FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Cellular Time Activation Networks, a novel approach applied to photovoltaic anomaly detection

David da Silva Moreira Freire

WORKING VERSION

Mestrado Integrado em Engenharia Eletrotécnica e de Computadores

Orientador: Cláudio Domingos Martins Monteiro

June 30, 2023

Resumo

A proliferação de centrais fotovoltaicas de dimensão industrial levou à necessidade de métodos para detetar e classificar falhas nos seus componentes, sendo que estas que podem ter impactos económicos significativos. Neste trabalho, explora-se o estado da arte das ferramentas de deteção de falhas e estimação do estado aplicadas ao campo dos sistemas PV, com foco na compreensão do seu funcionamento, identificando-se pontos fortes e possíveis limitações. Relevar-se-á quais os métodos baseados em aprendizagem computacional mais utilizados. Ainda assim, reconhece-se o contributo dos diversos domínios para colmatar este tipo de problema, desde a teoria dos grafos a processamento de sinal, aprendizagem profunda e aprendizagem quântica. Efetuam-se comparações e propostas de melhoria aos algoritmos existentes, e desenvolvida uma nova abordagem para abordar o tema de deteção de falhas. Com a retrospeção das ferramentas contemporâneas de maior sucesso, e pela oferta de uma nova abordagem, o objetivo deste trabalho é fornecer aos operadores de instalações fotovoltaicas o aumento na fiabilidade e eficiência dos seus sistemas. Além disso, há a possibilidade de que a ferramenta desenvolvida seja aplicável para outros problemas de coesão de dados, impactando positivamente os diversos tipos de domínios de sistemas orientados a dados.

Abstract

The increase in utility-scale photovoltaic power plants has led to the need for effective methods for detecting and classifying component faults, which can have significant economic impacts. This work assesses the current state of fault detection and state estimation tools in the field of PV systems, focusing on understanding how these tools work and identifying their strengths and limitations. It is seen that machine learning makes up the majority of state-of-the-art fault detection and classification algorithms. Still, many fields have contributed to this problem, from graph theory to signal processing, deep learning, and quantum machine learning. Consequently, this work compares and proposes improvements to existing approaches or a novel technique developed to address this issue. By examining the most successful tools to date and offering new solutions, the intention is to help PV plant operators improve the reliability and efficiency of their systems. The developed methodology is also expected to become a generalistic data cohesion algorithm, positively impacting other data-driven fields.

Agradecimentos

Aliquam id dui. Nulla facilisi. Nullam ligula nunc, viverra a, iaculis at, faucibus quis, sapien. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Curabitur magna ligula, ornare luctus, aliquam non, aliquet at, tortor. Donec iaculis nulla sed eros. Sed felis. Nam lobortis libero. Pellentesque odio. Suspendisse potenti. Morbi imperdiet rhoncus magna. Morbi vestibulum interdum turpis. Pellentesque varius. Morbi nulla urna, euismod in, molestie ac, placerat in, orci.

Ut convallis. Suspendisse luctus pharetra sem. Sed sit amet mi in diam luctus suscipit. Nulla facilisi. Integer commodo, turpis et semper auctor, nisl ligula vestibulum erat, sed tempor lacus nibh at turpis. Quisque vestibulum pulvinar justo. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Nam sed tellus vel tortor hendrerit pulvinar. Phasellus eleifend, augue at mattis tincidunt, lorem lorem sodales arcu, id volutpat risus est id neque. Phasellus egestas ante. Nam porttitor justo sit amet urna. Suspendisse ligula nunc, mollis ac, elementum non, venenatis ut, mauris. Mauris augue risus, tempus scelerisque, rutrum quis, hendrerit at, nunc. Nulla posuere porta orci. Nulla dui.

Fusce gravida placerat sem. Aenean ipsum diam, pharetra vitae, ornare et, semper sit amet, nibh. Nam id tellus. Etiam ultrices. Praesent gravida. Aliquam nec sapien. Morbi sagittis vulputate dolor. Donec sapien lorem, laoreet egestas, pellentesque euismod, porta at, sapien. Integer vitae lacus id dui convallis blandit. Mauris non sem. Integer in velit eget lorem scelerisque vehicula. Etiam tincidunt turpis ac nunc. Pellentesque a justo. Mauris faucibus quam id eros. Cras pharetra. Fusce rutrum vulputate lorem. Cras pretium magna in nisl. Integer ornare dui non pede.

David Freire

*"You should be glad that bridge fell down.
I was planning to build thirteen more to that same design"*

Isambard Kingdom Brunel

Contents

List of Figures

List of Tables

Abreviaturas e Símbolos

ANN	Artificial Neural Network
CNN	Convolutional Neural Network
CXN	Cell Complex Neural Network
DC	Direct current
DL	Deep Learning
DNN	Deep Neural Network
LSTM	Long short-term memory
MCD	Minimum Covariance Determinant
ML	Machine Learning
PV	Photovoltaic
RBFNN	Radial basis function neural network
RMM	Recurrent Neural Network
SC	Short Circuit
SRC	Sparse Representation Classifier
STC	Standard Test Conditions
SVM	Support Vector Machine

Chapter 1

Introduction

The XIX century marked a significant shift in the world's perception of energy resources as the desire to invest in renewable energy sources to power modern societies grew. This transition was driven by the need to reduce dependency on fossil fuels, mitigate the effects of global warming, and slow climate change. Renewable energy sources offer a range of benefits, including reduced greenhouse gas emissions, increased energy security, and air quality. Solar photovoltaic energy is a desirable renewable energy source due to its abundance, accessibility, and environmental benefits. While solar photovoltaic energy has proven to be both cost-efficient and environmentally friendly, it also comes with unprecedented challenges, such as its intermittent nature, low electrical inertia, complex forecasting, and geographic-dependent operating conditions. Despite these challenges, recent reports [1] show that the economic benefits of investing in renewable energy outweigh the complications, as there is an increasing global investment trend in these sources.

The general construction of PV farms, particularly on the utility-scale, has led to a need for effective maintenance and monitoring to ensure maximum efficiency and operational reliability. Towards this, various algorithms and routines are used to monitor the state of PV farms and identify any potential issues that may arise. Fault detection is crucial to this process, allowing PV farm operators to identify and address problems quickly. Detecting faults and identifying the necessary steps can prevent or minimize downtime and ensure optimal performance. Given the importance of maintaining high levels of operation, knowing if action is needed to restore or fix components from an anomalous scenario is desirable for reducing investment risk and maximizing profits.

Integrating intermittent energy resources into modern electric grids has led to stricter requirements for connecting such power systems to ensure safe grid operating conditions. As a result, companies that own or plan to build photovoltaic farms must comply with these requirements and have adequate power electronics and monitoring/control capabilities. Failure to meet these requirements can result in sanctions or fines for the responsible party, as well as potential impacts on system availability, asset value, and disturbance propagation to the grid. To minimize these risks and maximize the value of their assets, companies may opt to implement fault detection and state estimation tools. These tools allow for the early detection and resolution of potential issues and can prevent or minimize downtime. The need to create or improve existing fault detection and state estimation tools, and the search for the most effective methodologies for addressing these issues, drive research in this field.

Having laid the basis for why there must be system behavior assessment in utility-scale PV plants, it is necessary to understand what business concepts are crucial to this field. In the course of this work, the presented topics will go over the following questions:

- What components mostly fail in photovoltaic power systems?
- What is the average frequency of faults?
- What fault detection/state estimation tools exist for photovoltaic power systems?
- What are the most successful ones?
- What's their structure? Are they mostly centralized or decentralized?
- What are their computational costs/efficiency?
- What is the expected magnitude of precision and confidence?
- Which key performance indicators can evaluate the success of these tools?
- What are their implementation difficulties?

With these questions uncovered, the main objective is to adapt or design a novel algorithm/approach to fault detection based on modern artificial intelligence solutions. However, this can be split into finer goals:

- Identify and study existing fault detection tools for photovoltaic power systems.
- Adapt or develop a new tool.
- Apply and test the new tool in real case study PV assets.
- Validate the developed methodologies by comparison to reference tools.

Before reviewing state-of-the-art fault detection tools, types of failures in photovoltaic systems need to be understood: find which components usually fail, which ones fail more often, and how often. For this, it is necessary to understand such components' physical and electrical properties and the modeling techniques used to characterize them. There will be an assessment of utility-scale power plants architecture through literature, alongside the detection objective of state-of-the-art fault detection tools applied in this field. Then, there shall be an extensive analysis and review of what tools have been designed and used in this field. In this step, critical evaluation of the literature is a must for understanding the tool's scope, ease of implementation, and understanding that the data sets available for this work are compatible. Having selected the most prominent ones, they're to be qualitatively and quantitatively compared to each other in their application context so that the results allow objective evaluations. This process requires implementing these tools, following the guidelines in the respective article/book/report, verifying their metrics, and checking if the achieved results resemble the same as the literature suggests. It will require gathering data sets, which can either be artificially generated through simulation or provided by an enterprise that services photovoltaic plant owners.

There's a desire that, in the end, the developed work helps achieve an improved method for fault detection and state estimation in photovoltaic power systems, resulting in a production-ready software application agile enough to deploy for multiple PV assets. It's intended that the algorithm specializes in data cohesion as a means of anomaly inference, allowing asynchronous and self-healing data transfers between the considered components. Depending on the new algorithm's characteristics, it could result in an approach capable of generalization and application to other engineering systems, benefiting more than just PV systems. No matter the chosen methodology, fault detection will, in most cases, result in an economic benefit, catastrophe prevention, and safety increase.

Chapter 2

Fault detection in Utility Scale Photovoltaic Plants

2.1 Utility-Scale Photovoltaic System's Architecture

Utility-scale photovoltaic (PV) power plants are large-scale systems connected to the electrical grid, having installed capacities ranging from kilowatts peak (kWp) to megawatts peak (MWp). These systems typically consist of many PV panels interconnected through power electronics to aggregate and inject power into the grid. The number and type of components in a PV power plant depend on the plant's scale and topology, with different configurations possible for large-scale applications, including central inverters, string inverters, and multi-string inverters [2]. The physical installation of PV modules can include solar tracking apparatuses, such as single and dual-axis trackers [3], which add to system complexity and change production behavior. Understanding the architecture and components of PV power plants is vital for designing, operating, and maintaining these systems, as it helps optimize their performance and reliability.

Figure 2.1 presents a typical utility-scale PV plant architecture using the central inverter (or possibly multi-string inverter) configuration. It is noticeable that many system components may fail in one or more ways, which is why monitoring and fault detection algorithms are essential to maintain state estimation. The main subsystems considered in this work are the following:

- Solar photovoltaic panels (with or without bypass diodes).
- Tracking mount.
- Electrical cabling.
- Inverter(s) (mostly with Max Power Point Trackers).
- AC Transformer(s).
- Protection components (circuit breakers, fuses, surge protectors, etc.)

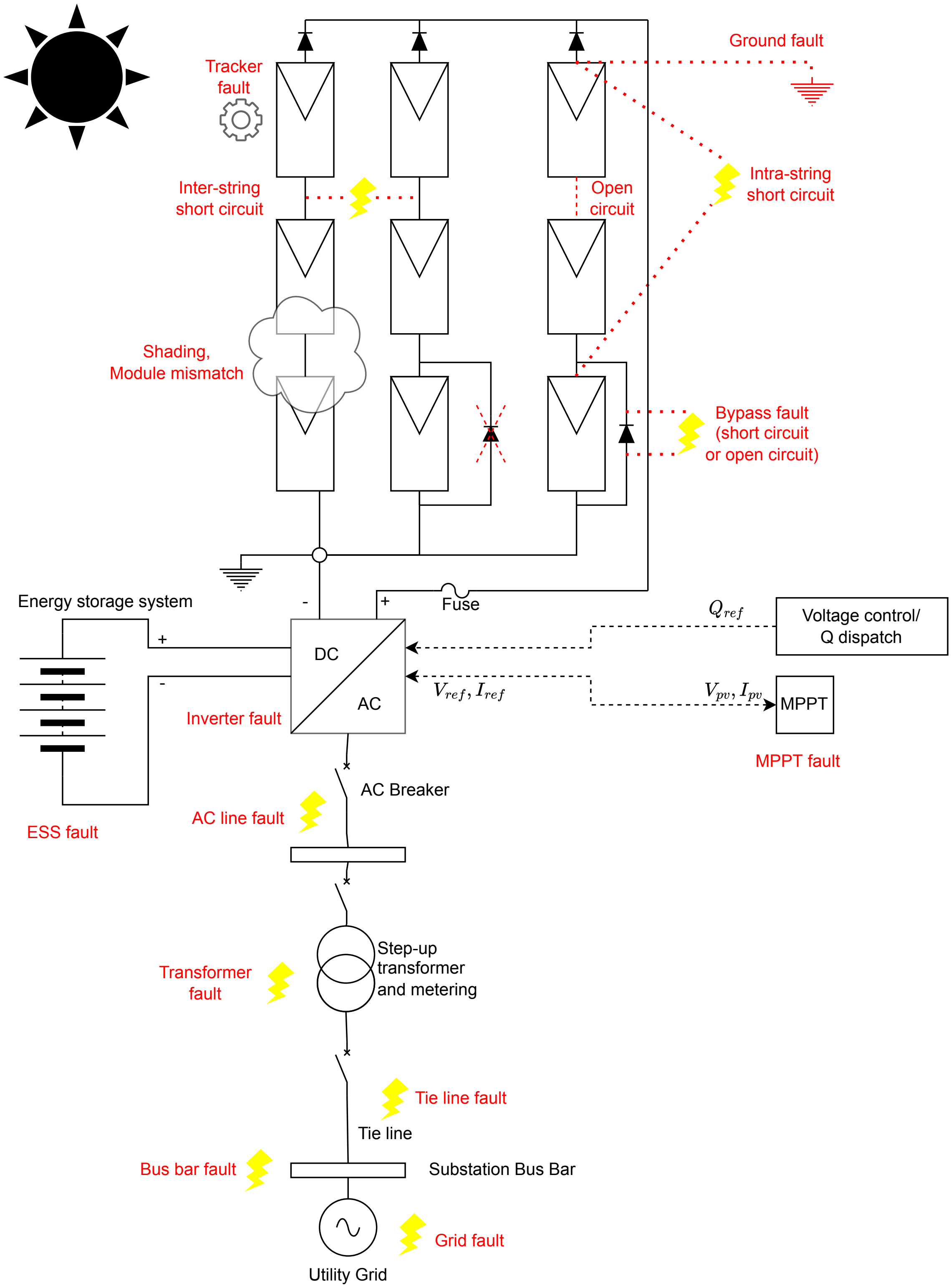

Figure 2.1: Representation of utility-scale PV plant components and some possible faults.

Most of these components have intrinsic variables, such as voltage and current values, that can help determine their operation states. Given that the utility grids (and the associated electricity market) integrate large-scale PV assets, some of the before-mentioned components require constant monitoring and control, achieved with adequate embedded systems and sensor infrastructure [4]. Since monitoring utility-scale PV assets relies on the investment and technologies employed, engineers must consider data availability when developing data-driven algorithms. Thanks to the continuous advancements in communication technologies, namely in IoT (Internet Of Things), data acquisition is becoming faster, more reliable, and more precise. Not only is this fundamental for real-time asset assessment, but it also allows better training of fault detection algorithms. However, on the industrial scale (in the order of MWp production), having sensors embedded in every PV module comes with a high economic cost. Inverters are the components that usually possess monitoring capabilities, though the grid-tie connection should also be equipped with sensors. These can be considered the primary sources of information from utility-scale PV plants, with the most accurate, fast, and reliable data acquisition.

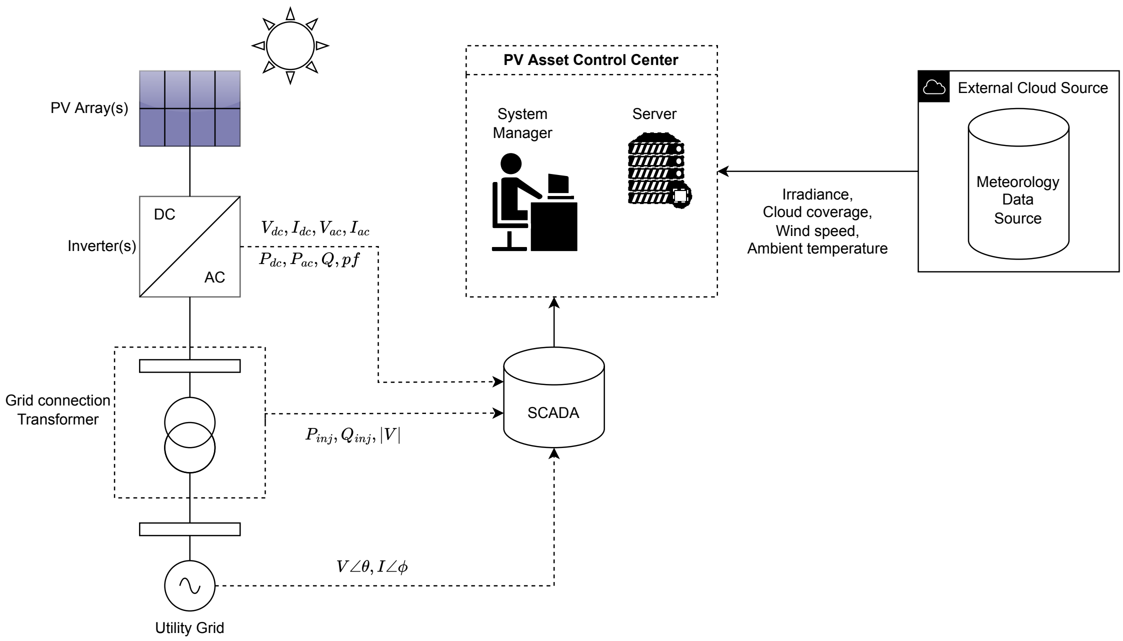

Figure 2.2: Typical data flow of utility-scale PV power plants.

Figure 2.2 represents a simplified data flow representation of a grid-tied PV system's most commonly available state variables, with most of them suggested by the IEC 61724 standard [5]. An external meteorological data source is defined since the PV system manager usually needs climate information for (at least) forecasting purposes.

2.2 Faults in Photovoltaic Systems

Several types of faults can occur in utility-scale photovoltaic (PV) power plants, which impact the performance and reliability of the system negatively. Unfortunately, some are very challenging to detect and protect the electrical installation against, thus requiring sophisticated detection algorithms [6]. Besides the economical price, their occurrence may even cause safety hazards, such as fires [7], thus the urgency in detecting or preventing such events early.

According to [6], these faults can fit into three categories: electrical, mechanical, and environmental. Electrical malfunctions include short circuits, open circuits, and inverter failure, affecting the PV panels' power output and the system's overall efficiency. Mechanical faults include broken panels, damaged cables, and defective inverters, which can lead to system downtime and reduced performance (although not mentioned, solar tracker failures could also belong in this category). Environmental faults include extreme weather events, such as hail or strong winds, which can damage the PV panels and other components [8].

The authors in [9] cover a comprehensive review of most types of faults studied in the ambit of detection and classification algorithms. However, authors in [10] have a more succinct fault categorization that better fits this work's scope. They categorize all the major PV system faults into either DC-side or AC-side. Figure 2.3 represents this detailed categorization with a tree-like structure.

Although also prone to failure, most literature on fault detection and classification for photovoltaic systems does not encompass solar tracking faults: most studies cover fixed PV systems. The supervision and assessment of these subsystems' correct functioning can be sensor-based [11] or image-based. Some authors developed fault detection methods for these apparatuses [12], using image processing on aerial photography to determine modules' slopes. This category of failures should be better supported when developing electrical data-driven algorithms since they can significantly affect the system's efficiency. Hence, this work will attempt to include said fault category in the proposed fault detection methodology.

Throughout the literature [13], some of the most noted faults in the context of fault detection are:

- Shading: partial coverage of a PV array or module, temporary or not. It might result in a Hot Spot fault.
- Soiling: dirt accumulation, blocking sunlight from reaching PV Cells. It might also result in a Hot Spot fault.
- Short circuit: either line-line or line-ground.
- Open circuit: connection breakage between modules.
- DC arc fault: electricity plasma arc formed on broken connections.

According to a 2017 survey conducted on five utility-scale PV plants in Italy [14], the authors observed failure rates from <1% to 3% in the majority of plants and 81.8% in the worst scenario. The high failure rate of the latter had a demonstrated cause that originated from manufacturing mistakes: snail trails. Besides this phenomenon, hot spot faults and bypass diode faults/disconnections were among the most common.

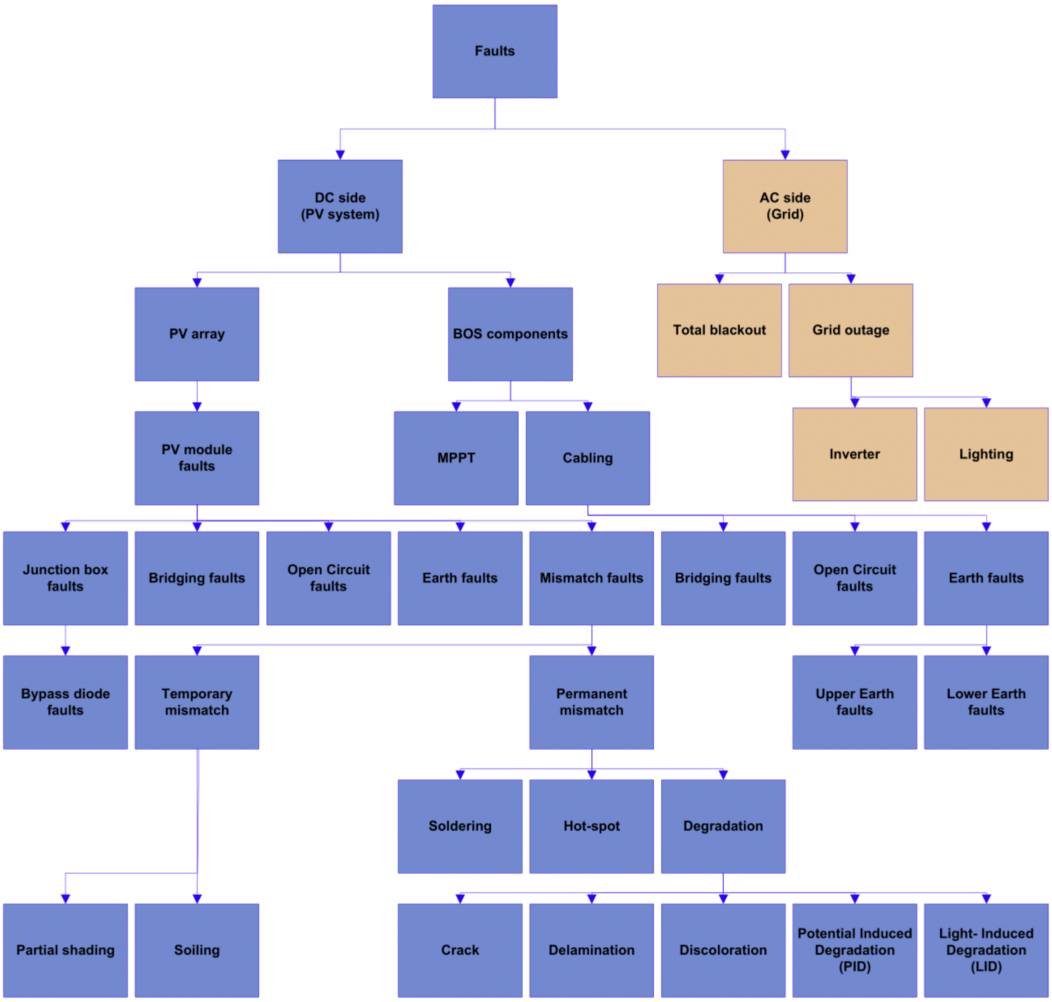

Image source and copyright: [6].

Figure 2.3: "Failures in grid-connected PV systems."

Alongside manufacturing failures, installation, planning, and other external effects can be the root cause for many of the presented faults [15].

Having the distribution of fault types from real-life scenarios is quite helpful for formulating fault detection algorithms. It allows for better generation/selection of training data and class decisions. In figure 2.4, it is possible to observe the failure type distribution for 24.254 inspected modules. Soiling, shading, and mechanically related failures were not as prominent, with only a group share of around 6%. It is relevant to note that discoloration represents almost a quarter of all faults.

Although the study had a limited geographic scope, with only a few power plants diagnosed, it allows for a more realistic view of the common scenarios encountered in typical operational ground-mounted utility-scale PV power plants.

Due to the difficulty of classifying some of these faults, given their similarity on the consequent effect in the system, it will be seen in further sections that most fault detection algorithms only

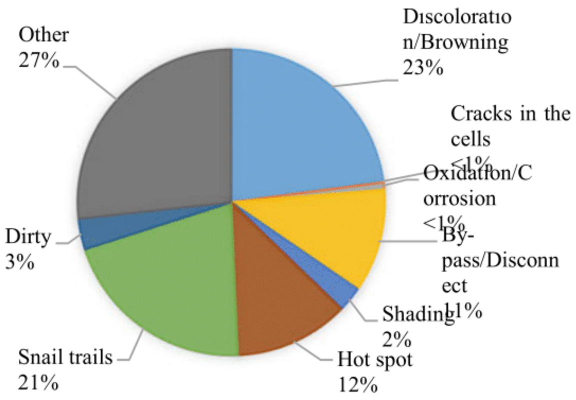

Image source and copyright: [14].

Figure 2.4: "Circle chart related to the module defects in the 5 plants (over the total number of failures)."

endeavor to classify between two to five types of reviewed faults.

2.3 Modeling photovoltaic's physical/electrical behavior

Photovoltaic cells are the fundamental components of photovoltaic panels. They are made from semiconductor materials like silicon and absorb photons that generate electric current. Their electrical behavior is characterizable using the current-voltage (I-V) equation 2.1. This equation, which represents a fundamental relationship governing the operation of PV cells, can be used to predict their performance under various operating conditions, such as differing solar irradiance and temperatures.

$$I = I_{ph} - I_d \times (e^{\frac{q \times (V_{pv} + I_{pv} \times R_s)}{n \times k \times T}} - 1) - \frac{V_{pv} + I_{pv} \times R_s}{R_p} \tag{2.1}$$

I_{ph} (A) is the light-generated current; I_0 (A) is the reverse saturation current; V_{pv} is the module's terminal voltage; I_{pv} is the module's output current; R_s (Ω) is the series resistance; R_p (Ω) is the shunt resistance; n (adimensional) is the diode ideality factor; k (J/K) is the Boltzman constant; T (K) is the cell temperature; q (C) is the electron charge;

For state estimation, it is crucial to accurately model PV modules' performance from the DC side of power converters. This information is vital for designing and optimizing PV power systems, as it enables predicting PV module performance under different conditions, as mentioned before. Accurate PV module models are also essential for state estimation and fault detection, as they provide critical information about their health and performance, allowing for early identification of potential issues. Moreover, they can be used to optimize the control and operation of PV power systems, improving their efficiency and reliability [13].

Physical and empirical models broadly categorize the several state-of-the-art methods for modeling photovoltaic modules [13]. Physical models lie on the fundamental physical principles governing PV modules' operation. They typically require detailed knowledge of the PV module's electrical and optical properties, such as its current-voltage (I-V) characteristics, spectral response, and temperature dependence. These models can accurately predict the PV module's performance under a wide range of operating conditions, but they may be complex and computationally intensive to implement [16]. On the other hand, empirical models are based on experimental data and are typically more straightforward to implement. However, they may not be as accurate as physical models, especially under conditions that differ significantly from those used to generate the experimental data (usually STC) [13]. Some examples of state-of-the-art physical models for PV modules include the single-diode model (the five-parameter model) and the two-diode model [17]. In contrast, one of the most used state-of-the-art empirical models is the Sandia model [13]. The choice of modeling method will depend on the specific application and the required level of accuracy and complexity; in some cases, there can be a combination of physical and empirical models.

In the case of utility-scale PV systems, detailed knowledge of the module's electrical and optical properties of empirical data may be limited, and building a model is only possible by recurring to the datasheet information. A complex model that requires more detailed information may not be feasible in such cases, and a simpler model that relies on fewer input parameters is more appropriate. Given the excellent trade-off between complexity and accuracy, the single-diode model suits this use case.

2.3.1 The five-parameter model

Figure 2.5 presents the single-diode model representation of the photovoltaic module. According to the five-parameter model, the unknown parameters are determined by fitting the model to experimental data or using data from the PV module's datasheet. The single-diode model can predict the PV module's performance under a wide range of operating conditions while maintaining reasonable accuracy. However, remembering that the single-diode model is a simplified representation of the PV module, it will have poor accuracy under certain situations compared to the more representative two-diode model [17].

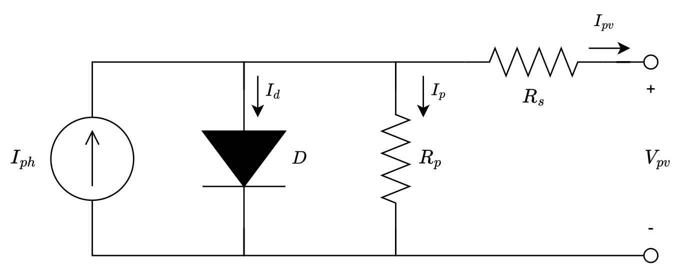

Figure 2.5: Single-diode model for photovoltaic modules.

2.4 Literature on Fault Detection and Classification for Photovoltaic Systems

The parent field of fault detection is anomaly detection (also known as outlier detection), a highly studied subject in the scope of statistics [18], applied in many scientific areas. Classification is also well-studied in this field, with applications in numerous scientific contexts, from medical diagnosis to airport safety [19]. Consequently, adaptations of generic tools and ad hoc methodologies have originated to aid in solving fault detection and classification problems in photovoltaics.

According to [4], the tools dedicated to PV fault detection and state estimation mostly come from mathematical/statistical methodologies, machine learning, and deep learning applications. Regarding the three general problem-solving principles mentioned before, it's known that machine learning and deep learning are the most popular and successful ones for recent applications that ought to solve complex problems. However, this categorization is somewhat limited, with contemporary literature suggesting an abundance of developed methodologies from different backgrounds, thoroughly reviewed in [9] and [10]. In [9], the authors consider two principal fault detection and classification algorithm branches: image-based and electrical-based; while [10] also distinguishes numerical-based techniques. Image-based refers to aerial or visual capture of the PV array by photography and thermal imaging, commonly used along with artificial intelligence algorithms for assessing the photovoltaic module's state. Although the contribution and importance of such methods are appreciable, this work will mainly focus on the electrical-based and numerical-based ones, as the use case of the developed tool is bound to this type of data.

Categorizing methodologies becomes fuzzy, considering that some literature mixes physical behavior models with machine learning, statistics, and signal processing. Figure 2.6 is an attempt to present a structure inspired by the review made by [9], [10], and this work's, with a focus on the more relevant techniques (for this work's scope). Hybrid models are ubiquitous since combining robust statistical, signal processing, ML, or DL models and PV's electrical characterization can achieve more remarkable results. Hence, a better representation than figure 2.6 would be an incomprehensible mesh of connections representing the permutations between category aggregation.

To not wander in the literature, there must be a decision on which methodologies to revise. The developed tool in this work must meet certain real-life constraints, such as data availability, frequency, accuracy, PV system configuration, and context. Therefore, the (qualitative) potential estimation for each methodology will be based on the capability of adapting the proposed algorithms to the same expected restrictions. This evaluation process confines the methodology review to emphasize the ones thought to be most capable of implementation in a real scenario. Therefore, the following sections will not cover an extensive literature review, as it is not intended to repeat the works of [9] and [10], only presenting interesting or adequate methodologies related to this work's scope.

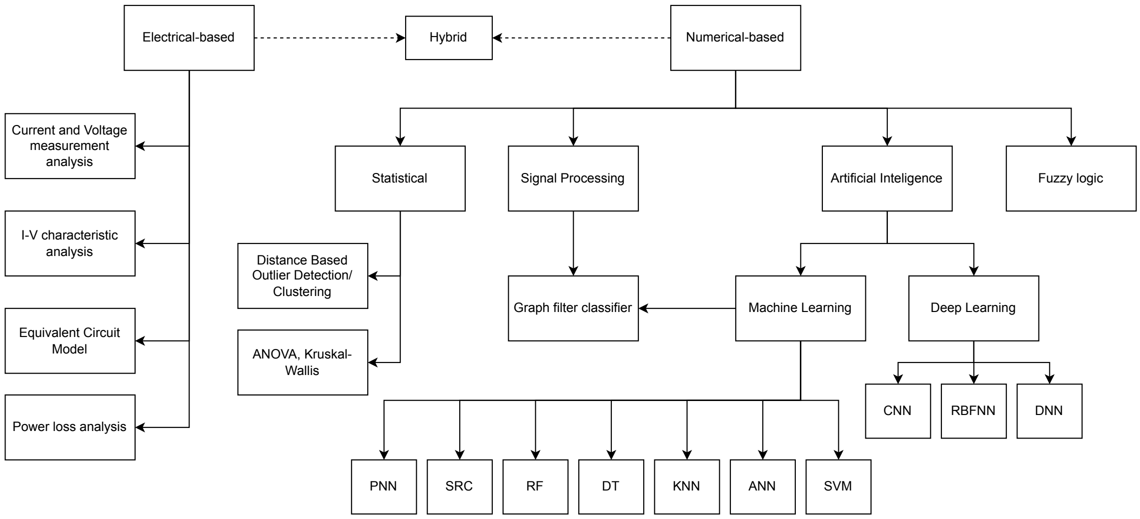

Figure 2.6: Representation of some of the methodologies employed in fault detection for PV systems.

2.4.1 Statistical and Signal Processing Algorithms

Statistical methodologies look into historical data to find the characteristics of how samples relate to the population (interpolation). These methodologies yield good results in case studies of PV farms that have been logging data for a considerable time, suffering in the cases that do not. Therefore, they are limited in that it is required to have curated data sets of historical significance for relevant features of the studied systems.

The literature on statistical and signal processing fault detection algorithms for PV is mostly quite dated ([20], [21], [22]), given that more recent machine learning methods have become increasingly attractive in this matter. Nonetheless, anomaly (or outlier) detection statistical algorithms can be used for fault detection in PV systems by identifying unusual patterns or deviations from normal behavior in the data collected from the PV system. Distance-based methods, such as the Euclidean, Mahalanobis, and MCD-based distances [13], may be adequate. Although simple, these techniques might only work for detecting outliers in the context of PV systems if they are scale-invariant (due to the different magnitude in the system's state variables) and resilient to outlier contamination (which only MCD-based distance is capable of). In [22], the authors applied Analysis of Variance (ANOVA) and Kruskal-Wallis test for inverter failure detection, with the only downside of only being able to identify outliers in a sub-array resolution, i.e., not for specific string or module failures.

Some algorithms consider incoming data from PV systems as signals, allowing the adaptation of signal processing theory to develop ad hoc algorithms. Coming up with a relatively simple algorithm, the authors in [23] propose a power-based fault detection method that only requires delayed samples of the PV array's power output and a threshold. It reasons that since the power output of PV systems can't vary beyond a given point, considering a very short-term period (milliseconds), significant perturbations in this variable can be associated with faults. Although the simplicity and ease of implementation, it's clear that the success of this method requires feeding the algorithm with relatively high-frequency data, which would only be feasible on-site (and with specialized monitoring equipment).

In [24], the authors successfully formulated a graph signal processing algorithm for fault classification that yields increasingly better results when there is a considerable amount of labeled data, although its training is only semi-supervised. The results outperformed other standard machine learning methods for the same training data, given 30% or more of labeled data. On another note, the data utilized came from the PVWatts [25] dataset, and the PV system is on a small scale (ASU testing facility [26]) possessing a monitoring density and capability that can be considered irrealistic for utility-scale. This same data source is present in many other reviewed works.

The authors in [27] displayed another excellent use for graph theory, although not specifically for fault detection: they implemented a consensus-based distributed approach to minimize the impact of noise in acquired data from the PV array. By formulating a data propagation algorithm that resulted in measurement convergence, they achieved higher accuracy for state estimation.

With both graph theory-based algorithm proposals, this field sparks interest in its usage for the upcoming formulated methodology, given that it would be desirable to achieve an algorithm that features fault detection alongside data consensus.

2.4.2 Machine Learning Algorithms

Machine learning is the trending way of solving increasingly complex and non-linear problems, as neural networks (or other learning structures) can better model complex, non-trivial, and nonlinear relations between data. Still, they are as good as the training data, with many designs requiring a lot of representative learning examples to achieve good results. Their output can also be very obfuscated (depending on the technique), meaning that many methods do not allow a direct interpretation of the relationship between inputs and outputs. This "black-box" characteristic, specifically of neural networks, is considered a disadvantage. Besides, extrapolating data remains a challenge when classically using these structures. Still, they have immense applications for PV systems, from MPP (Max Power Point) estimation to power forecasting, soiling, and fault prediction.

In [28], the authors utilize an ANN to classify short circuit and hot spot faults. This algorithm achieved an outstanding 98.4% classification accuracy, yet the data originated from *MatLab/Simulink* simulations and only considered two classes of faults. Because the inputs were the variation of voltage and current ($\frac{dV}{dt}$, $\frac{dI}{dt}$), the algorithm required data sampling with relatively high frequency (>5Hz). The present work will not regard such methodologies as background for the upcoming tool since requiring high-frequency simulated data while covering only two fault types is quite far from a real utility-scale PV system scenario.

The trend of utilizing simulated data (sometimes without even added noise) has been a target of criticism in [29]. Accordingly, this work also emphasizes that the literature shows many proposed ML (and other types of) techniques that fall into this concept, which makes selecting appropriate methodologies to base future work on a challenging task.

The proposed ANN solution in [30] is remarkable by the diversity of fault classification achieved: STC, short circuit, varying temperature, partial shading, complete shading, degraded modules, ground fault, and arc fault. It presents one of the most fault class coverage with high accuracy, considering the literature that utilizes synthetic noiseless data. Hence, the cyber-physical conceptualization and data preprocessing (clustering) demonstrated can be admired, but not forgetting that validation data came from a relatively unrealistic setting.

In [31], there is a captivating proposal of utilizing an autoencoder and pruned neural network to separate the tasks of detecting and classifying faults, which resulted in one of the most performant ML approaches in the literature. The algorithm classifies five states: degraded, shaded, soiled, short circuit, and STC, utilizing nine inputs representing voltage, current, power, and irradiance available from the MPPT, datasheet, or meteorological sources. While the neural network pruning adds complexity, it resulted in a better generalized and lighter-weight trained model suitable for faster detection times. Even though using data from a small-scale PV system, the presented algorithm and its assumptions may make it possible to adapt and implement in an industrial scenario.

Regarding performance, the work in [32] proposes a sparse representation classifier (SRC) that evaluates if the system has line-to-line or line-to-ground faults for varying operating conditions. Although a drop in accuracy occurred for extreme circumstances, it is impressive that the algorithm identifies faults in such varied operating conditions: 10 to 50 degrees ambient temperature, 200 to 1000 W/m^2 irradiance, 10 to 60 % of mismatch, and 0 to 25 Ω of fault resistance. The feature extraction step was also very impressive, which could be a determining factor in the method's performance. Unfortunately, this work also does not validate results with experimental data and only uses simulation as a source. However, the demonstrated computational performance, both in terms of training cost and utilization speed, its usage without the need for training for parameter tuning, the straightforward implementation, and consistent convergence, suggests the potential for this alternative in the face of other ML methodologies. The authors also emphasize that sparse representation might be utilized alongside different learning algorithms for classification, opening the door to many possible future implementations.

An exciting yet far-fetched proposal was made in [33], where a quantum neural network (QNN) is formulated for PV fault classification. The QNN was trained for predicting just two scenarios: faulty or standard, but required up to four days of training, resulting in 93.89% accuracy. For comparison, the classical ANN took twenty seconds to train and achieved 95.39% accuracy. Although the methodology showcases the potential of quantum computing for this field, its preliminary results still distance itself from the traditional methods.

An abundance of ML methods have been tested and reviewed in this field ([9],[10]), utilizing structures such as SVM, KNN, RF, etc. Nonetheless, the results of [31]-[32] sparked the most interest in this work's scope.

2.4.3 Deep Learning Algorithms

The field of deep learning is a branch of machine learning, with the term "deep" referring to amplified machine learning structures that ought to understand data patterns through more complex and intertwined artificial neuron connections. A simple example of a deep learning model would be the design of an artificial neural network with multiple hidden layers (DNN), with the intuition that each of these "extra" layers achieves feature/pattern recognition in a cascade. Other DL structures include the LSTM, CNN, and RBFNN. They have been explored alongside classical machine learning techniques for PV fault detection, although the known disadvantage is a usually high computational cost and relatively tricky implementation. These techniques are typically applied to image-based solutions [34] since they require classification based on 2D data from various image acquisition equipment [35], [36]. Given the 1D characteristic of raw electrical data, little literature considers these techniques for fault detection, as it implies an extra step of increasing dimensionality. However, there are some promising results in doing so [29].

In [29], not only is a DL technique presented for fault detection and classification, but there is also the best attempt at comparative evaluation against other methodologies. As mentioned, much of the literature presents results solely based on particular datasets comprising simulated noiseless data, invalidating any significant quantitative comparison.

Authors in [37] use a CNN model based on the pre-trained AlexNet for classification and feature extraction, allied with a classical ML model also for classification. The classified faults were arc fault, partial shading, open circuit, and short circuit. While the experiments utilized simulation data, adding noise and an abundance of heterogenous operating conditions better resembled a real scenario. Considering the same noisy data, other tested methodologies present 22-70% average accuracies, with the proposed fine-tuned AlexNet CNN reaching a maximum of 70.45%. This work presents one of the best benchmarks in the literature, with decent coverage of other state-of-the-art ML and DL algorithms, while demonstrating the most realistic results and a sophisticated methodology proposal.

2.4.4 Proposed method's scope

While classical fault detection resides in the synchronous and direct evaluation of state and climate variables, realistic industrial scenarios can have data from various types, sources, and acquisition rates. It's also important to realize that monitoring equipment can register erroneous information, and current communication technology is also susceptible to delays and data loss. With this in mind, recent developments in the intelligent composition of deep learning structures aligned with graph theory spark some interest in their application to this field, such as the new deep learning technique named Cell Complex Neural Networks [38]. The motivation for choosing such a structure comes from its data propagation and consensus capability. The propagation techniques utilized in a CXN appeal to graph theory, dividing a system into other subsystems and components (nodes, also called cells in [38]) that share information. Even if the direct application of this structure might not be feasible or grant better results in the context of fault detection, its modification to meet the scope's needs could result in a robust and efficient solution. Further investigation of this state-of-the-art tool will unroll throughout the development of this work in an attempt to adapt this knowledge to the PV fault detection field.

According to the reviewed methodologies, the proposed tool should pertain to the DL or the hybrid category since, while having a central component of DL, it may also require modeling the PV system's components. The intention of proposing such a novel approach is to contribute to the deep learning methodology ecosystem, explicitly formulated for electrical-based PV fault detection and classification. As mentioned, it aims at an asynchronous and online application, which differs from most current methods and presents a novel DL paradigm considering current knowledge. This work also desires to bring a comprehensive benchmark between popular methods (likewise [29]), utilizing a richer dataset with samples from tangible utility-scale PV assets, allowing accuracy assessment in a realistic scenario.

Reference and year	**Data Source**	**Inputs**	**Proposed methodology**	**Classified Faults (alongside STC)**	**Validation data realism**	**Computational cost**	**Notes**	**Drawbacks**
[29] 2020	Simulated PV System, added noise	Irradiance, Temperature, Short circuit current, Open circuit voltage, PV current, MPP current, MPP voltage, MPP power. Boost converter Maximum current, Voltage and power.	Pre-trained CNN (AlexNet) for feature extraction and classification	Arc fault, Partial shading, Fault during partial shading, Open circuit, Line to line SC	Moderate	High	Resilient against noisy data. Outperforms classical ML methodologies.	Requires data samples from the MPPT boost converter.
[32] 2020	Simulated PV System, no added noise	MPP voltage, MPP current, Short circuit current, Open circuit voltage, Irradiance	Sparse representation classifier	Line to line SC Line to ground SC	Low	Low	Very fast learning speed compared to classical ML structures. Straightforward implementation. Good feature extraction process.	Validation data was very idealistic. Only classifies line to line and line to ground faults.
[24] 2020	PVWatts dataset	MPP voltage, MPP current, Short circuit current, Open circuit voltage, Irradiance, Fill factor, Temperature, Gamma ratio, Maximum power	Graph signal processing	Shading Degraded modules Soiling Short circuit	High	Low	Semi-supervised, allows usage of unlabeled data for training. Better accuracy relatively to other ML methods for less labeled data. Low training cost.	-
[31] 2021			Autoencoder for detection and pruned neural network for classification			Medium	Separate the task of detection from classification, allowing for other combinations. Good performance method considering the algorithms complexity. Pruning creates an ANN less prone to overfitting.	Requires more complex training phase, for two different networks, and utilizing a dropout algorithm for pruning.

Table 2.1: Comparison of literature that inspired this work.

Table 2.1 represents a summary that compares four of the most inspiring reviewed proposals.

Chapter 3

CellTAN Application

3.1 Case study

The experiments validating CellTAN's behavior incorporate two neighboring grid-tied string inverters from the same PV farm with common satellite data. Their only known characteristics are:

- Inverter one: 12.5kW nominal power, 14.4kW peak power. Installed January 1st, 2013.
- Inverter two: 15kW nominal power, 15.84kW peak power. Installed January 1st, 2013.

Variable	Source	Unit	Label
AC side power	Inverter (1 & 2)	W	ac_power
AC side current	Inverter (1 & 2)	A	ac_current
AC side voltage	Inverter (1 & 2)	V	ac_voltage
DC side power	Inverter (1 & 2)	W	dc_power
DC side current	Inverter (1 & 2)	A	dc_current
DC side voltage	Inverter (1 & 2)	V	dc_voltage
Global tilted irradiance	Satellite	W/m^2	global_tilted_irradiance
Global horizontal irradiance	Satellite	W/m^2	global_horizontal_irradiance
Cloud coverage	Satellite	%	cloud_coverage
Air temperature	Satellite	ºC	temperature

Table 3.1: Available variables from two inverters and a satellite.

Table 3.1 represents the available variables and corresponding labels used to identify them in the cell's inputs and graphs. These variables are sampled every 10 minutes from May 31, 2020, at 5:00 am to April 30, 2023, at 7:30 pm (although having some gaps). We utilized data from 2020 until the end of 2022 for the cells' knowledge base, and any information from 2023 onwards is considered new and used for testing. Since there is no production at night, the database does not store values for this period. Not accounting for the night as missing samples, we have around 98% of data availability.

Analyzing and cleaning raw inverter and satellite data is essential to take full benefit of CellTAN's capabilities. As seen in its development, having a clean knowledge base contributes to correctly identifying anomalous situations. Therefore, the following sections focus on these two steps, contributing to understanding the anomalies' domain and frequency of occurrence.

3.1.1 Data analysis

Before data visualization, and regarding the variables in table 3.1, we eliminate those that will not benefit the CellTAN. We determined that AC side voltage is insignificant since the grid mandates it in a grid-tied inverter. We have decided to only use the measure of power instead of using the AC side current measure since, in conjunction with voltage, it provides the same information. To simplify things further, we do not need to consider the power on the DC if considering both the DC side current and voltage measures.

We examine all variables related to satellite data to determine which ones could be useful, not making any premature assumptions.

3.1.1.1 Power

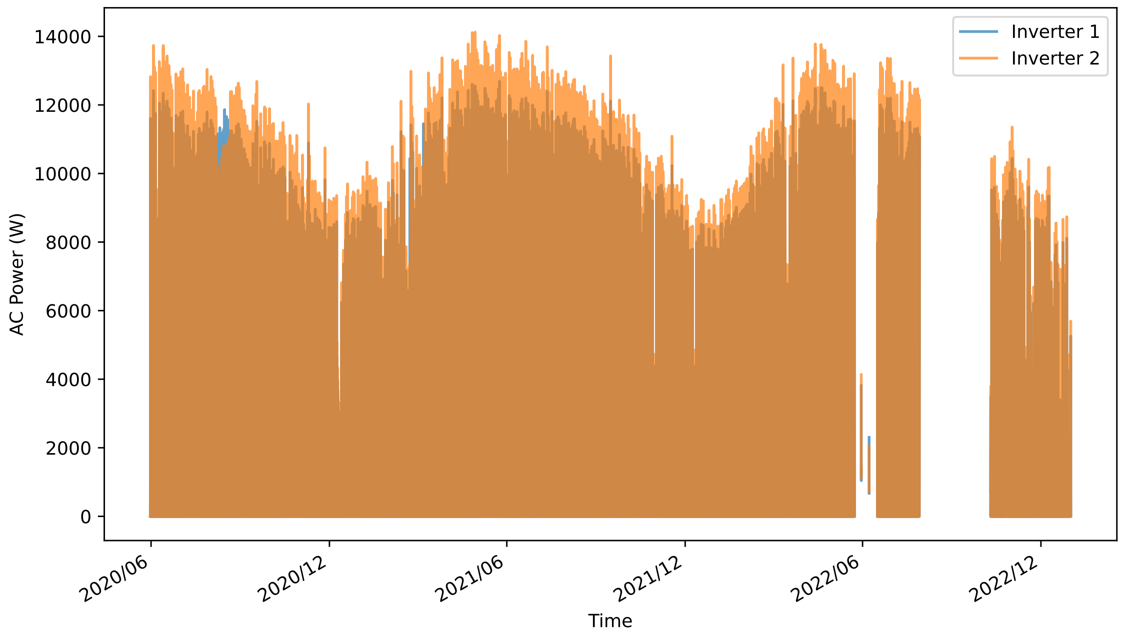

Figure 3.1: Inverter AC side power from 2020 to 2022, used for the knowledge base.

Figure 3.1 shows the power profile of the two studied inverters. Right away, we notice that the power of inverter two caps at around 14kW, while inverter one usually maxes at 12kW. This information is coherent with their ratings. We can notice two relatively large chunks of missing data, with the gaps occurring in mid to late 2022.

Figure 3.2 represents the power profile on the portion of data used for testing. When performing a closer inspection (with more zoom), we could hand-pick some fault occurrences in both

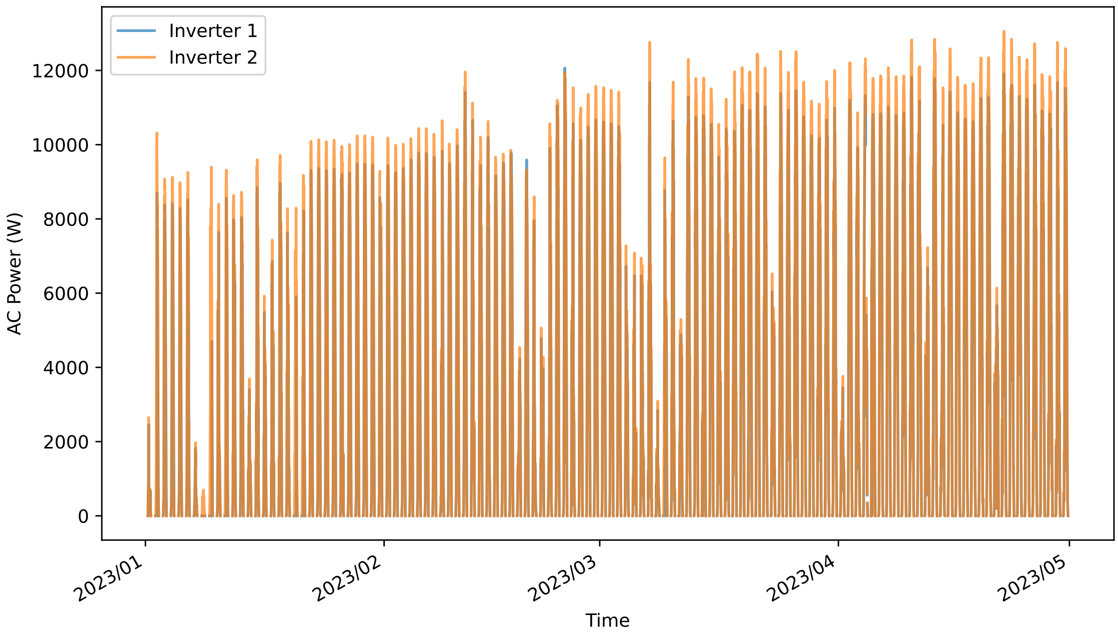

Figure 3.2: Inverter AC side power from 2023-01-01 to 2023-01-05, used for testing.

datasets, with the majority being one inverter off while the other continues regular operation. However, these will be more noticeable during different types of data analysis, such as pair plotting. Regardless, the cases that will matter are in the test data since these scenarios will not exist in the knowledge after cleaning. In Section 3.1.5, you will find a selection of carefully chosen scenarios.

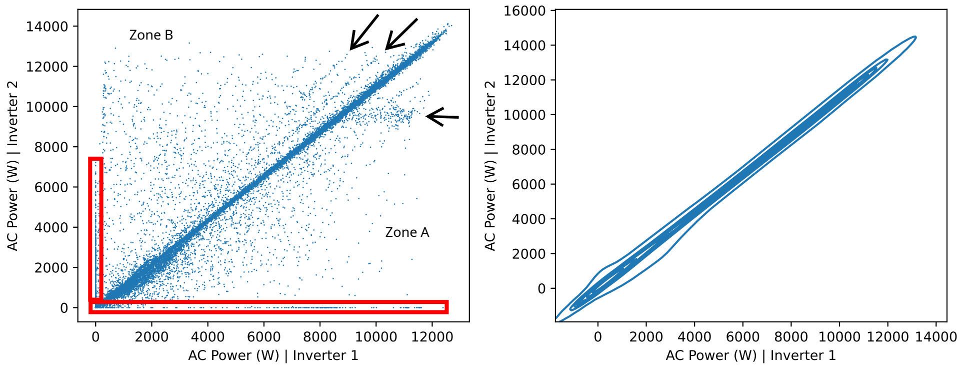

Figure 3.3: Pair plot of AC power from both inverters (2020 to 2022), using scatter (left) and KDE (Kernel Density Estimation) (right).

Figure 3.3 lets us better understand the relationship between the two inverters. As expected, since they are neighboring, they have a strong trend line, leading them to a high Pearson coefficient: 0.97753. However, the noise from outliers is noticeable in the scatter. We define Zone A as the zone where inverter two underperforms compared to one and Zone B as the opposite. The black arrows in the graph display secondary trend lines in Zone B, indicating scenarios of inverter one performing consistently less than expected. Another arrow also points out a cloud in Zone

A (right under the trend line) of the opposite scenario. Furthermore, the red rectangles highlight instances where one inverter was functioning while the other was not. CellTAN must flag these situations, so we should remove them from the knowledge base. The KDE visualization confirms that most samples lie close to the primary trend.

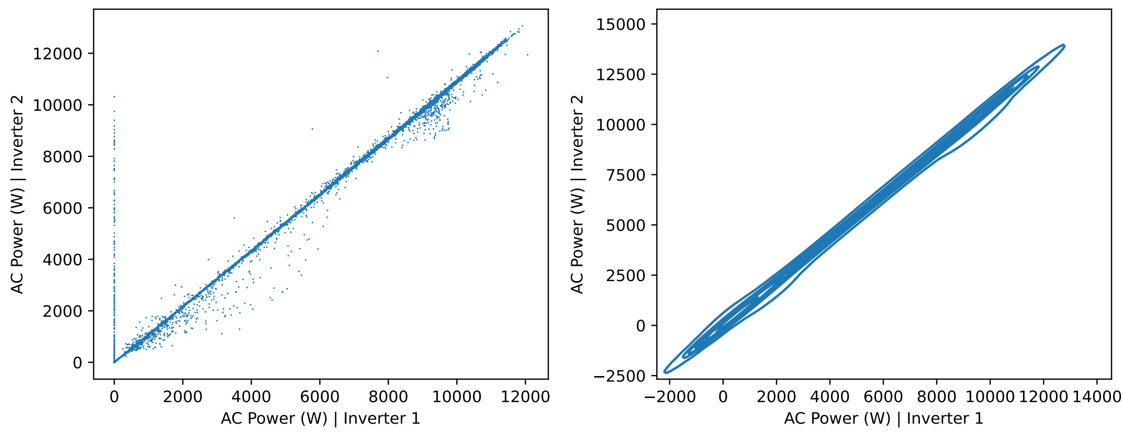

Figure 3.4: Pair plot of AC power from both inverters (2023), using scatter (left) and KDE (Kernel Density Estimation) (right).

From 3.4, it is clear that test data has fewer outliers than the previous. Nonetheless, there are many occurrences of inverter one being inoperational. Besides, there are also a considerable amount of samples below the trend line, meaning the underperformance of inverter two.

3.1.1.2 Voltage and Current

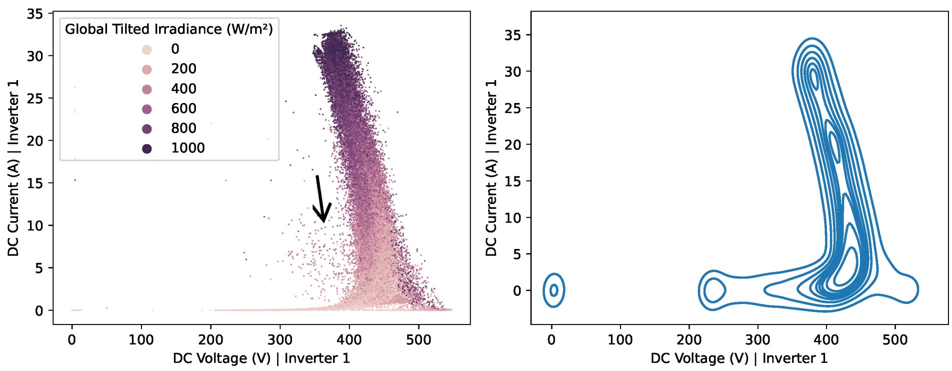

Figure 3.5: Pair plot of DC side voltage and current from inverter one (2020-2022), using scatter (left) and KDE (Kernel Density Estimation) (right).

Regarding DC side voltage and current, figures 3.5 and 3.6 demonstrate the operating range of the inverter's MPPT. Both kickstart production at around 400 V and operate until close to 600 V. Between this range, the central column of samples represents voltage-current points relative to the knee of the strings' I-V curve (see figure B.1). Some instances are outside this normal operation range (outliers), especially in the zone marked by the black arrow. This zone has particular interest

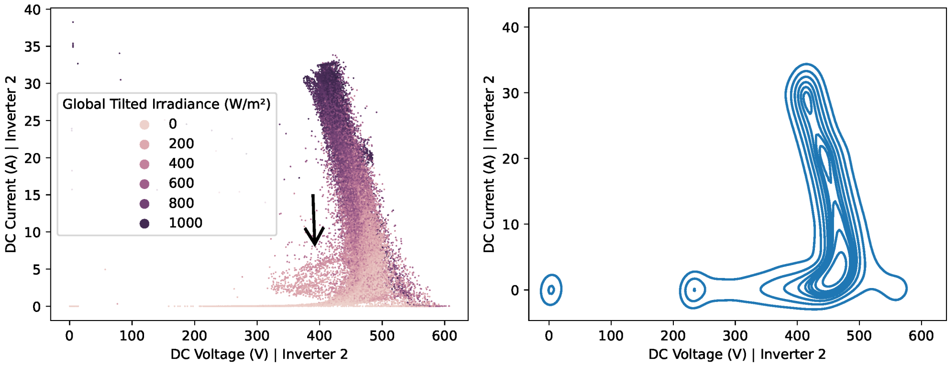

Figure 3.6: Pair plot of DC side voltage and current from inverter two (2020-2022), using scatter (left) and KDE (Kernel Density Estimation) (right).

given that it represents scenarios of underperformance, which could mean the occurrence of faults. It is denser in figure 3.6, meaning that inverter two has more underperforming situations, which was not completely clear from the previous analysis (figure 3.3).

We added to appendix B.2 these same plots but for test data (figures B.2 and B.3).

3.1.1.3 Satellite

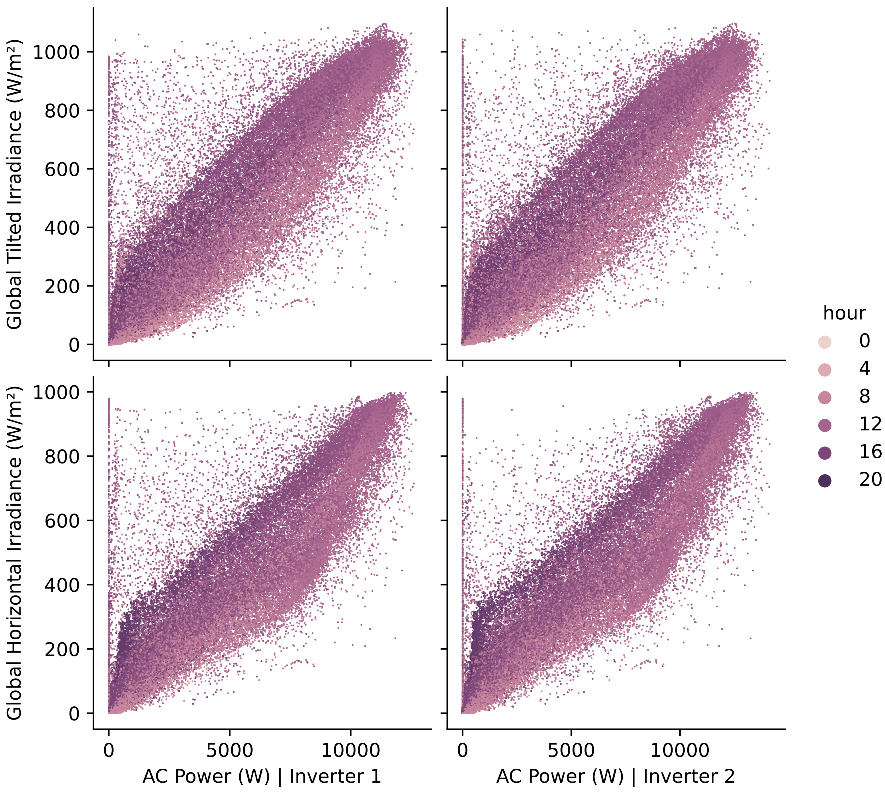

Figure 3.7: Scatter pair plot of the AC power, tilted and horizontal global irradiance for both inverters (2020 to 2022).

Figure 3.7 shows the relationship between irradiance and AC power. We expected a positive correlation, and it exhibits such. However, the large radius around the central trend line demonstrates cycles around it that resemble some kind of hysteresis (especially with horizontal irradiance). By adding a color that displays the hour in each sample, we can affirm that these paths occur due to the fixed nature of the installed PV panels, having a characteristic curve from low to high irradiance in the sunrise and another from high to low during the sunset. Because they produce more with less sunlight in the morning, we can infer that they are oriented slightly towards the east.

Although most instances appear inside the sunrise-sunset paths, there are some outliers. The most notable are the ones of non-zero irradiance with zero production. Either error in satellite data or some anomaly in the inverters causes these odd scenarios, so we target them for cleaning the knowledge base.

Regarding the rest of the meteorological variables (cloud coverage and temperature), we deemed them unnecessary since they do not demonstrate a direct relationship with inverter behavior (see figure B.4). We added KDE visualizations and plots for the test period in appendix B.2.

3.1.2 Data Cleaning

3.1.2.1 Power

...

3.1.2.2 Satellite

...

3.1.2.3 Voltage and Current

...

3.1.3 Photovoltaic Plugin

...

3.1.4 CellTAN Configuration

...

3.1.5 Simulation and Results

...

3.1.6 Scaling up

...

Appendix A

CellTAN Development

A.1 Statistical tests for measure of association

A.1.1 Pearson's chi squared test

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \tag{A.1}$$

where:

χ^2 : Chi-squared statistic

O_i : Observed frequency for category i

E_i : Expected frequency for category i

k : Number of categories or cells in the data

A.1.2 Fischer's exact test

$$p = \frac{\binom{a}{x}\binom{b}{y}}{\binom{N}{n}} \tag{A.2}$$

where:

p : p-value of the test

a : Number of successes in group A

b : Number of successes in group B

x : Number of successes of interest in group A

y : Number of successes of interest in group B

N : Total number of observations

n : Number of observations in group A

A.1.3 Odds ratio

$$OR = \frac{a \cdot d}{b \cdot c} \tag{A.3}$$

where:

OR : Odds ratio

a : Number of successes in group A

b : Number of failures in group A

c : Number of successes in group B

d : Number of failures in group B

A.1.4 Phi coefficient

$$\phi = \sqrt{\frac{\chi^2}{N}} \tag{A.4}$$

where:

ϕ : Phi coefficient

χ^2 : Chi-squared statistic

N : Total number of observations

A.1.5 Contingency coefficient C

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \tag{A.5}$$

where:

C : Contigency coefficient

χ^2 : Chi-squared statistic

N : Total number of observations

A.1.6 Theil's U

$$U(x|y) = \frac{H(x) - H(x|y)}{H(x)} \tag{A.6}$$

Entropy of variable x:

$$H(x) = -\sum_{i=1}^{n} p(x_i) \log(p(x_i)) \tag{A.7}$$

Conditional entropy of variable x given variable y:

$$H(x|y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log\left(\frac{p(x_i, y_j)}{p(y_j)}\right) \tag{A.8}$$

A.2 Technology stack

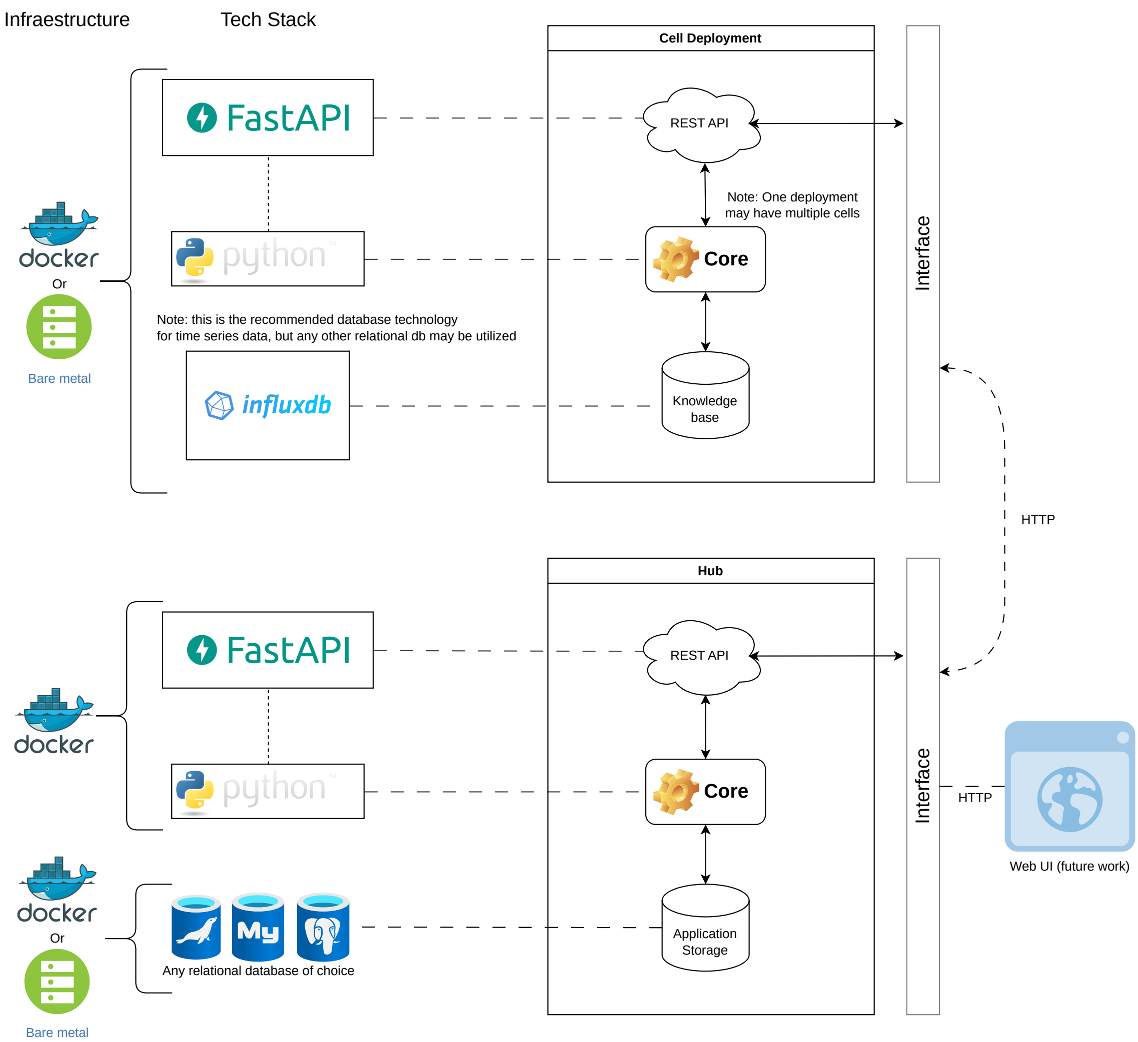

Figure A.1: Technology stack of the Cell and Hub of CellTAN.

A.3 Cell configuration

Appendix B

CellTAN Application

B.1 MPPT Curve

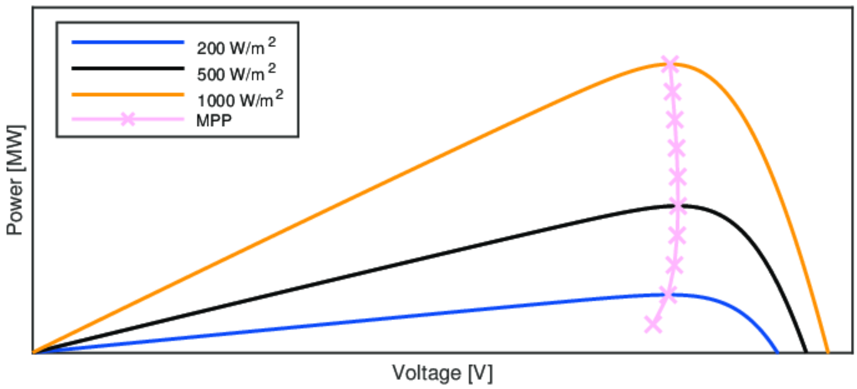

Image source and copyright: [39].

Figure B.1: "PV panel power characteristics as a function of the DC voltage and solar irradiance."

B.2 Data analysis

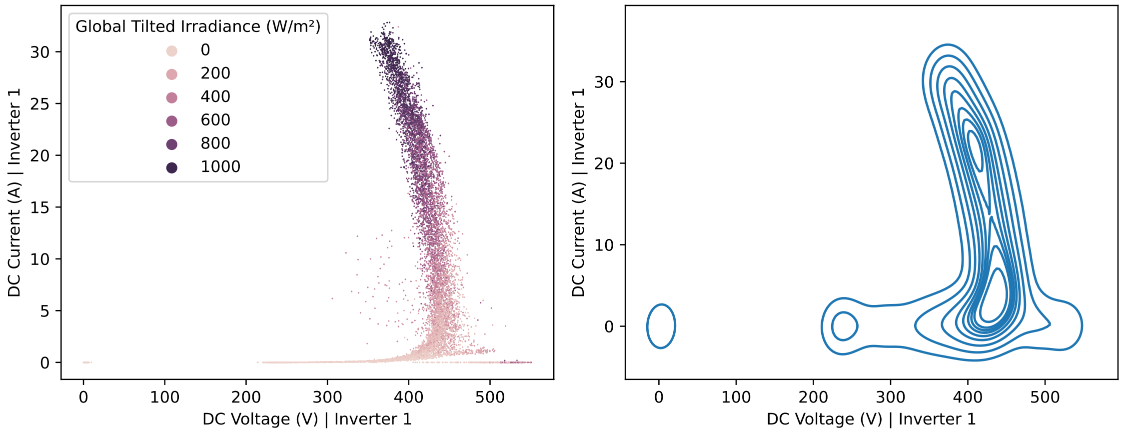

Figure B.2: Pair plot of DC side voltage and current from inverter one (2023), using scatter (left) and KDE (Kernel Density Estimation) (right).

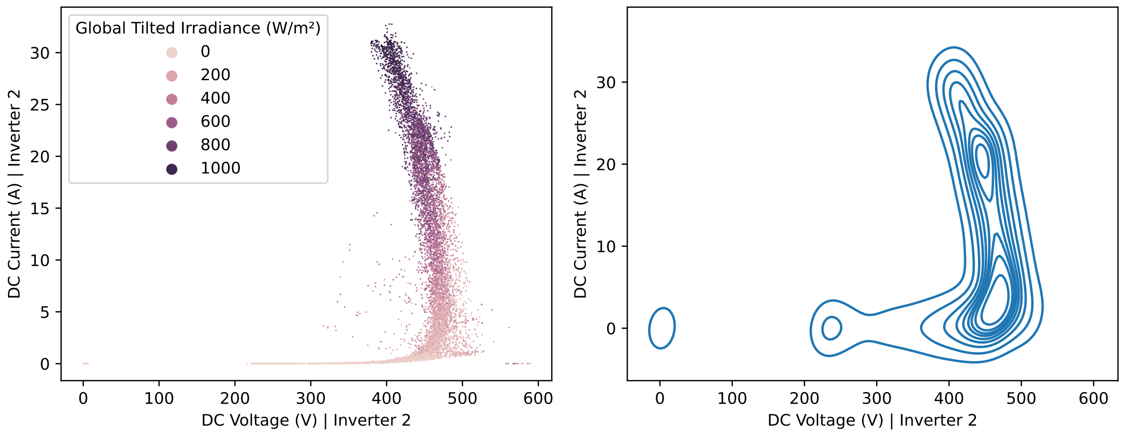

Figure B.3: Pair plot of DC side voltage and current from inverter two (2023), using scatter (left) and KDE (Kernel Density Estimation) (right).

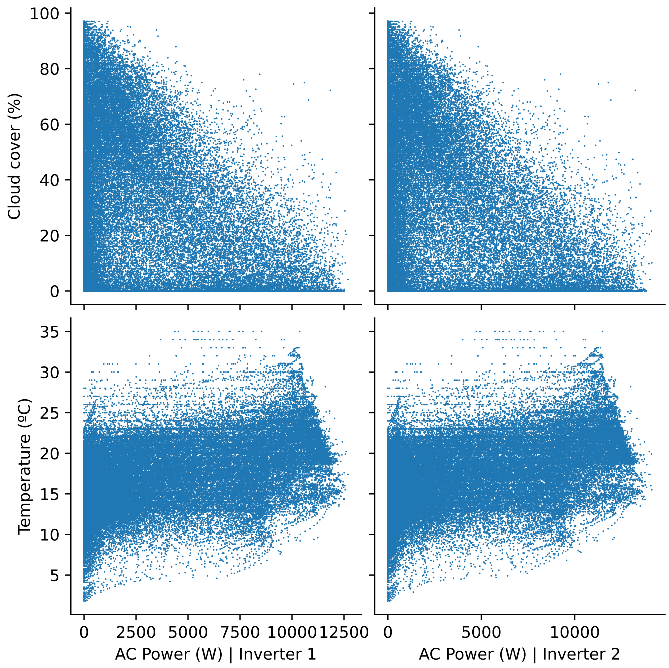

Figure B.4: Scatter pair-plot of AC power from the two inverters with cloud coverage and temperature (from satellite).

References

[1] *Renewable Capacity Statistics 2021.* 2021.

[2] A. Cabrera-Tobar, E. Bullich-Massagué, M. Aragüés-Peñalba, and O. Gomis-Bellmunt, "Topologies for large scale photovoltaic power plants," 6 2016.

[3] A. H. I. Mourad, H. Shareef, N. Ameen, A. H. Alhammadi, M. Iratni, and A. S. Alkaabi, "A state-of-the-art review: Solar trackers," *2022 Advances in Science and Engineering Technology International Conferences, ASET 2022*, 2022.

[4] A. Mellit and S. Kalogirou, *"Handbook of Artificial Intelligence Techniques in Photovoltaic Systems".* Academic Press, 2022. `https://www.sciencedirect.com/book/9780128206416/handbook-of-artificial-intelligence-techniques-in-photovoltaic-systems`.

[5] "Iec 61724-1:2021 rlv | iec webstore." `https://webstore.iec.ch/publication/70170`. Accessed: 2023-01-07.

[6] D. S. Pillai and N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in pv systems," 8 2018.

[7] M. K. Alam, F. Khan, J. Johnson, and J. Flicker, "A comprehensive review of catastrophic faults in pv arrays: Types, detection, and mitigation techniques," 5 2015.

[8] H.-Y. Li, *Assessment of Photovoltaic Module Failures in the Field.* 05 2017.

[9] Y. Y. Hong and R. A. Pula, "Methods of photovoltaic fault detection and classification: A review," 11 2022.

[10] A. Livera, M. Theristis, G. Makrides, and G. E. Georghiou, "Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems," 4 2019.

[11] A. Stepanov, A. Sokolovs, and L. Dzelzkaleja, "Solar tracker supervisory system," *2014 55th International Scientific Conference on Power and Electrical Engineering of Riga Technical University, RTUCON 2014*, pp. 79–83, 12 2014.

[12] T. G. Amaral, V. F. Pires, and A. J. Pires, “Fault detection in pv tracking systems using an image processing algorithm based on pca,” *Energies 2021, Vol. 14, Page 7278*, vol. 14, p. 7278, 11 2021.

[13] H. Braun, S. T. Buddha, V. Krishnan, C. Tepedelenlioglu, A. Spanias, T. Takehara, T. Yeider, M. Banavar, and S. Takada, “Signal processing for solar array monitoring, fault detection, and optimization,” 2011. summary:
chapter 2 describes physics and electrical behaviour of the PV cell
chapter 3 has description of the most common faults
chapter 4 has statistical signal processing-based techniques for determining the presence
of a fault.

chapter 2:
“The most common metric comparing measured module performance to predicted behavior
is the performance ratio”
Modelling performance:
Sandia Model: very accurate (1testing
The five-parameter model: it utilizes only data provided by the manufacturer at standard test conditions and does not require additional measurements to derive parameters.

[14] F. Grimaccia, S. Leva, A. Dolara, and M. Aghaei, “"survey on pv modules’ common faults after an o&m flight extensive campaign over different plants in italy",” *IEEE Journal of Photovoltaics*, vol. 7, pp. 810–816, 5 2017.

[15] “Fire safety for pv systems - sunny. sma corporate blog.” `https://www.sma-sunny.com/en/fire-safety-for-pv-systems/`. Accessed: 2023-01-07.

[16] M. Kumar and D. V. S. K. Rao, “Modelling and parameter estimation of solar cell using genetic algorithm,” *2019 International Conference on Intelligent Computing and Control Systems, ICCS 2019*, pp. 383–387, 5 2019.

[17] R. Godina, E. M. Rodrigues, E. Pouresmaeil, and J. P. Catalão, “Simulation study of a photovoltaic cell with increasing levels of model complexity,” *Conference Proceedings - 2017 17th IEEE International Conference on Environment and Electrical Engineering and 2017 1st IEEE Industrial and Commercial Power Systems Europe, EEEIC / I and CPS Europe 2017*, 7 2017.

[18] N. R. Prasad, S. Almanza-Garcia, and T. T. Lu, “Anomaly detection,” *ACM Computing Surveys (CSUR)*, vol. 14, pp. 1–22, 7 2009.

[19] M. Ilas and C. Ilas, “Towards real-time and real-life image classification and detection using cnn: a review of practical applications requirements, algorithms, hardware and current trends,” pp. 225–233, 10 2020.

[20] S. Buddha, H. Braun, V. Krishnan, C. Tepedelenlioglu, A. Spanias, T. Yeider, and T. Takehara, “Signal processing for photovoltaic applications,” *2012 IEEE International Conference on Emerging Signal Processing Applications, ESPA 2012 - Proceedings*, pp. 115–118, 2012.

[21] Y. Zhao, F. Balboni, T. Arnaud, J. Mosesian, R. Ball, and B. Lehman, "Fault experiments in a commercial-scale pv laboratory and fault detection using local outlier factor," *2014 IEEE 40th Photovoltaic Specialist Conference, PVSC 2014*, pp. 3398–3403, 10 2014.

[22] S. Vergura, G. Acciani, V. Amoruso, and G. Patrono, "Inferential statistics for monitoring and fault forecasting of pv plants," *IEEE International Symposium on Industrial Electronics*, pp. 2414–2419, 2008.

[23] H. Iles and Y. Mahmoud, "Power based fault detection method for pv arrays," *IECON Proceedings (Industrial Electronics Conference)*, vol. 2021-October, 10 2021.

[24] J. Fan, S. Rao, G. Muniraju, C. Tepedelenlioglu, and A. Spanias, "Fault classification in photovoltaic arrays using graph signal processing," *Proceedings - 2020 IEEE Conference on Industrial Cyberphysical Systems, ICPS 2020*, pp. 315–319, 6 2020. "We propose here a graph signal processing based semi-supervised learning technique, which achieves good performance in fault classification with relatively limited data."
Classified faults: standard test conditions (STC), shaded modules, degraded modules, soiled modules, and short circuit conditions.
Also uses PVWatts dataset
MOST SIMILAR WITH THE THESIS.

[25] A. P. Dobos, "Pvwatts version 1 technical reference," 2013.

[26] S. Rao, D. Ramirez, H. Braun, J. Lee, C. Tepedelenlioglu, E. Kyriakides, D. Srinivasan, J. Frye, S. Koizumi, Y. Morimoto, and A. Spanias, "An 18 kw solar array research facility for fault detection experiments," *Proceedings of the 18th Mediterranean Electrotechnical Conference: Intelligent and Efficient Technologies and Services for the Citizen, MELECON 2016*, 6 2016.

[27] S. Katoch, G. Muniraju, S. Rao, A. Spanias, P. Turaga, C. Tepedelenlioglu, M. Banavar, and D. Srinivasan, "Shading prediction, fault detection, and consensus estimation for solar array control," *Proceedings - 2018 IEEE Industrial Cyber-Physical Systems, ICPS 2018*, pp. 217–222, 6 2018. They have sensors per each panel, not realistic for utility scale pv plants
good information about consensus in graphs
consensus is used for determining average values along the entire array, based on each sensor's measurements.

[28] A. Kumari, A. Shekhar, and M. S. Kumar, "An artificial neural network-based fault detection technique for pv array," *2022 2nd International Conference on Emerging Frontiers in Electrical and Electronic Technologies, ICEFEET 2022*, 2022. Uses ANN for classification Classified Faults: short circuit faults and hot spot faults Inputs: dI/dt and dV/dt Results: 98.4accuracy Pros: is that it can use the dI/dt and dV/dt as network inputs to assess faults, yields high accuracy. Cons: uses high frequency data from current and voltage of the panels, sampled every few milliseconds.

[29] F. Aziz, A. U. Haq, S. Ahmad, Y. Mahmoud, M. Jalal, and U. Ali, "A novel convolutional neural network-based approach for fault classification in photovoltaic arrays," *IEEE Access*,

vol. 8, pp. 41889–41904, 2020. "utilizes deep two-dimensional (2-D) Convolutional Neural Networks (CNN) to extract features from 2-D scalograms generated from PV system data in order to effectively detect and classify PV system faults."

"A survey study conducted in 2010 showed that such faults can reduce the generated power of photovoltaic systems annually by about 18.9to noice. Heavy deep learning model.

[30] S. Rao, A. Spanias, and C. Tepedelenlioglu, "Solar array fault detection using neural networks," *Proceedings - 2019 IEEE International Conference on Industrial Cyber Physical Systems, ICPS 2019*, pp. 196–200, 5 2019. Covers a lot of faults: 8 in total.

[31] S. Rao, G. Muniraju, C. Tepedelenlioglu, D. Srinivasan, G. Tamizhmani, and A. Spanias, "Dropout and pruned neural networks for fault classification in photovoltaic arrays," *IEEE Access*, vol. 9, pp. 120034–120042, 2021. Uses an autoencoder machine learning framework
Emphasizes the performance
Autoencoder detects the fault, and neural network classifies it
"We consider the approach of fault detection and classification by monitoring the electrical signals such as maximum power point tracking (MPPT) parameters"
They use the THE PVWatts DATASET
BEST ARTICLE SO FAR!

[32] H. Kilic, B. Khaki, B. Gumus, M. Yilmaz, and P. Palensky, "Fault detection in photovoltaic arrays via sparse representation classifier," *IEEE International Symposium on Industrial Electronics*, vol. 2020-June, pp. 1015–1021, 6 2020. Detect faults: DC short circuit faults of PV array. line to line and line to ground
Has a lot of sources and compares a lot of them!
"The aim of the proposed method is to detect the faults under low-mismatch, low-irradiance and high-impedance conditions."
"The common faults in PV systems are short and open circuit faults as well as panel mismatch and module failures"
Relatively light algorithm that doesn't require tuning.

[33] G. Uehara, S. Rao, M. Dobson, C. Tepedelenlioglu, and A. Spanias, "Quantum neural network parameter estimation for photovoltaic fault detection," *IISA 2021 - 12th International Conference on Information, Intelligence, Systems and Applications*, 7 2021. QNN are still in early stages of development
Not as good as classical Neural Network
Promising for the future.

[34] I. Høiaas, K. Grujic, A. G. Imenes, I. Burud, E. Olsen, and N. Belbachir, "Inspection and condition monitoring of large-scale photovoltaic power plants: A review of imaging technologies," 6 2022.

[35] A. K. V. de Oliveira, M. Aghaei, and R. Rüther, "Automatic inspection of photovoltaic power plants using aerial infrared thermography: A review," 3 2022.

[36] X. Li, Q. Yang, Z. Lou, and W. Yan, "Deep learning based module defect analysis for large-scale photovoltaic farms," *IEEE Transactions on Energy Conversion*, vol. PP, pp. 1–1, 10 2018.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” *Communications of the ACM*, vol. 60, pp. 84–90, 6 2012.

[38] M. Hajij, K. Istvan, and G. Zamzmi, “Cell complex neural networks,” 2020.

[39] A. Lunardi, L. F. Normandia Lourenço, E. Munkhchuluun, L. Meegahapola, and A. Sguarezi, “Grid-connected power converters: An overview of control strategies for renewable energy,” *Energies*, vol. 15, p. 4151, 06 2022.