse-Str. 1, Halle 06120, Germany
[2] DiSUN, Deutsche Solarservice GmbH, Mielestr. 2, Werder 14542, Germany
[3] DENKweit GmbH, Blücherstr. 26, Halle 06120, Germany
[4] saferay holding GmbH, Rosenthaler Str. 34/35, Berlin 10178, Germany
[5] Leipziger Energiegesellschaft mbH & Co. KG, Burgstraße 1, Leipzig 504109, Germany

Abstract – This study examines the performance and vulnerability of large-scale, grid-connected PV systems in relation to inverter faults that can be assigned to the insulated-gate bipolar transistor component. The work presents an interdisciplinary approach, utilizing methodologies from material science, data analysis, statistics, and machine learning to investigate the causes and effects of these faults. The aim is to investigate the defect mechanisms, to identify repeating problems, and to analyze their impacts for a system portfolio in the size of 64 MWp. A root cause analysis identified the failure pattern through material diagnostics of several power modules from inverters previously installed in the field. Prolonged exposure to high temperatures led to the degradation of the semiconductors, resulting in a breakthrough due to the short-term release of excessive heat. In parallel, an impact analysis was carried out based on historical monitoring data, evaluating the behavior of a faulty inverter during curtailment. Due to the sharp increase in curtailment occurrences, a correlation of the effect was observed across nearly the entire portfolio. Finally, a method combining artificial neural network and clustering was proposed to automatically detect this recurring and propagating error pattern.

Keywords: **inverter failure, material diagnostics, data analysis, fault detection, curtailment, machine learning**

1. Introduction

The optimal integration of renewable energies into the electricity supply network is crucial for reducing greenhouse gas emissions and ensuring a sustainable energy supply. This study focuses on the significant role played by photovoltaic (PV) inverters in connecting solar modules and the public power grid. Beyond converting direct current (DC) into grid-compatible alternating current (AC), these inverters play a crucial role in ensuring the efficient and reliable integration of PV-generated electricity into the grid. In addition, inverters are also responsible for optimizing the power output by adjusting (tracking) the voltage and current, ensuring that the PV system operates at its maximum power point. Failures or malfunctions of inverters can lead to partial or complete shutdowns of PV systems, resulting in substantial financial losses for operators and investors, combined with increasing maintenance and operating costs. To ensure that PV systems can be operated continuously, it is necessary to understand the defect mechanisms of inverter failures and to reduce their extent and impact during operation.

The scale of PV deployment continues to grow, with billions of systems installed globally and an increasing number of large-scale solar farms [1]. Although operation and maintenance (O&M) is supporting operations, in general, the demand for quality and efficiency of these services is struggling to keep pace with this rapid expansion. Particularly in large-scale PV systems, where dozens or hundreds of inverters operate simultaneously, the risks posed by inverter failure are magnified. Under such conditions, an efficient O&M service is critical to mitigate technical risks and to achieve optimal energy output [2].

* Corresponding author: Stephanie.Malik@csp.fraunhofer.de

In utility-scale PV systems, inverters handle significant amounts of power and are designed to maintain high efficiency while coping with variable environmental conditions, such as changes in solar irradiance, temperature, and grid stability. In these systems, grid-connected inverters also have the added responsibility of managing interactions with the grid, including synchronizing frequency and voltage, and providing reactive power support. The importance of utilizing renewable energy from PV or wind, inverter needs to meet the demand for robust and reliable inverter performance. Studies [3, 4, 5] report that around 50% of failures in PV systems are related to inverters. Also, the high number of O&M tickets (43–70%) caused by inverter defects require a considerable amount of time and personnel effort [3]. Inverters are generally considered to have a shorter operational lifetime compared to PV modules [6]. Nowadays, where PV modules are designed to last for over 30 years, inverters may need calibration, replacement, or repair every 10 to 15 years. Additionally, inverters are subject to various internal and external stresses, such as thermal fluctuations, grid influences, as well as environmental conditions, including dust and humidity. Operation under these conditions over a longer period can lead to failures and performance degradation over time. The most common types of inverter failures are overheating, material fatigue of components, control and communication failures, grid disturbance, and electrical overload [7, 8].

Thermal cycling is one of the primary contributors to material aging in power electronic devices. As inverters operate, power transistors such as IGBT frequently switch on and off, generating heat. The cycling heating can cause thermal expansion and contraction in the semiconductor material and surrounding packaging. Over time, the repeating thermal cycling can lead to thermal fatigue, where the materials, such as solder, wire bonds, encapsulating materials, begin to deform, dissolve, degrade or crack. These microscopic changes gradually enhance during continuing operating, may leading increased electrical resistances, reduced thermal conductivity, and eventually component failure. In addition, thermal issues can be accelerated by harsh environment or insufficient cooling. [8, 9]

Environmental factors, such as humidity, snow, or cold temperatures, can lead to moisture ingress into the inverter housing, especially for older devices. Corrosion of metal components and conductive paths compromises the integrity of electrical connections and increases the probability of short circuits or electrical leakage.

An industrial survey explores the importance of further research to fill the gap within the reliability of specific components within power electronic devices. In inverters, such critical components are electrolytic or film capacitors, power semiconductors, transistors, protection devices (overvoltage protection or fuses), and cooling systems. [7]

There is a great need in O&M to detect the deviating standard behavior of inverters in large system portfolios as early as possible. In addition, the aim is to create a greater understanding of the causes of systemic inverter failures. Deeper insights into the behavior of inverters, combined with advanced failure detection techniques, can allow for predictive and preventative maintenance, minimizing unplanned downtime and reducing the need for costly, reactive repairs. As inverter failures represent a significant portion of overall PV system failures, improving failure detection mechanisms can significantly enhance the overall reliability and economical return of investment of large-scale solar installations.

O&M services can consist of several actions: preventive maintenance, such as monitoring and regular inspections, corrective maintenance, for repairing and correcting critical problems, and predictive maintenance, deriving maintenance strategies based on historical operation data [10]. However, managing O&M in large-scale systems presents unique challenges due to the vast number of components, geographical and climatic deviations, and complex variety of system portfolios [11, 12].

Traditional monitoring approaches track and analysis key performance metrics, such as electrical and environmental parameters. When these parameters exceed predefined thresholds, such as string over current, the system flags a fault or triggers a shutdown. This reactive procedure only indicates failures but does not clearly detect or inform the operator in advance. On-site inspections are becoming more and more automated and provide additional measurements for evaluating observations. However, large-scale systems are still costly to thoroughly inspect and require time planning and human resources.

Corrective measures must be taken quickly and cost-effectively to avoid significant economic losses. Typical performance indicators in O&M are MTTD (mean-time-to-detect) and MTTR (mean-time-to-repair), which estimate the approximate time between the occurrence or discovery and its detection or remedy. It is in the interest of every stakeholder to keep these mean times as short as possible. Advanced methods based on monitoring data can provide support here. [10]

Data-driven methods can help to detect anomalies at an early stage, determine the possible root cause of problems, and recommend remedial action efficiently [13, 14, 15, 16]. These methods are applicable to all types of electrical measurements: module, string, and inverter data. In this study, the focus is primarily on inverter data.

Advanced methods range from analytical monitoring to more advanced approaches that leverage signal processing or machine learning. Nevertheless, several gaps remain within inverter diagnostics, particularly for large-scale plants: the scalability of detection systems, early detection and root cause analysis, pattern recognition, and stresses due to environmental impact factors.

This study examines the performance and vulnerability of large-scale, grid-connected PV systems in relation to inverter faults that can be assigned to the IGBT (insulated-gate bipolar transistor) component. The work presents an interdisciplinary approach, utilizing methodologies from material science, data analysis, and statistics to investigate causes and effects of these malfunctions to understand defect mechanisms, identify repeating problems, and analyze their impacts in portfolios. It will consider the topic of severity of inverter failures assigned to the IGBT component from two sides: *Root Cause Analysis* and *Impact Analysis*. The former explores failure causes through material diagnostics, drawing conclusions on potential failure mechanisms based on error messages and material fatigue of defective inverter components from the field. The latter employs pattern recognition in monitoring data – from the defective devices examined in the laboratory – to automatically detect recurring and propagating error patterns as well as to determine the associated operational risks.

2. Material and Methods

2.1. Root Cause Analysis

For this study, several power modules from field inverters were provided, which included both failed and functioning devices, to determine the degree of damage and the actual cause of failure. To proceed as sensitively as possible and to minimize the risk of the damage pattern being impaired by preparation steps, non-destructive characterization methods such as X-ray inspection and scanning acoustic microscopy (SAM) are generally used as a first step. These can already provide initial insights into the level of damage and help to localize the origin. The next step is to expose the damaged area, for example, by opening the housing or removing the encapsulation. In case of overload-related failures of power semiconductors, thermal energy is released to such an extent that often the level of damage no longer allows a specific detailed analysis of the failure root cause. If further analysis seems reasonable and promising, low-artifact preparation methods and subsequent microstructure analyses should be applied. The aim of the investigations is to obtain information on material interactions that have taken place (intrinsic or extrinsic), as well as thermal or mechanical loads that have occurred, the associated degradation mechanisms, and causes of failure. Classical metallographic methods (grinding and polishing using abrasive papers, cloths, and suspensions) and ion beam techniques are used for cross-sectioning. Last named are reduced in their preparation volume but minimize the risk of artefacts being introduced. The range of analytical procedures and methods is also complex and extends from imaging procedures such as light microscopy and scanning electron microscopy (SEM), through structural analysis methods to element identification and mechanical property determination.

2.2. Impact Analysis

2.2.1. Data Pre-Analysis and System Portfolio

This study is grounded in extensive data sets, encompassing monitoring data from large PV systems, with a total capacity of 64.4 MWp and spanning multiple years of operation. These data sets, which include high temporal resolution information on inverter operation, establish a robust foundation for analysis. The initial step involved a rigorous assessment of data quality, employing stringent criteria to ensure the plausibility, completeness, and consistency of the available data.

The investigation focused on 102 sub-systems, each equipped with a central inverter of the same type in the 500 to 700 kWp range, distributed across 3 large PV parks, monitored over a period of more than ten years (from 2012 to 2023) in moderate climate conditions. The 3 PV parks are installed near each other in Germany. The data foundation (see **Table 1**) comprised inverter measurement data and status information, as well as irradiance measured by pyranometers in module plane with a high temporal resolution of one minute. The systems are equipped with crystalline modules. Ambient temperature data are downloaded from the national weather service [17], specifically for a larger regional city.

It was assumed that malfunctions or defects in PV systems slightly change the general structure of the operational data, from minor to severe deviations. Possible yield losses do not have to occur immediately but may arise later. Faults without these changes in operating data are unlikely to contribute to yield losses in the near future. Consequently, inverter faults or error conditions also cause deviations that can be found in operating data. With the help of information and protocols from the operators, error messages from the monitoring, and time series evaluation methods, the inverter data were examined for special behavior.

Table 1. Overview of the available parameters per inverter (area: inverter) and per system (system status and environment).

Area	**Parameter**	**Abbreviation**	**Unit**
inverter	AC reactive power	Q_{AC}	kvar
	AC active power	P_{AC}	kW
	AC current L1	I_{AC} L1	A
	AC current L2	I_{AC} L2	A
	AC current L3	I_{AC} L3	A
	DC voltage	V_{DC}	V
	DC power	P_{DC}	kW
	DC current	I_{DC}	A
	temperature of heatsink of inverter	T_{Matrix}	°C
system status	inverter state PV state system state inverter error curtailment	n/a	n/a
environment	irradiance in plane of array	G_{POA}	W/m²
	sun position	azimuth, elevation	°
	ambient temperature	T_{amb}	°C

2.2.2. Modeling of Regular System Operating Behavior with ANN

The methodology used here to predict a regular system operating behavior of one inverter, in terms of DC and AC power, is an artificial neural network, more precisely feedforward networks or multilayer perceptron (for simplicity, in the following term as ANN).

A feedforward ANN is composed of different layers, which themselves consist of different numbers of nodes (so-called neurons). Each individual neuron is completely connected with all neurons of the adjacent layers. The connections between the neurons (so-called weights) take over more or less the function of storing the structural dependency between given input and output information. Furthermore, an activation function within a neuron transforms the incoming signal into the final output signal. The weights of an untrained ANN are predefined systematically or randomly. To adapt the weights to the data, a training process must be carried out using an initial data set (training data). During the training, the weights are iteratively adjusted by an optimization procedure until a termination criterion is reached. The parameters used for modeling are summarized in **Table 2**.

It is important that the training data already contains all the information needed for later application. This data is made up randomly of 70% of the data to be learned. The remaining data is considered as test data to verify that the training is sufficiently

accurate and still applicable to a comparable data set. The training data was checked in advance regarding data quality. Obvious outliers and malfunctions due to systematic measurement errors or technical failure within the PV system are removed utilizing standard filtering [18]. The training of the ANN model was done with the Keras library in Python 3.10.

The quality of the resulting model was determined for the training and testing period by using the root-mean-square-error (RMSE), see Eq. (1). Here, the error was calculated between predicted ($P_{predicted}$) and measured inverter power ($P_{measured}$), considering each timestamp i in the selected period $\{1, \ldots, N\}$. This equation can be used separately for both DC and AC power.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{measured}(i) - P_{predicted}(i)\right)^2} \qquad (1)$$

The resulting ANN is used as a reference for comparison with actual measured power, to detect deviating behavior. The measured and predicted power is used as further cluster parameters to differentiate between regular and irregular behavior.

2.2.3. Differentiation of Regular and Irregular Operating Behavior using OPTICS Clustering

The clustering method OPTICS (Ordering Points To Identify the Clustering Structure) was utilized in this study [19, 20]. The clustering algorithm was applied using functionalities of the Scikit-learn software library (machine learning functions in the Python programming language) [21], version 1.5.2. OPTICS is based on the density of data points, i.e. the number of neighboring data points within a certain radius and can recognize clusters of any shape. It has the advantage that it is robust against noise, such as statistical outliers, identifies noise points separately and does not require a fixed number of clusters. In comparison to similar methods, such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise), OPTICS can set the appropriate radius (reachability) for each possible cluster automatically to encompass a given minimum set of points.

The aim of using OPTICS in this study was to differentiate between regular and irregular patterns in the behavior of inverters. Clustering methods sort data with similar characteristics into equal groups (so-called “clusters”). As a result, data points with widely differing characteristics are distributed into different clusters. The procedures differ in their approach, e.g. based on the metric used, the number and form of clusters, as well as robustness and scalability. In general, they group the data set holistically and without external targets and prior knowledge, which means that they belong to the area of unsupervised learning. The results of this procedure are the allocation of the data points to different clusters (without prior knowledge of the data structure), the respective data density and amount of data per cluster, the cluster boundaries, the temporal accumulation, and the number of outliers.

Compared to other density-based methods, the method is characterized by its flexibility, which enables it to recognize clusters with different data densities in multi-dimensional space. Three main parameters are used: the surrounding radius (ε), the steepness parameter (ξ), which determines the minimum steepness on the reachability diagram, and the minimum number of reachable points (minPts) required to form a cluster. In OPTICS, different from DBSCAN, the surrounding radius ε is not fixed, but can vary depending on the nature of the data set and can be regulated by an upper limit.

Besides the determination of these parameters, the choice of the distance metric which is used for clustering is also crucial. Within this study the Mahalanobis distance [22] was applied. The difference is that unlike the Euclidean distance, where all variables are considered equally important and independent, the Mahalanobis distance considers the correlations between the variables. Since most of the electrical and environmental parameters are not independent of each other, the Mahalanobis distance was used to consider the correlation within defective and non-defective behavior. It is expected that this will account for the majority of the data, i.e. there is a sufficient large distribution available in the initial data set.

In the present approach, clustering was used twice. The goal of the first step was to distinguish the regular from the irregular behavior of the inverter. In the second step, only the data set identified as noise was clustered again. The parameters associated with the clusters are presented in **Table 2**.

Table 2. Data set and parameters used for ANN modeling and OPTICS clustering.

Step	**Data set**	**Parameter**
ANN modelling	Input: ▪ G_{POA} ▪ T_{amb} ▪ sun position (azimuth, elevation) Output: ▪ P_{DC} ▪ P_{AC} Time frame: 1 year (10 min interval); 2021	Hyper parameter: ▪ Batch size 128 ▪ Activation function: ReLU ▪ Optimiser ADAM ▪ Loss function: Huber ▪ Epochs: 200
1st clustering	G_{POA} sun position (azimuth, elevation) T_{amb} Q_{AC} P_{AC} I_{AC} L1 I_{AC} L2 I_{AC} L3 V_{DC} P_{DC} I_{DC} T_{Matrix} P_{DC} predicted (from ANN) P_{AC} predicted (from ANN) Time frame: 1 year (10 min interval); 2022	metric: Mahalanobis minPts: 3,000 ξ: 0.001
2nd clustering	G_{POA} T_{amb} V_{DC} I_{DC} P_{DC} T_{Matrix} Time frame: noise cluster (-1) of 1st clustering	metric: Mahalanobis minPts: 30 ξ: 0.01

3. Results

The following section presents the results from the two perspectives root cause analysis and impact analysis, using the methods described in the previous chapter.

3.1. Root Cause Analysis

Several inverters were considered in the data analysis. Regarding inverter 57 of PV plant 2, it was possible to establish a correlation down to component level. A power module that was used in inverter 57 was characterized in the laboratory. "**Figure 1" (Fig. 1)** shows the failed power module in comparison to another one after field ageing of inverter 64 of PV plant 2.

power module after field ageing (inverter 64) **power module after failure (inverter 57)**

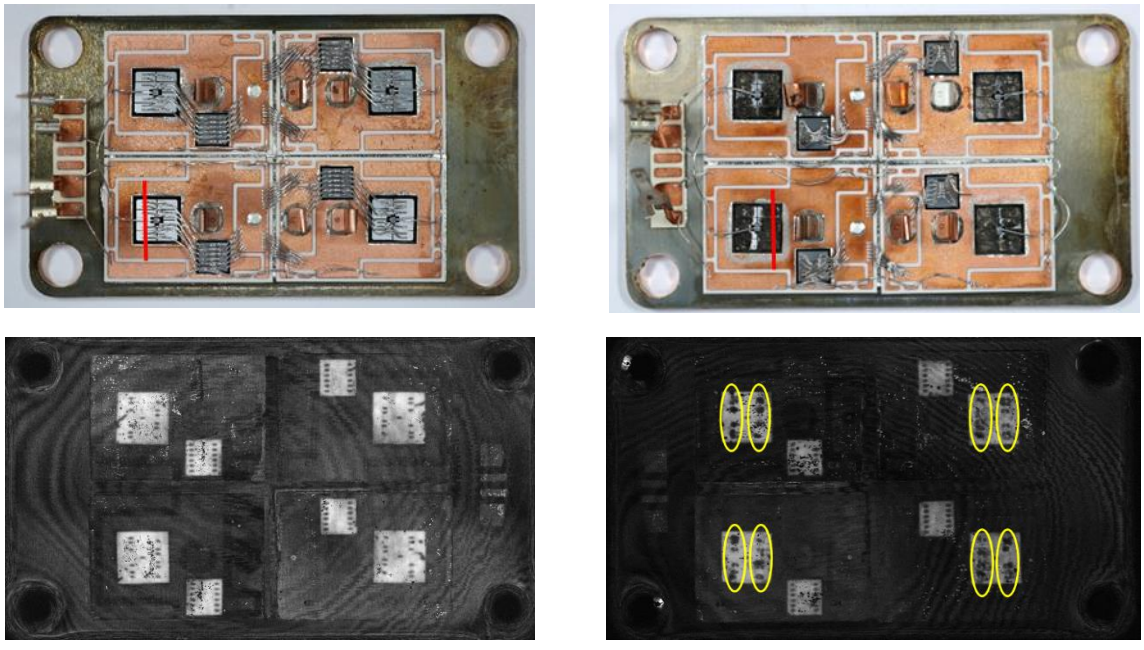

Fig. 1. Characterization of failure mechanisms in inverter components, top: power modules without housing and encapsulation, bottom: non-destructive SAM analysis (wire bond plane) before removal of housing and encapsulation

Even the non-destructive examination of the modules using SAM before opening the housing and removing the silicone gel encapsulation shows clear differences between the failure and reference module and provides indications of the damage pattern. Impairments can be seen in the plane of the chip-sided wire bond connections (see "**Figure 1**" (**Fig. 1**) bottom line). Changes can also be seen in the chip soldering. After opening the housing and chemically removing the silicone gel encapsulation, the differences are more obvious (see "**Figure 2**" (**Fig. 2**)). The emitter metallization of the IGBT components is completely molten and the associated wire bond connections to the diode chip are also destroyed. In opposite, the metallization of the gate pad as well as the gate wire bond itself are still intact. Furthermore, re-melting of the solder connections of the chips and the pins of the connection terminals can be seen. This suggests that the enormous temperature that occurred was dissipated very quickly through the semiconductor and via the chip-solder into the copper metallization of the DCB.

power module after field ageing (inverter 64) **power module after failure (inverter 57)**

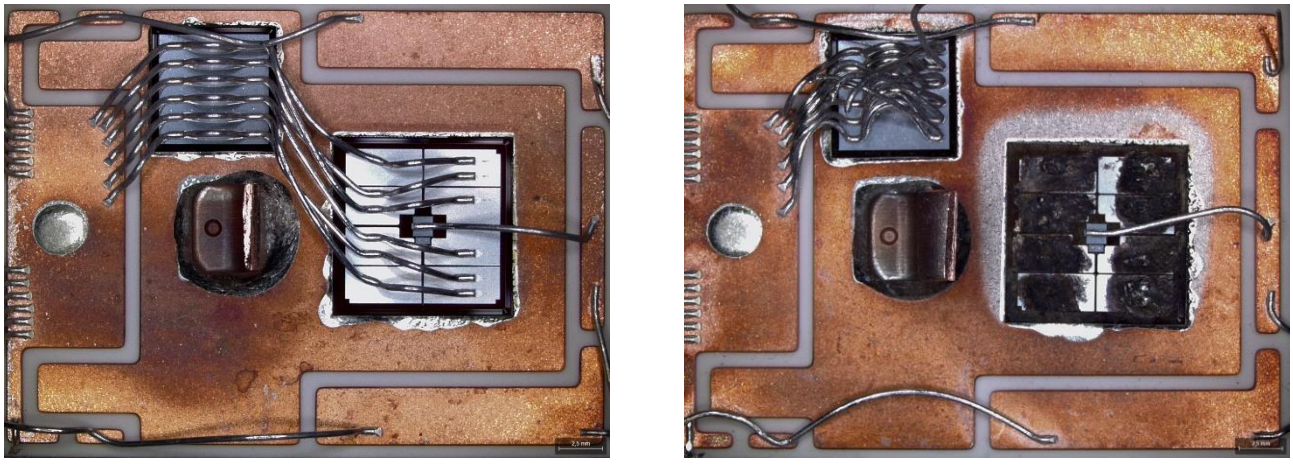

Fig. 2. Overview of the IGBTs shows the changes at emitter metallization and wire bond connections after failure

Following the optical inspection of the components, the microstructure analysis was carried out using cross-section preparation of one IGBT per module. The analyzed positions are marked on the respective IGBTs of the modules in "**Figure 1**" (**Fig. 1**) (red lines positioned over the lower left IGBTs). After cross-sectioning, both IGBTs can be seen comparatively with light microscope images in "**Figure 3**" (**Fig. 3**). For the failed module (inverter 57), the destruction of the IGBT can be seen in the cross-section. The complete top-side metallization is molten and as expected, the silicon material too. The light microscope image of the field-aged module (inverter 64) showed no significant conspicuous features.

power module after field ageing (inverter 64) **power module after failure (inverter 57)**

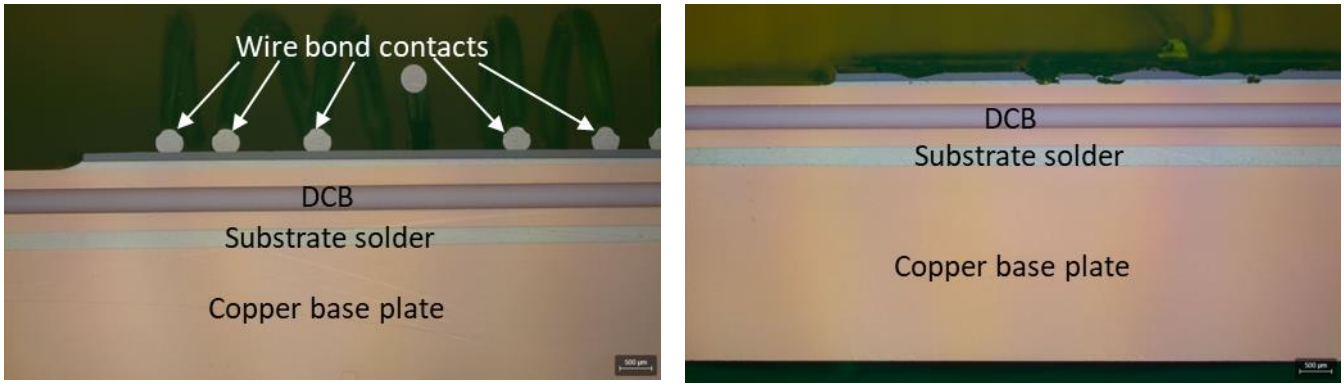

Fig. 3. Cross sectional view of the bonding wires and solder contact using light microscope imaging of the IGBTs

The destruction of the chip layer was already visible by means of non-destructive analysis with SAM and in the following optical inspection (see "**Figure 1**" (**Fig. 1**)). In "**Figure 4**" (**Fig. 4**), the massive damage in the chip area of the failed module (inverter 57) is shown in cross-section using SEM. The wire bond contacts are destroyed and not visible, and the partially missing chip and solder area up to the DCB substrate are also recognizable. Due to the damage, the module was exposed to enormous heat, causing the melting of chip and solder materials locally. The element distribution analysis (see "**Figure 4**" (**Fig. 4**) – right image with yellow frame) shows a material mixing of silicon (IGBT), aluminium (chip metallization), tin (solder), and silver (solder) within the melting zone. The copper of the DCB is located below the solder.

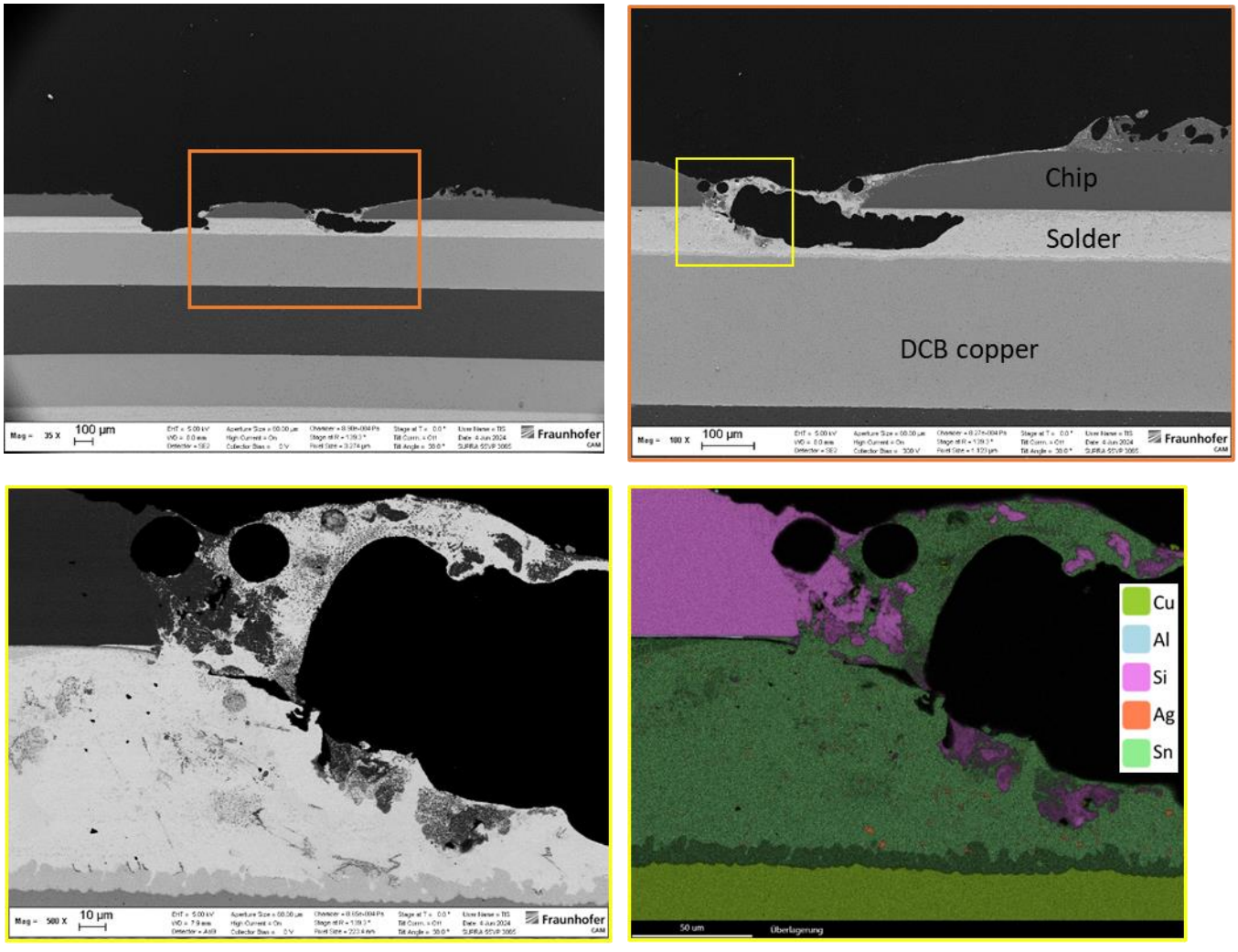

Fig. 4. Cross-sectional view of the melting area using SEM imaging and EDX for elemental analysis of the IGBTs from the power module after failure

The failed module showed that all four IGBT components were equally affected, but the associated diodes showed no damage, and the peripheral extent of damage was less drastic overall. It can be assumed that prolonged temperature impact during the application led to the degradation of the semiconductors and, consequently, to an increasing overload. In the end, a breakthrough occurs, and the final damage is caused by the short-term release of massive heat, which also leads to localized melting. The temperature overload can also be seen in the SEM images of the microstructure analysis.

"**Figure 5**" (**Fig. 5**) shows the cross-section comparison using SEM. In both modules, the solder layer shows small cracks at the edges of the die-side contacts, which may indicate thermomechanical stress cycles during the time the module was in operation. Both modules were active in the inverter for approximately 12 years, showing cracks in comparable levels. These small defects are not in a critical condition that would affect the reliability of the IGBT component. The intermetallic phase growth at the interfaces to the chip backside metallization and the copper metallization of the DCB also show no significant differences, which also indicates a comparable thermal load during the operating time. The solder layer is significantly thicker on the field- aged module (inverter 64) than on the failed module (inverter 57). This difference is due to production-related reasons and not to signs of ageing.

power module after field ageing (inverter 64) **power module after failure (inverter 57)**

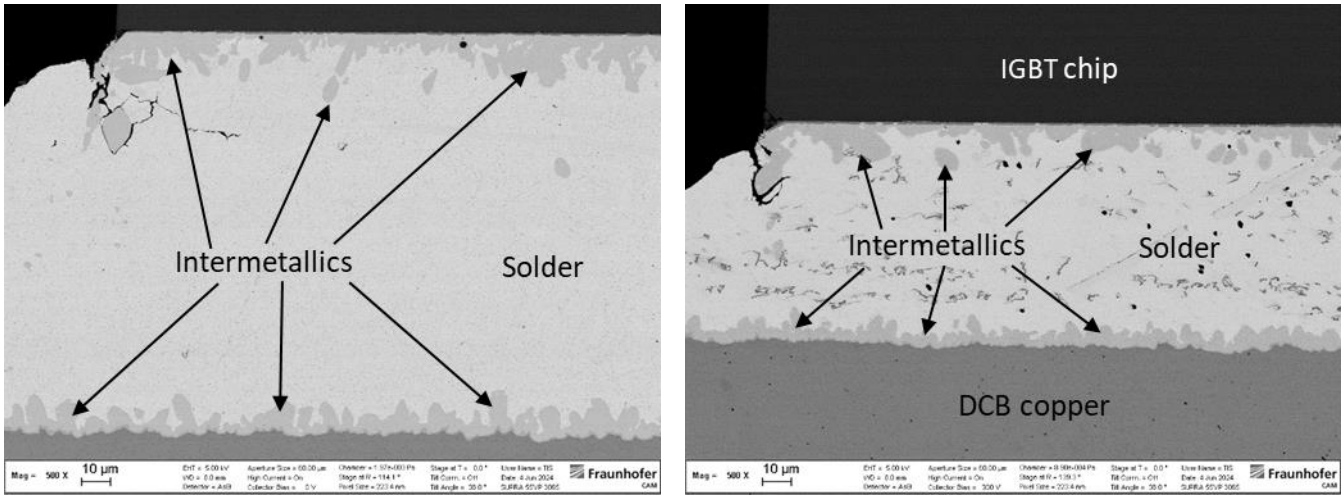

Fig. 5. Cross-sectional view of the solder contact using SEM imaging of the IGBTs

The analysis of the field-aged power module (#01726) from inverter 64 showed no critical ageing damage. Two further aged modules from the same inverter with field ageing (approx. 12 years) were inspected using SAM and then visually examined on the opened module to verify the condition. The SAM overview images of the modules, as well as optical images after opening the housing and removing the silicone gel encapsulation, can be seen in "**Figure 6**" (**Fig. 6**). No apparent aging damage was found on either module, as was already the case with the previous analysed field-aged module (#01726).

power module after field ageing (inverter 64 - #01729) **power module after field ageing (inverter 64 - #01730)**

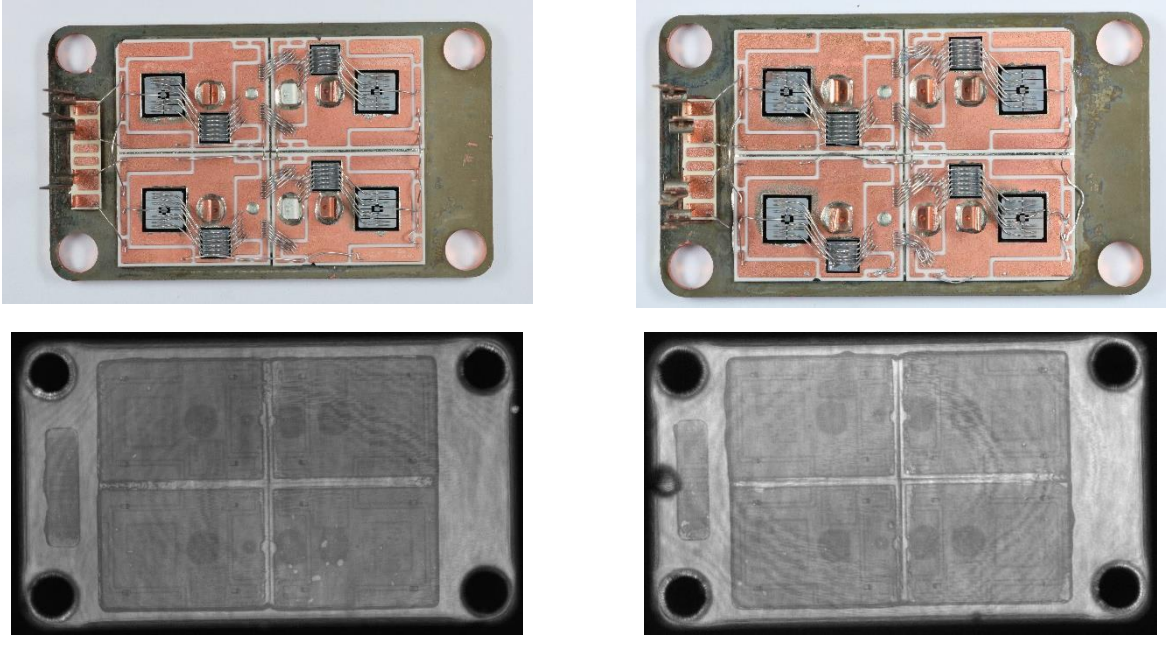

Fig. 6. Characterization of failure mechanisms in inverter components, top: power modules without housing and encapsulation, bottom: non-destructive SAM analysis (overview module) before removal of housing and encapsulation

3.2. Impact Analysis

3.2.1. Pattern of Deviating Inverter Behavior

Based on data pre-evaluation, an unusual pattern showed irregular inverter behavior, as displayed in "**Figure 7**" (**Fig. 7**). The characteristic of this pattern is that during a grid operator-induced curtailment to 0%, the inverter switches multiple times between different operating states, as seen in the parameter "PV state". The inverter would normally switch to a specific "PV state" and remain in this state until the curtailment is over (see inverter 8). Additionally, the AC currents of inverter 7 fluctuate between approx. 90 A and 0 A and the AC reactive power as well in small ranges. After the curtailment ended, the inverter showed an IGBT switching error message (see parameter "Inverter error").

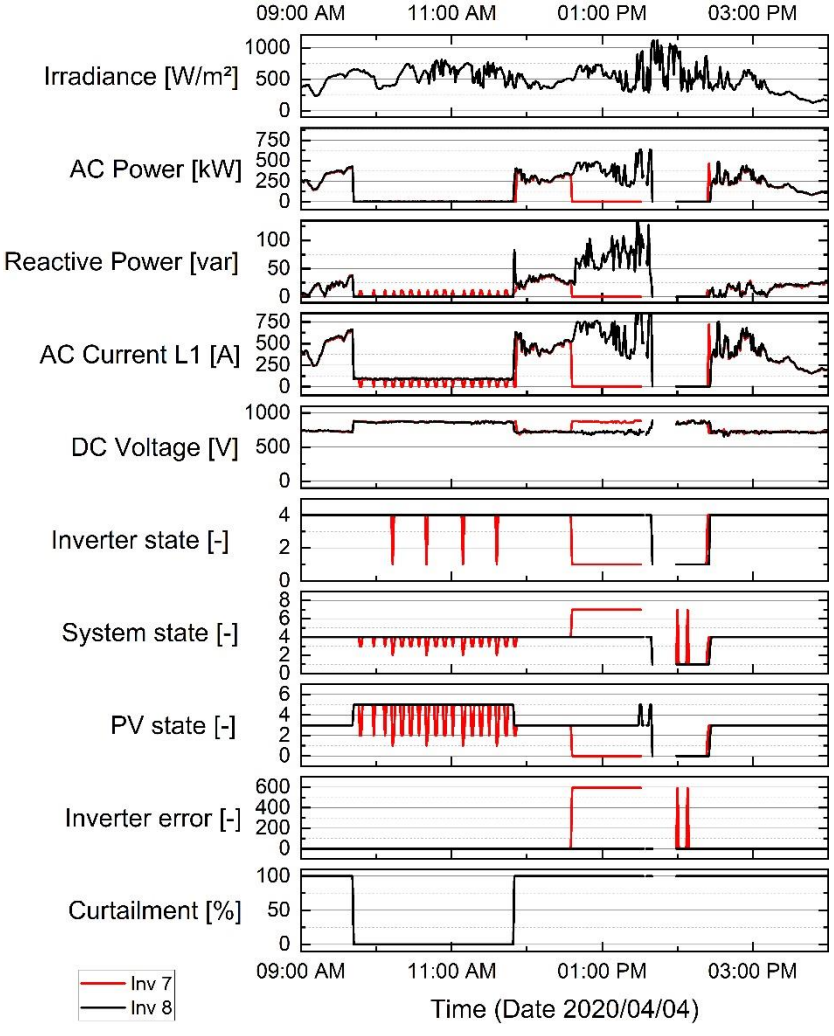

Fig. 7. Comparison of individual electrical measured variables and status information of inverter 7 (Inv 7) and a comparable inverter 8 (Inv 8) in the same PV plant 1, on 04.04.2022.

3.2.2. Impact on System Portfolio

To estimate the influence of the deviating inverter behavior on the entire system portfolio, various aspects were investigated.

1. How often does curtailment take place?
2. How many inverters show the observed pattern?
3. Do faults which point to the IGBT module occur in a temporal relationship to the curtailment?

To answer the first question, the states of the curtailment were summed up for each PV plant with a time resolution of one minute. "**Figure 8**" (**Fig. 8**) (left) shows the number of curtailments determined for PV plant 2, subdivided by month and year. In addition, a further distinction was made here between curtailment due to the grid operator and an economically motivated direct marketing. As can be seen in the diagram, curtailment does not play a major role until the beginning of 2021 but occurs very frequently from this point onwards. Furthermore, it can be clearly seen that the PV plant was very rarely shut down for economic reasons. The observations are also similar for PV plant 1 and PV plant 3, which can be found in the appendix in "**Figure 13**" (**Fig. 13**) and "**Figure 14**" (**Fig. 14**) of this article.

The second question, how many inverters show this irregular behavior, can be answered by checking the "PV state", which varies during curtailment. A moving average with a window of two consecutive time steps was determined and used for this purpose. Whenever this number does not correspond to the desired status, there is a deviating behavior. If a deviating behavior occurs, these time steps are summed up. The resulting number, aggregated per month and year, is demonstrated in "**Figure 8**" (**Fig. 8**) (right) for every inverter in PV plant 2. Correlated to the occurrence of curtailment, this behavior also appears very seldom until 2021 and only with a small number of inverters. From this point onwards, however, almost all systems follow this behavior. The same observation also applies to the inverters in PV plant 1 and PV plant 3, the diagrams of these are included in the appendix in "**Figure 13**" (**Fig. 13**) and "**Figure 14**" (**Fig. 14**).

This leads to the third question: how often error messages relate to the IGBTs occur during and after curtailment in this data set. To check this, 7-days periods, starting with the day of curtailment, were separated from the entire data set for all curtailment situations that occurred, and the error messages that appeared in these periods were summed up. The period of 7 days was chosen to capture as many different weather situations as possible, in which the systems would perform differently.

All error messages occurring in these periods are shown in "**Figure 9**" (**Fig. 9**), with the corresponding short description. The error messages that appear in the list "no assignment possible" cannot be mapped to a description in the manufacturer's documentation, which makes it impossible to determine the cause. The number of error messages relating to grid quality or grid faults is remarkable and can be seen across all inverters, but these are not associated to the curtailment or IGBT failures. These error messages are self-canceling as soon as the grid quality is restored, but they prevent the inverters from feeding into the grid. Other messages are, for example, notifications when shutdowns have occurred due to maintenance work or faults in data transmission and software. The error messages that point to the IGBT component are of greater interest here like: "IGBT overcurrent", "IGBT switch-on error", "Overtemperature error". As can be seen in the diagram, these errors occur in some inverters, whereby even small numbers of occurrences can mean critical states. The same overview has also been created for the inverters in PV plant 1 and PV plant 3 and is documented again in the appendix in "**Figure 15**" (**Fig. 15**) and "**Figure 16**" (**Fig. 16**). The answers to these three questions, based on the presented and explained diagrams and their correlations, show a clear relevance for the entire system portfolio.

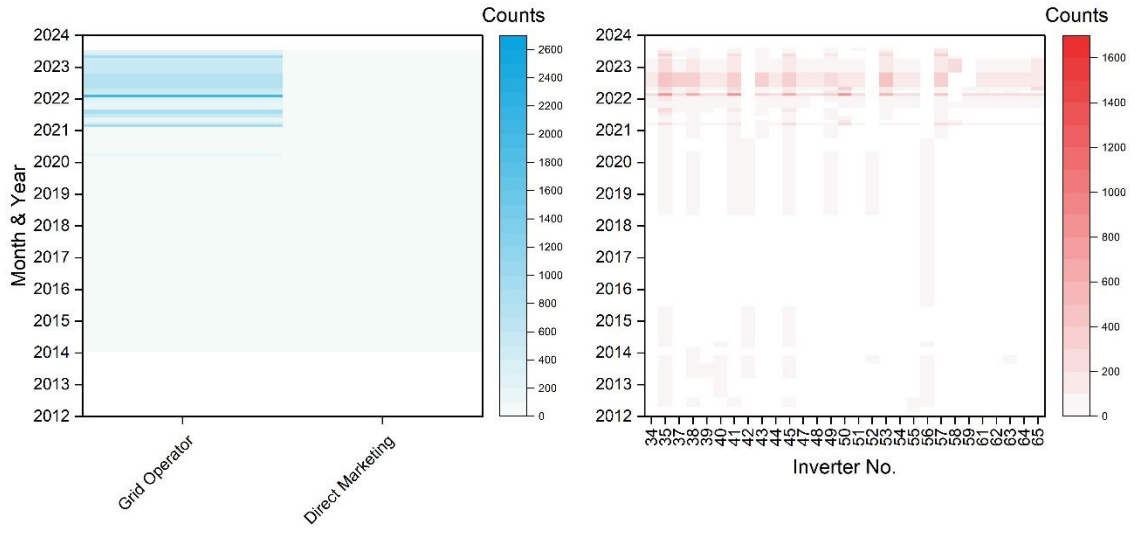

Fig. 8. Correlation of curtailments-related influences on inverters in PV plant 2. Counts are determined based on data set with resolution of 1 min. **Left:** differentiation between the causes of curtailment and their monthly counts. **Right:** detected frequent status changes of inverter in a short period of time after curtailment.

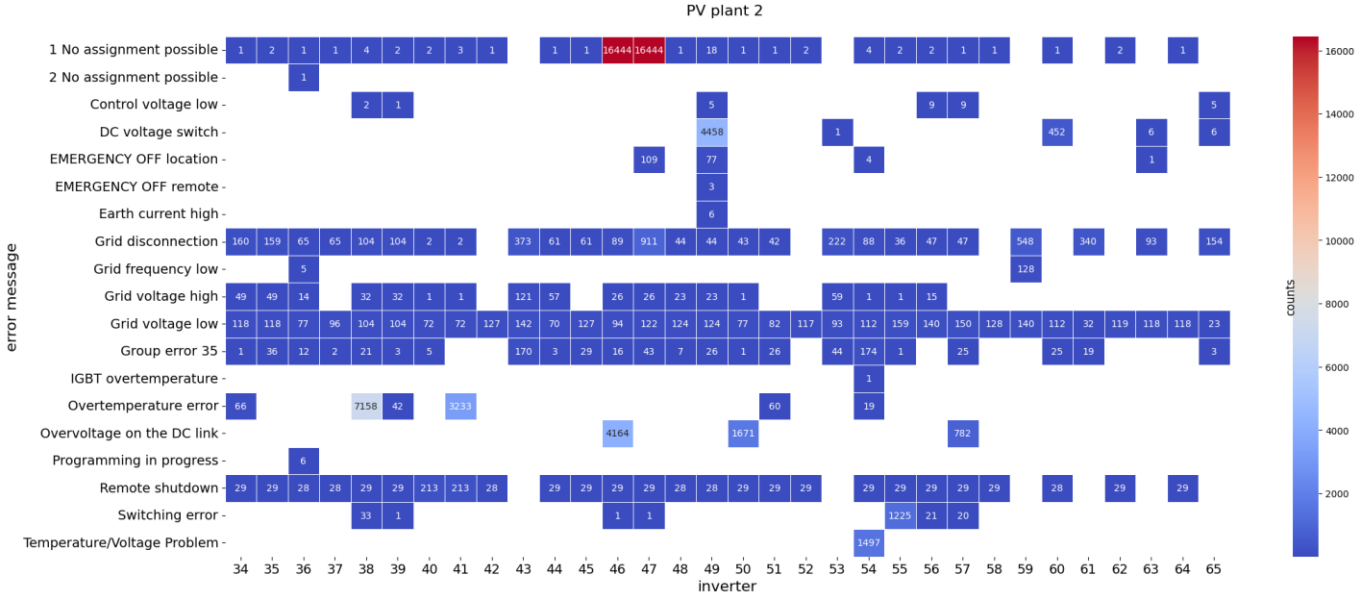

Fig. 9. Number of error messages occurring after curtailment in PV plant 2.

3.2.3. Detection of Pattern in Monitoring Data

The results of the different applied machine learning methods, described in the chapter "Material and Methods", are shown below using data from an exemplary inverter from PV plant 2 for two consecutive years.

In the first step, the normal operating behavior of the inverter was modelled using an ANN. The data set and the associated parameter settings are shown in **Table 2**. "**Figure 10**" (**Fig. 10**) shows the ratio of measured to predicted power for DC and AC

for the period used for modeling. The certainty in terms of RMSE is for the training set 15.5 kW and for the testing set 15.8 kW. The model can now be used to determine the expected output for the following year, using the ambient conditions as input.

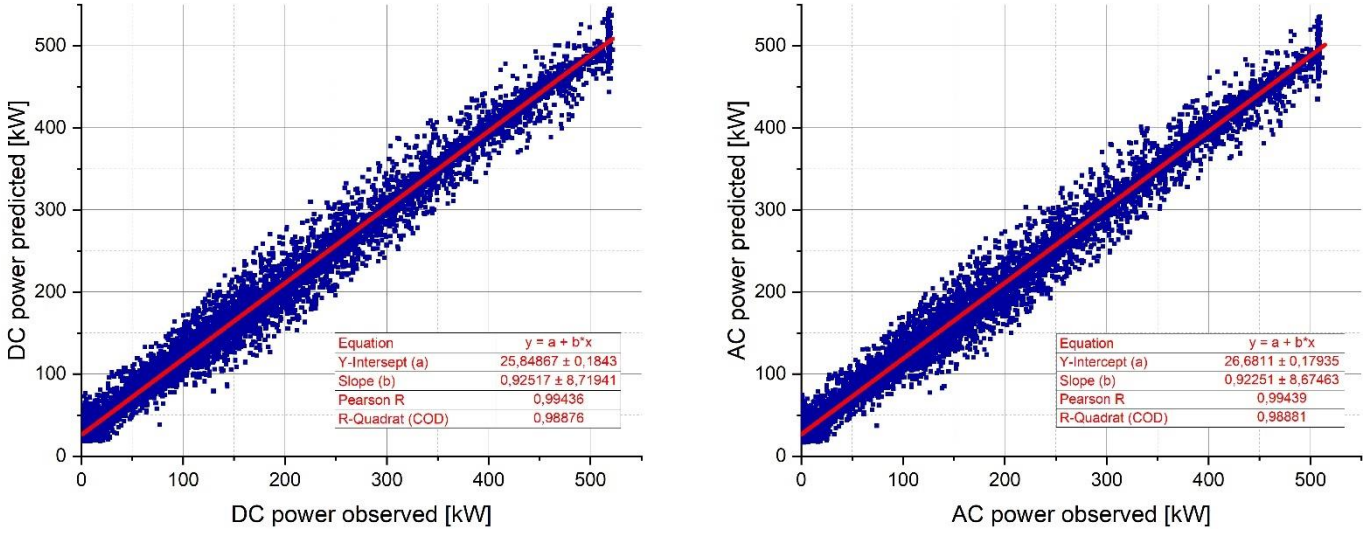

Fig. 10. Training of ANN. Comparison between measured (observed) and predicted power of year 2021.
Left: DC power training accuracy. **Right:** AC power of training accuracy.

This modeled performance can now be included as a reference in the cluster dataset to influence the formation of the clusters. This leads to the results of the first clustering. The parameters and settings used are listed in **Table 2**. The results of the first cluster step are illustrated in "**Figure 11**" (**Fig. 11**). The corresponding clusters for the ratio of the measured to the predicted power are colored in gray and blue. The blue cluster represents the normal operating behavior, and the points marked in gray represent deviating behavior (outlier). Among other things, the two boundary areas are more noticeable with outliers, for measured power levels around zero and for power levels above 500 kW. The deviations of the predicted outputs in the power range above 500 kW are because these high outputs were not available in the year of training and an extrapolation for the model is not possible. This shows the limits of the ANN quite clearly.

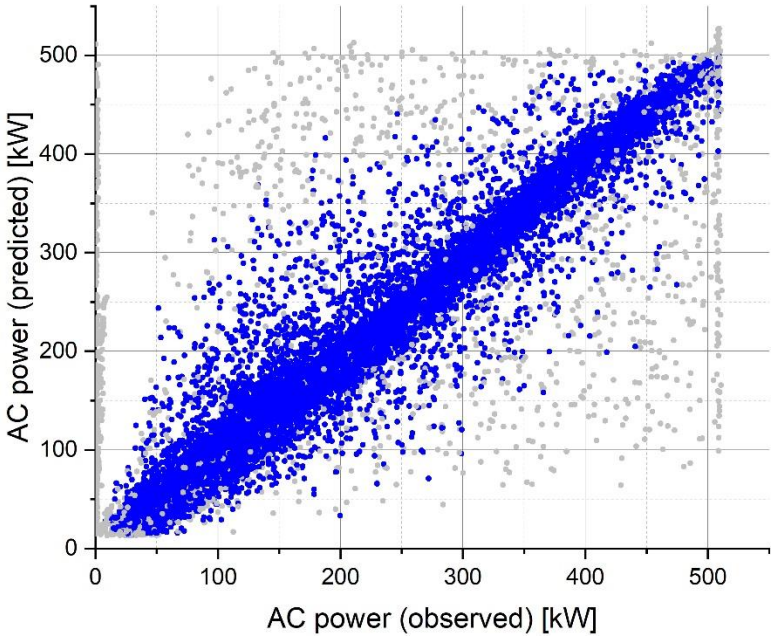

Fig. 11. Result of 1st clustering. **Blue:** Main cluster representing normal inverter behavior (89% of total data set). **Gray:** Irregular behavior or outliers (11% of total data set).

To further differentiate the deviating behavior, only the outliers are now included as a data set in the second cluster step. In addition, the clustering parameters are further reduced. An overview of the parameters and settings is provided in **Table 2**. The results of the 2nd clustering are shown in "**Figure 12**" (**Fig. 12**).

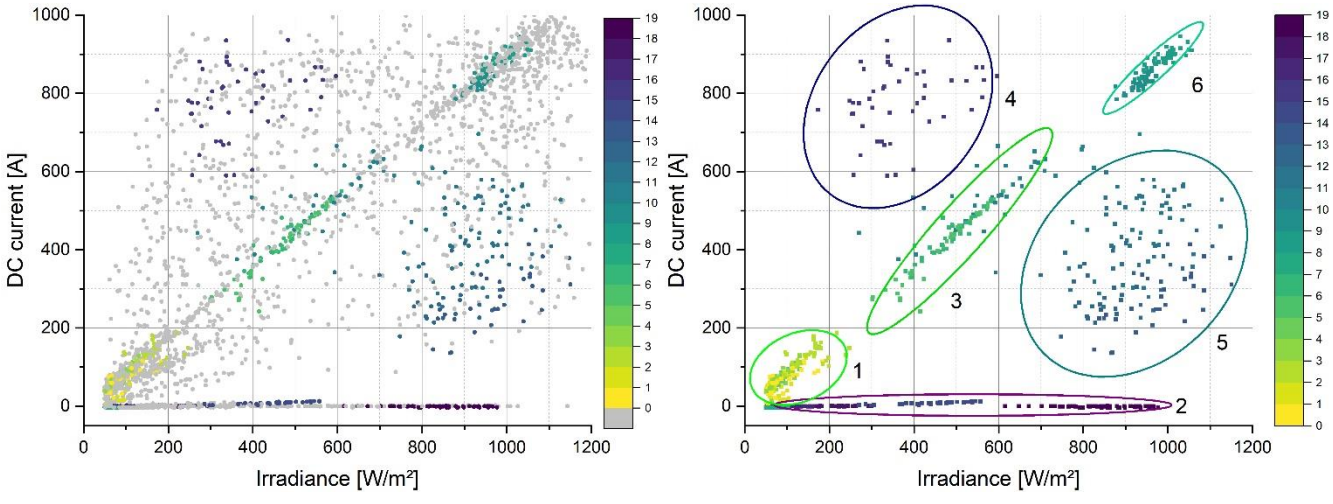

Fig. 12. Result of 2nd clustering.
Left: Entire data set including outliers (gray). **Right:** Same as left but excluding outliers and defined cluster areas.

The clusters form different separated areas: 1, 2, 3, 4, 5, 6. The formation of the cluster in area 2 is particularly important here, as it is related to defects or curtailment. These individual cluster areas can be explained using specific visualizations, ideally in 2 or 3 dimensions. Such helpful diagrams are collected in "**Figure 17**" (**Fig. 17**), in Appendix. The results from the 2nd clustering obtained in this study, based on the systems evaluated, can be described as follows:

Area 1 (consists of cluster 0, 1, 2, 3, 4): The deviations occur primarily in winter or spring, accompanied by low irradiance values. The data occurs more frequently towards midday, which means that this could be due to a possible performance reduction caused by partial shading or snow. The ambient temperatures are below 5°C (except cluster 0). T_{Matrix} is different for each cluster: 15–35°C (clusters 0 and 2), 37–47°C (3), and 50–57°C (4). Cluster 4 is in the normal T_{Matrix} operating range.

Area 2 (Cluster 7, 14, 16, 17, 18, 19): Shutdowns due to defects or derating, visible in DC current and DC power at zero, even though sufficient irradiance would be available. The high number of clusters is due to different voltage levels that lie in the open-circuit voltage range (> 650 V). Cluster 19 is also separated due to high ambient temperature of > 35°C.

Area 3 (Cluster 5, 6, 10): The clusters occur on a few days, more often in the morning and midday hours. The ambient temperature is below 5°C, but the T_{Matrix} is normal 50–60°C. Furthermore, these clusters show less scattering for example compared to area 1. A separated voltage range is appearing: 610–640 V. To summarise, these clusters can be classified as normal operation at cold temperatures and low irradiances.

Area 4 (Cluster 15): This cluster occurs mainly in the summer mornings, at ambient temperatures between (15–30°C) and normal matrix temperature (45–60°C). The voltage range is 525–580 V and the power is high, despite moderate irradiance values. There could be higher weather dynamics here, e.g. passing clouds.

Area 5 (Cluster 11, 12, 13): These clusters are similar to area 4, the ambient temperature is in the range of (10–30°C), with a normal temperature-voltage behaviour (the values are on the voltage-temperature regression line). The matrix temperature is in the normal range of (50–60°C), with a lower voltage range of (480–580 V). The main cluster assessment would be: weather-related deviations under normal behaviour.

Area 6 (Cluster 8, 9): These clusters occur at high irradiance levels at midday, with warm ambient temperatures 25–35°C (difference: cluster 9 > 30°C, cluster 8 < 30°C). The voltage values are lower, whereas the matrix temperature is in a normal range (52–58°C). These are special cases that occur during normal operating behaviour at high ambient temperatures and low operating voltages.

As illustrated by the detailed assessments of the clusters, clustering is not a self-sustaining process; rather, the subsequent interpretation is crucial, particularly during method development. This evaluation aids in distinguishing between critical and non-critical states. By defining parameters within suitable frameworks – even in a multidimensional context – the search for these specific boundaries in the next clustered dataset can be subsequently automated.

4. Discussion

The results of the two-sided analysis fit together and complement each other. The assumed mechanism, based on material diagnostics, is that a prolonged temperature impact during the application led to the degradation of the semiconductors and thus to an increasing overload. Moreover, on the data analysis side, a pattern of deviating inverter behavior that causes the inverter to enter in undefined states. The crucial question now is how this behavior of the inverter can be corrected. This deviation has an enormous impact on the entire system portfolio in the long or short term. The curtailment itself cannot be avoided. However, where should the error be found? Is it in the inverter software or in a faulty signal transmission? In terms of intervening in the inverter software, it is not possible because the product is no longer being developed. Regarding signal transmission, there are ten power modules mounted per phase, and the control signal is looped through all of them. The greatest deviations would be expected at the last segment in the chain. Unfortunately, there are limits to the data analysis here, as no high-resolution information is available on the functional groups involved. The available results only manage to narrow this problem but cannot solve it. Possible corrective measures must be evaluated and implemented in the PV plant itself.

Acknowledgments

No further acknowledgement.

Funding

This research was funded in the "robStROM" project by the Federal Ministry for Economic Affairs and Climate Action in Germany as part of the 7th Energy Research Program (funding code: 03EE1163B).

Conflicts of interest

Stephanie Malik, David. Daßler, Dharm Patel, Carola Klute, Robert Klengel, Andreas Dietrich, Kai Kaufmann and Matthias Ebert has received funding in the project "robStROM" from the Federal Ministry for Economic Affairs and Climate Action in Germany as part of the 7th Energy Research Program (funding code: 03EE1163B).

The institutions of Deutsche Solarservice GmbH, Fraunhofer Institute for Microstructure of Materials and Systems and DENKweit GmbH have received funding in the project "robStROM" from the Federal Ministry for Economic Affairs and Climate Action in Germany as part of the 7th Energy Research Program (funding code: 03EE1163B).

All authors certifies that they have no financial conflicts of interest (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangements, etc.) in connection with this article.

All monitoring data from the photovoltaic systems used in this project are anonymized in the sense of location, owner, etc.

The typus of the inverter for the root cause analysis is also anonymized in this paper.

Data availability statement

Data associated with this article cannot be disclosed due to legal reason.

Author contribution statement

Conceptualization:	Andreas Dietrich, Carsten Hennig, Danny Wehnert
Methodology:	David Daßler, Stephanie Malik, Carola Klute, Robert Klengel
Software:	Kai Kaufmann, David Daßler, Stephanie Malik, Dharm Patel
Validation:	David Daßler, Stephanie Malik
Data Curation:	Kai Kaufmann, Dharm Patel
Writing – Original Draft Preparation:	Stephanie Malik, David Daßler, Carola Klute, Robert Klengel
Writing – Review & Editing:	Dharm Patel, Matthias Ebert
Visualization:	Carola Klute, Robert Klengel, David Daßler
Supervision:	Andreas Dietrich, Kai Kaufmann, Carsten Hennig, Danny Wehnert, Matthias Ebert
Project Administration:	Andreas Dietrich

Appendix

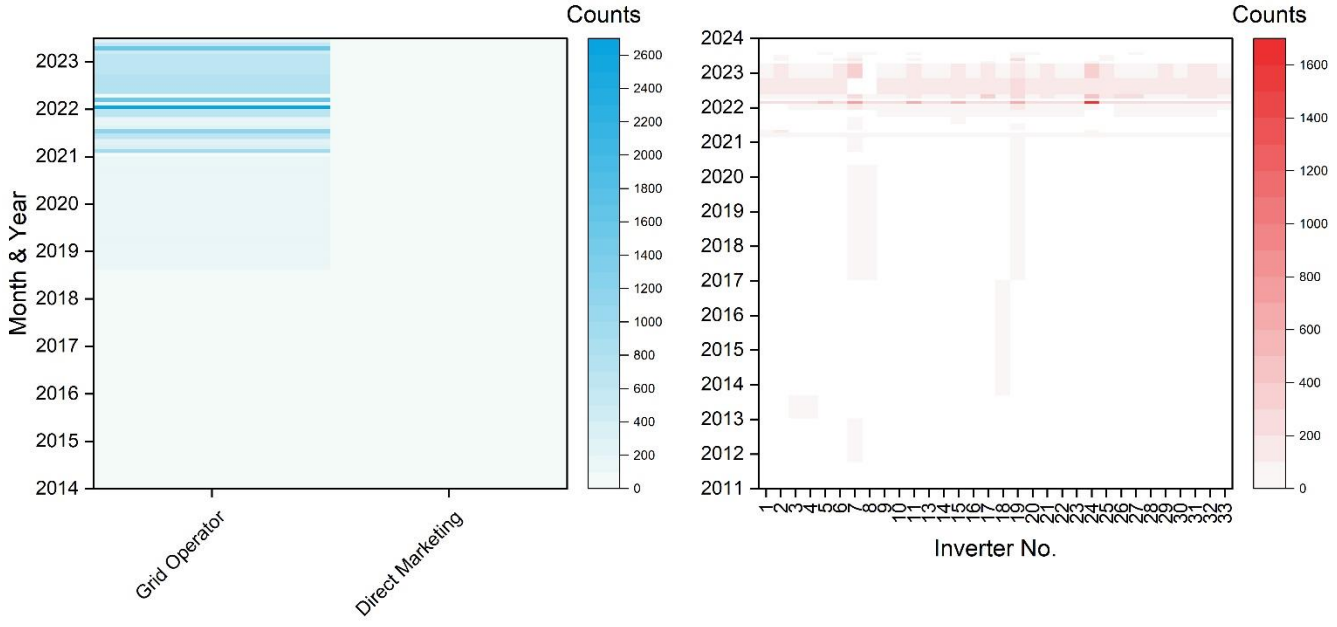

Fig. 13. Comparison for 2nd set of inverters (PV plant 1): correlation of curtailments-related influences on inverters. **Left:** differentiation between the causes of curtailment and their monthly counts. **Right:** detected frequent status changes of inverter in a short period of time after curtailment.

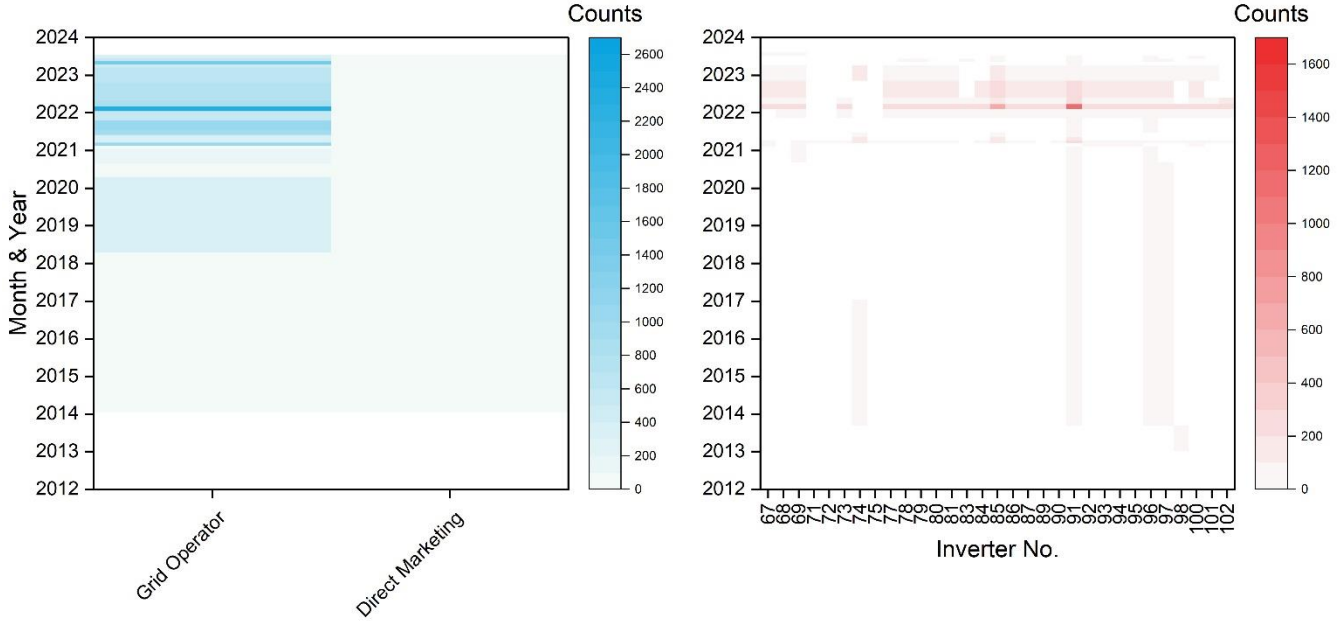

Fig. 14. Comparison for 3rd set of inverters (PV plant 3) : correlation of curtailments-related influences on inverters. **Left:** differentiation between the causes of curtailment and their monthly counts. **Right:** detected frequent status changes of inverter in a short period of time after curtailment.

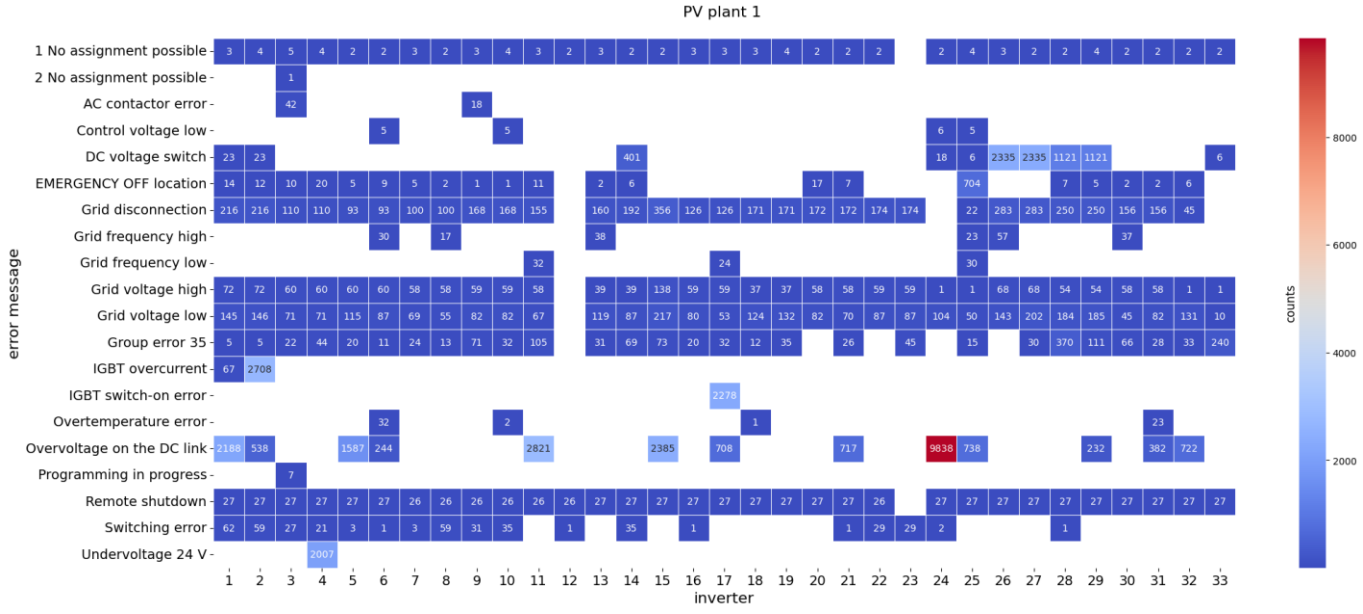

Fig. 15. Number of error messages occurring after curtailment in PV plant 1.

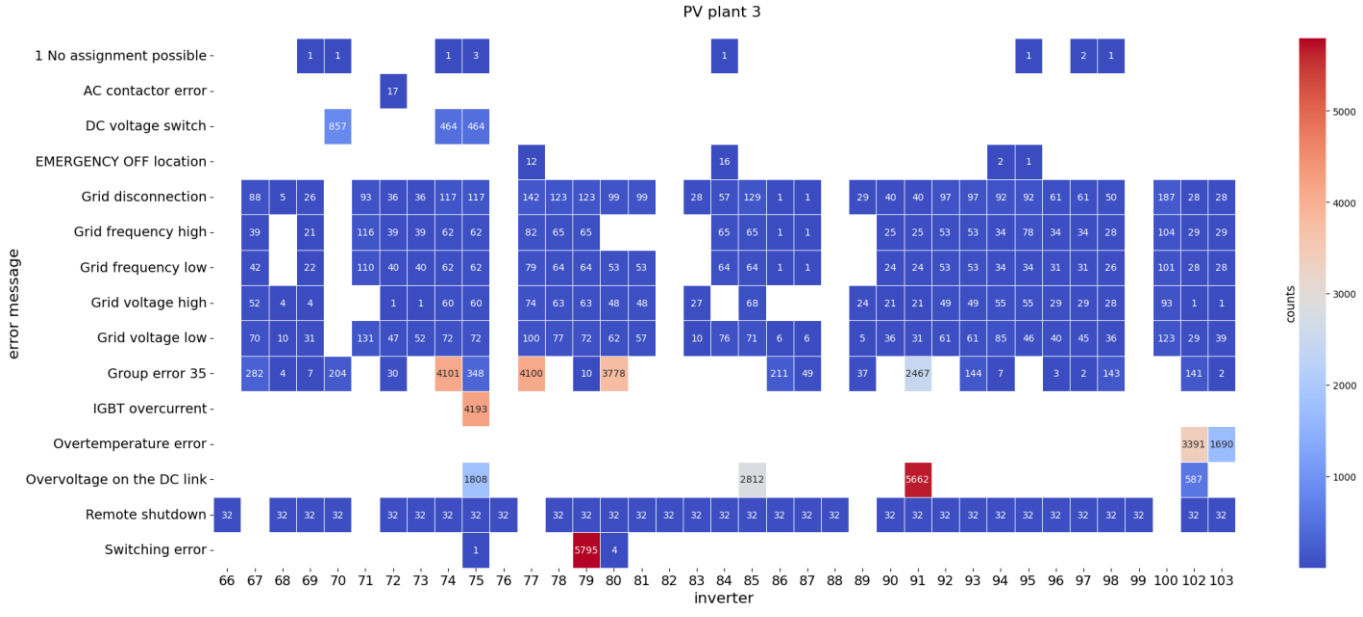

Fig. 16. Number of error messages occurring after curtailment in PV plant 3.

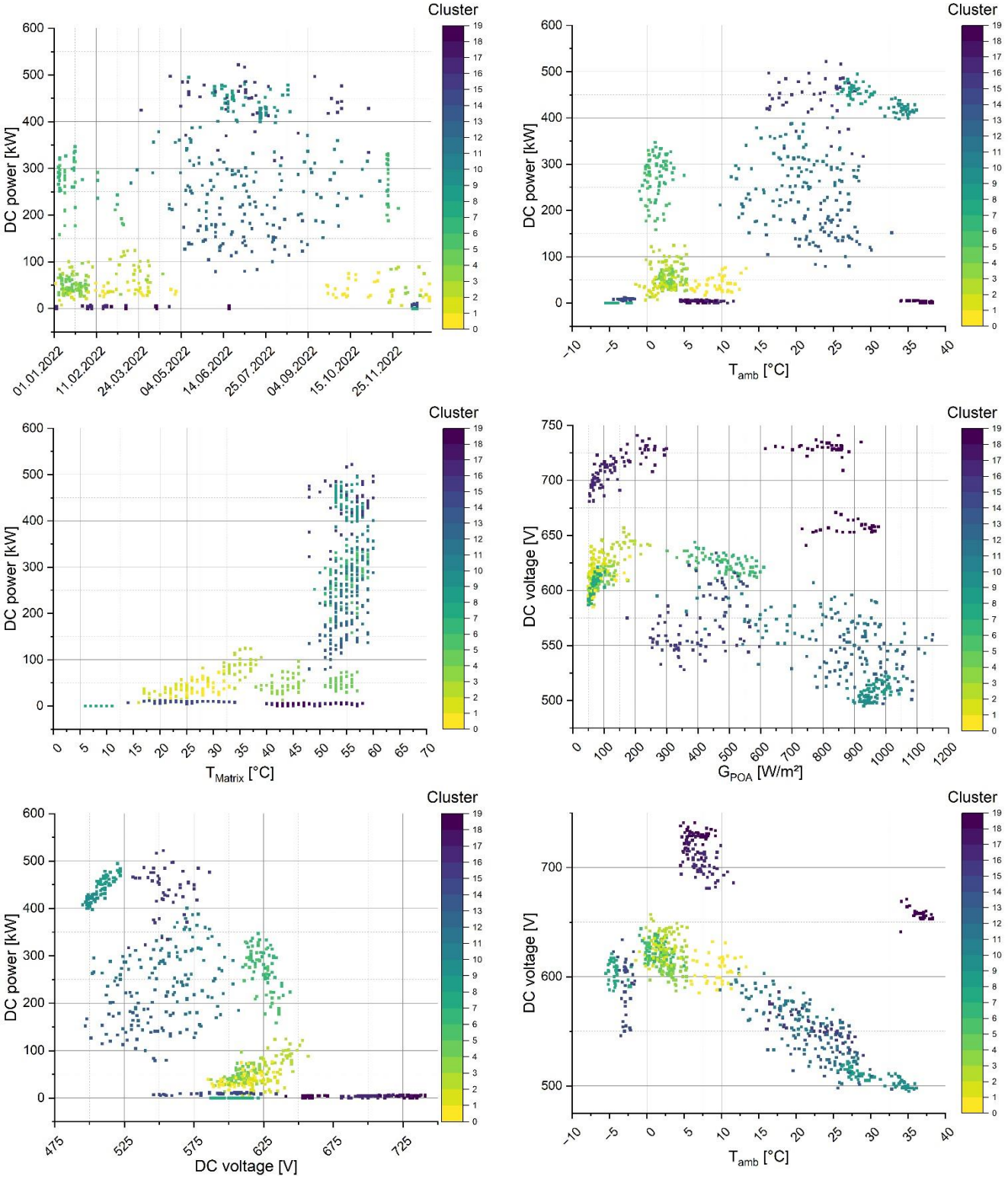

Fig. 17. Supplementary types of representation of the results from 2nd Clustering.

Glossary

Supplementary material

References

[1] SolarPower Europe (2024): Global Market Outlook for Solar Power 2024-2028. June 2024. ISBN: 9789464669169.

[2] J. Leloux. Mapping the Relevance of Digitalization for Photovoltaics. Intersolar Conference 2024. June 2024.

[3] P. Hacke, S. Lokanath, P. Williams, A. Vasan, P. Sochor, G. TamizhMani, H. Shinohara, S. Kurtz. A status review of photovoltaic power conversion equipment reliability, safety, and quality assurance protocols. Renewable and Sustainable Energy Reviews 82 (2018) 1097-1112. DOI: http://dx.doi.org/10.1016/j.rser.2017.07.043

[4] A. Golnas. PV System Reliability: An Operator's Perspektive. IEEE JOURNAL OF PHOTOVOLTAICS, VOL. 3, NO. 1, JANUARY 2013. DOI: 10.1109/JPHOTOV.2012.2215015

[5] J.M. Freeman, G.T. Klise, A. Walker, O. Lavrova. Evaluating Energy Impacts and Costs from PV Component Failures. WCPEC-7. 2018. DOI: 10.1109/PVSC.2018.8547454

[6] VDMA e.V. International Technology Roadmap for Photovoltaic (ITRPV). 2022 Results. 14 Edition, April 2023.

[7] J. Falck, C. Felgemacher, A. Rojko, M. Liserre, P. Zacharias. Reliability of Power Electronics Systems. An Industry Perspective. 2018. DOI: 10.1109/MIE.2018.2825481.

[8] M. Shahzad, K.V.S. Bharath, M.A. Khan; A. Haque. Review on Reliability of Power Electronic Components in Photovoltaic Inverters. 2019 International Conference on Power Electronics, Control and Automation (ICPECA). 2020. DOI: 10.1109/ICPECA47973.2019.8975585.

[9] A. Sangwongwanich, Y. Yang, D. Sera, F. Blaabjerg. Mission Profile-Oriented Control for Reliability and Lifetime of Photovoltaic Inverters. IEEE Transactions on Industry Applications (Volume: 56, Issue: 1, Jan.-Feb. 2020). DOI: 10.1109/TIA.2019.2947227. 2019

[10] SolarPower Europe. Operation & Maintenance. Best Practice Guidelines. Version 5.0. ISBN: 9789464444247. December 2021.

[11] IEA-PVPS. Guidelines for Operation and Maintenance of Photovoltaic Power Plants in Different Climate. Report IEA-PVPS T13-25:2022. 2022. ISBN 978-3-907281-13-0

[12] A. Livera, M. Theristis, L. Micheli, E.F. Fernández, J.S. Stein, G.E. Georghiou. Operation and Maintenance Decision Support System for Photovoltaic Systems. IEEE Access (Volume: 10). 2022. DOI: 10.1109/ACCESS.2022.3168140

[13] D. Daßler, S. Malik, S.B. Kuppanna, B. Jäckel, M. Ebert. Innovative Approach for Yield Evaluation of PV Systems Utilizing Machine Learning Methods. IEEE 46. PVSC. Chicago. Juni 2019. DOI: 10.1109/PVSC40753.2019.8981367

[14] D. Daßler, S.B. Kuppanna, S. Malik, R. Schmidt, M. Ebert. Training and Evaluation for Yield-Driven Detection of Losses in PV Systems Utilizing Artificial Neural Networks. IEEE 47th Photovoltaic Specialists Conference (PVSC). Virtual. 2020. DOI: 10.1109/PVSC45281.2020.9300490

[15] G.D. Rupakula, D. Daßler, S. Malik, M. Ebert, R. Schmidt. Automatic Fault Detection and Classification in PV Systems by the Application of Machine Learning Algorithms. 38th European Photovoltaic Solar Energy Conference and Exhibition. 2021. DOI: 10.4229/EUPVSEC20212021-5DO.1.2

[16] D. Daßler, S. Malik, R. Gottschalg, M. Ebert. Effect of Availability and Quality of Data on the Detection of Defects Utilizing Artificial Neural Networks in PV System's Monitoring Data. 8th World Conference on Photovoltaic Energy Conversion. Milan. September 2022. DOI: 10.4229/WCPEC-82022-4DO.1.5.

[17] Deutscher Wetterdienst. CDC (Climate Data Center). Accessed: 05/17/2024. URL: https://www.dwd.de/DE/klimaumwelt/cdc/cdc_node.html

[18] S. Malik, D. Daßler, J. Fröbel, J. Schneider, M. Ebert. Outdoor data evaluation of half- /full-cell modules with regard to measurement uncertainties and the application of statistical methods. 29. European Photovoltaic Solar Energy Conference and Exhibition. Amsterdam. September 2014.

[19] Ankerst, Mihael, Markus M. Breunig, Hans-Peter Kriegel, and Jörg San-der. "OPTICS: ordering points to identify the clustering structure." ACM SIGMOD Record 28, no. 2 (1999): 49-60.

[20] Schubert, Erich, Michael Gertz. "Improving the Cluster Structure Ex-tracted from OPTICS Plots." Proc. of the Conference "Lernen, Wissen, Daten, Analysen" (LWDA) (2018): 318-329.

[21] Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

[22] Reprint of: Mahalanobis, P.C. (1936) "On the Generalised Distance in Statistics.". Sankhya A 80 (Suppl 1), 1–7 (2018). https://doi.org/10.1007/s13171-019-00164-5