TÉCNICO LISBOA

Failure diagnosis in photovoltaic systems: a pattern recognition approach based on artificial neural networks

Francisco Manuel Osório Eiras da Costa Santos

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors: Prof. Paulo José da Costa Branco

Eng. Eduardo Abdon Sarquis Filho

Examination Committee

Chairperson: Prof. Célia Maria Santos Cardoso de Jesus

Supervisor: Prof. Paulo José da Costa Branco

Members of Committee: Prof. Rui Manuel Gameiro de Castro

November 2019

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

To my grandmother, Virgínia Santos, for being a shining light and for her guidance through my life.

Abstract

This thesis aims a method of condition monitoring in photovoltaic systems using machine-learning techniques, more specifically artificial neural network learning. First, to approach from a solar cell level, a known model, the one-diode and five-parameter model is used to model it. Follow, an upgraded solar cell model that englobes the previous one but also capable of simulating mismatch faults is used to simulate the whole system. This model is employed to create a database with the following conditions: standard, short circuit and shading. The first mode has no faults, the second has several cells short-circuited, and the third one has some cells under the effect of a lower level of irradiance. Based on this model, input data for the neural network model has been obtained as the voltage at maximum power, current at maximum power, panel temperature and irradiance. Normalization of the inputs has been considered for a more accessible and better convergence and more precise classification results. The neural network is trained to obtain the functional relation between conditions and fault diagnostic preview. Following the training, an outdoor test of a photovoltaic panel is made. The essential inputs are presented to the trained neural network model for it to preview the supposed PV panel conditions, comparing with the real condition. The results achieved proved to be efficient, and it accurately distinguished standard from shading and high short-circuits from the standard condition. Several recommendations are also systematized to improve and develop this method of diagnosis.

Keywords

Condition monitoring, Diagnosis, Photovoltaic, Mismatch faults, Artificial neural network, Machine learning

Resumo

Esta tese visa um método de diagnóstico de comportamentos em sistemas fotovoltaicos usando a técnica de inteligência artificial. Aproximando-se do nível de célula solar, usando o modelo de díodo e cinco parâmetros, outro modelo, capaz de simular falhas de incompatibilidade, é utilizado para criar uma base de dados com os seguintes comportamentos: padrão, onde a condição operacional não apresenta falhas, curto-circuito, onde várias células estão em curto-circuito e sombreamento, onde algumas células estão sob o efeito de menor nível de irradiância. A partir disso, são obtidas as entradas para o modelo de inteligência artificial, sendo elas: tensão na potência máxima, corrente na potência máxima, temperatura do painel e irradiância. Para uma convergência e classificação mais fáceis e com maior precisão, é feita uma normalização dessas entradas. Após o treinamento da rede neural artificial, um modelo matemático é extraído. Um teste ao ar livre de um painel fotovoltaico é realizado com todos os modos. Em seguida, as entradas básicas são apresentadas ao modelo treinado e os comportamentos supostos são obtidos. Uma comparação final entre os comportamentos conhecidos e os produzidos pelo modelo é feita. Isso provou ser eficiente em um primeiro esboço do problema e distinguiu com precisão os modos padrão do sombreamento e altos curtos-circuitos do padrão.

Palavras-chave

Diagnóstico, Fotovoltaico, Falhas de incompatibilidade, Rede neural artificial

Acknowledgements

I would like to thank my dissertation coordinator, Paulo Branco, for accepting my candidature in this theme, my co-coordinator, Eduardo Sarquis, for his never ending patience in discussing my doubts and questions along the thesis and finding an answer to them. Also, to Nuno Melo, for these last months of companionship while we both worked on our thesis.

To Carolina Lopes for her true friendship through these years in college and for life. The second best thing university gave me was the master's degree, the first was her friendship.

For all my closest friends Frederico Arouca, Miguel Pinto, Nuno Diogo, Tomás Santana, I thank you for giving me fun and relaxing ways of distracting me of all pressure college matters.

To my parents, my sister and my brother for all the love, for always supporting me and believing in me this would not have happened if it were not for you.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AC	Alternate current
AI	Artificial intelligence
ANN	Artificial neural network
IEC	International electrotechnical commission
CEC	California energy commission
DAQ	Data acquisition
DC	Direct current
IST	Instituto superior técnico
M&D	Maintenance & diagnostics
LoG	Level of granularity
MPPT	Maximum power point tracker
MSE	Minimum square error
NOCT	Normal operating cell temperature
Nsc	Number of short-circuited cells
Nsh	Number of shaded cells
P&O	Perturb & observe
PC	Personal computer
PID	Potential induced degradation
PR	Performance ratio
PV	Photovoltaic
SC	Short-circuit
SH	Shade
STC	Standard test conditions

List of Symbols

α_{Isc}	Temperature coefficient of short circuit current
$\alpha_{sc,ref}$	Reference value for temperature coefficient of short circuit current
β	Coefficient for the power generated
β_{Voc}	Temperature coefficient of open circuit voltage
$\beta_{Voc,ref}$	Reference value for temperature coefficient of open circuit voltage
Θ_a	Ambient temperature
ΔP_l	Tolerance of the power on the module label
ΔP_m	Uncertainty of the power module measured
ΔT	Increment of temperature
a_1, a_2, a_3, a_4	Fitting parameters
adjust	Temperature coefficient adjustment factor
ce	Cross-entropy
E_g	Energy bandgap
$E_{g,ref}$	Reference energy bandgap
G	Irradiance
G_i	Instantaneous irradiance
G_{STC}	Irradiance at STC
H	Humidity
i	Number of strings in parallel
I	Current
I_{mpp}	Maximum power point current
I_{ph}	Photocurrent
$I_{ph_{cell}}$	Photocurrent of a cell
I_s	Saturation current
$I_{s,ref}$	Reference saturation current
I_{sc}	Short circuit current
I_{sd}	Bypass diode current
j	Number of substrings
k	Number of cells in a substring
K	Boltzmann's constant
M	Mechanical load
n	Diode ideality factor

N_s	Number of cells in series
P_{ac}	Alternate current power
P_i	Instantaneous power
P_l	Power on the module label
P_m	Power module measured
P_{mpp}	Maximum power point
P_{nom}	Nominal power
Psh	Percentage of shading
P_T	Virtual power point
q	Electron's charge
R	Resistance
R_s	Series resistance
$R_{s,ref}$	Reference series resistance
R_{sh}	Shunt resistance
$R_{sh,ref}$	Reference shunt resistance
S	Shading
T	Temperature
t	Target
TC	Thermal cycling
T_{cell_k}	Cell temperature
T_D	Bypass temperature
U	Ultraviolet radiation
V	Voltage
V_D	Diode voltage
V_{mpp}	Maximum power point voltage
V_{oc}	Open circuit voltage
V_{ref}	Operating voltage
V_t	Thermal junction voltage
y	Output

Chapter 1

Introduction

A brief overview of the growth of photovoltaic systems in nowadays energy production scheme. Reasons for its development and foremost problem in maintaining performance over long periods are also brought up in the motivation and scope for the thesis.

1.1 Motivation

Along the years, there has been a shift of paradigms in ways of producing energy, with observations and conclusions that non-renewable methods of transforming energy into electricity (coal, natural gas, nuclear) were polluting our world. Stopping and decreasing pollution increases the switch to renewable energies. Solar energy has seen increased investment in the last few years. It has some unique perks, installation in almost every place due to its flexibility in size, it is flexible in power sizing and noise-free [1]. According to [2], from 2010 to 2017, the installed capacity of solar energy in the world expanded from 50.000 MW to 400.000 MW, a 700% increase, demonstrated in Figure 1.1.
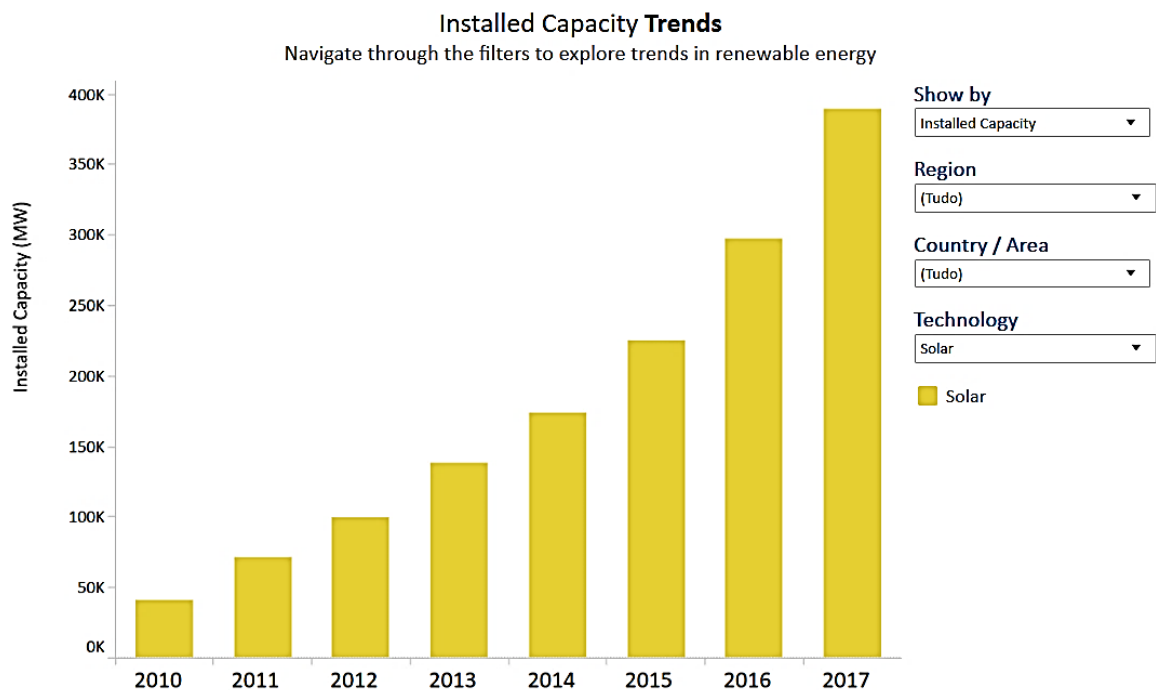
Figure 1.1 - Installed capacity of solar energy in the world [2]

A decrease in photovoltaic panels and equipment costs, special remunerations to renewable energies, more efficient panels all brought investment to this technology. Some disadvantages persist being availability one of the main problems. With all the developments made in new photovoltaic panels, the efficiency of converting energy has increased. However, these were not tested in real conditions, only in ideal ones and not for an extended period of such as their lifespan (25 years). These require a continuous operation to keep availability and compete with other forms of producing energy. When a fault occurs in a system, less energy is produced. Some faults may even damage the panels permanently, and all these flaws reduce the income. Also, there may be abnormal behaviour which is only temporary; however, if not detected, can lead to underperformance and in the future can develop into a fault. To maintain the performance, to the maximum, monitoring of the system is required [3].

1.2 Problem Description

Photovoltaic systems rely on continuous operation when solar irradiation is present, to keep availability to the maximum. If a fault arises a decrease in energy will appear, translating into less profit. To maintain the energy produced by the photovoltaic panels to the maximum monitoring of the system should be implemented and optimized [1].

Monitoring already exists in many forms, for example visual observation where a person inspects visually, only, the condition of the system (Figure 1.2) to detect shading or soiling in the surface of the panels. Another example is thermal observation, where through heat sensors, the panel's temperature, in their surface, is observed and compared to standard limits (Figure 1.3).

Figure 1.2 - Representation of visual observation of photovoltaic panels [4]
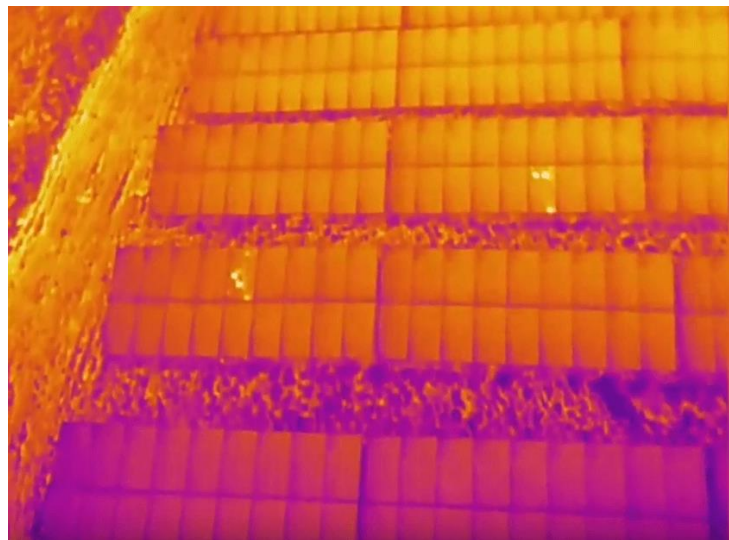
Figure 1.3 - Thermal image of photovoltaic panels [5]

These conventional methods are expensive, not very efficient and only detect significant malfunctions. Minor faults lower the energy causing major malfunctions with time and forcing higher investment for not resolving the fault earlier and with substitution of the equipment instead of repairing the minor flaw. More recently, a comparison between models which calculate, given the proper inputs, the expected energy that the system should be producing and then compare it with the actual energy retrieved.

In this thesis, a method for detection and diagnosis of failures is proposed. It monitors the electrical parameters (voltage, current) and detects if the output power is not in the expected level, then it determines which is the most probable fault. This method relies on the fact of every system now and in the last few years has a digital inverter making it simple to collect the electrical parameters. As can be seen in Figure 1.4, the basic structure of a photovoltaic system, where the inverter is an essential part of it to connect to the grid and AC loads.
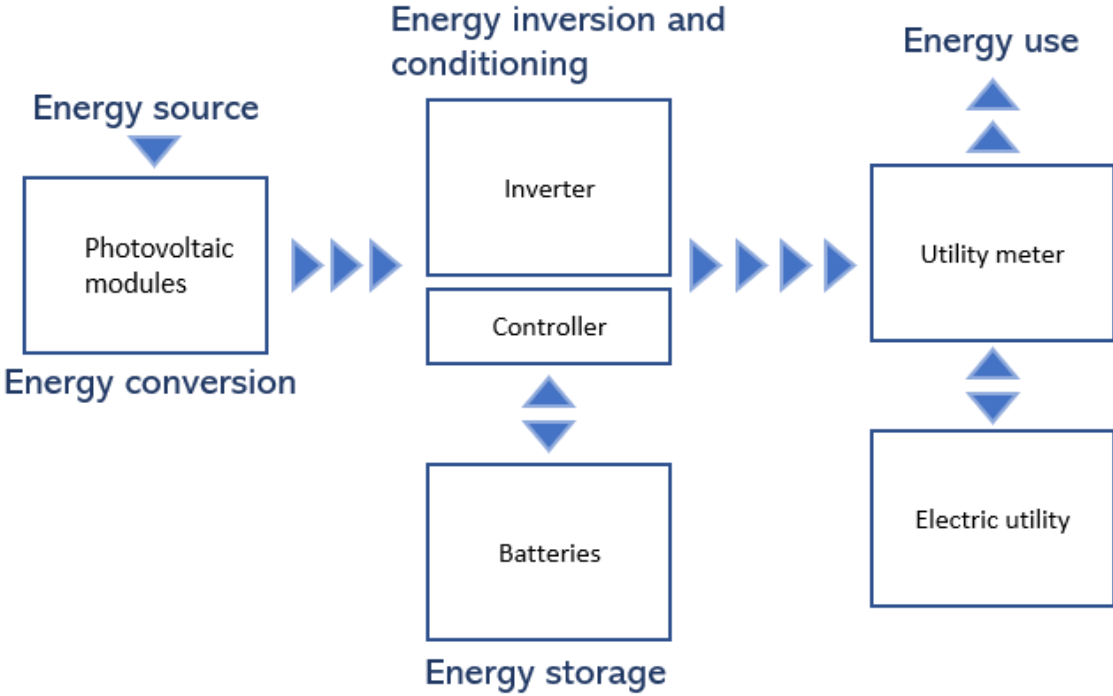
Figure 1.4 - Basic structure of a photovoltaic system [6]

Panels change their behaviour with irradiance and temperature in a way that both are critical in determining variations in power delivered by the panels. Therefore, these variables are inputs in the fault detection method to be developed. Collecting these weather parameters may resort to weather data servers, depending on the country, or to sensors installed near the system.

Collecting electrical parameters and weather data over long periods requires storage and analysis of the data. To determine if the system is working as it should a data processing method will compare typical values of power to the ones observed. If the system is considered to be with fault, then every parameter will be evaluated to determine which fault may be. First using a model developed to simulate

standard and faulty behaviours the values will be obtained and stored to use in the training of an artificial neural network. Through manipulation and optimization of the artificial intelligence method, these behaviours will converge into different patterns.

The method proposed should be able to determine if the system, knowing its topology and specifications, has faults, temporary abnormal behaviour or is working as intended. If the system is faulty, the approach will provide a probability of which fault may have occurred.

1.3 Objectives

The main objective of the work is to develop a method capable of detecting when a photovoltaic system is with a fault or not and give a probability of which faults may have occurred. The goal is to use only simple electrical parameters, such as voltage and current from the inverter and weather data, more specifically temperature and irradiance. These last provided by an online server or weather sensors located in the site of the system.

Specific objectives:

- ❖ Study and comprehension of the most common faults in photovoltaic systems. Observation of the behaviour of the electrical parameters when faults occur.
- ❖ Verify the theoretical knowledge of faults with a simulation model.
- ❖ Validate the simulation model with an experimental procedure.
- ❖ Using the simulation model, construct a database with references of standard, and faulty behaviours.
- ❖ Train an artificial neural network to discover if a system has a fault or not and classify it.
- ❖ Optimize the artificial neural network.
- ❖ Test the artificial neural network in a real system.

1.4 Thesis Structure

This thesis has seven chapters described as follows:

- ➢ Chapter 1 – Introduction: It presented a motivation for this work and a brief problem description.

- ➢ Chapter 2 – Theoretical review: Fundamental concepts related to photovoltaic panels and state of art.

- ➢ Chapter 3 – Experimental procedure: Model utilized in the simulation of the behaviours and experimental procedure related to the interpretation of the model.

- ➢ Chapter 4 – Database construction: Detailed development of the database.

- ➢ Chapter 5 – Artificial neural network: Comprehensive look in the training and optimization of the neural network.

- ➢ Chapter 6 – Testing an artificial neural network: Outdoor experimental procedure and comparison with the results obtained from the neural network.

- ➢ Chapter 7 – Conclusions: Overview of the results and recommendations for future work.

Chapter 2

Literature Review

This chapter provides essential knowledge of previous models and concepts needed to understand further chapters in the thesis. From an electrical point of view and a programming view. Demonstrates a historical evolution of electric measuring methods to give an understanding of the problem itself.

2.1 Fundamental Concepts

2.1.1 Photovoltaic effect

Photovoltaic panels generate DC power with the incidence of sunlight, which causes the photoelectric process. This photoelectric effect consists of the emission of electrons when light is applied to the material. In this specific case, in the PV systems, it is called the photovoltaic effect as the electrons are still contained in the material. In the panels, there is a junction between two types of semiconductors n-type and p-type called the p-n junction. There the electrons will move towards the p-side and holes move to the n-side, then the electrons are caught by the metallic bands in the cell, flowing through them, generating current and voltage. Allowing the system to transform the energy from light into electric energy. In Figure 2.1, the effect illustrates as it occurs inside the PV panel. It can be seen the solar radiation (light rays) hitting the panels and freeing the electrons, though still inside the panel, enabling the flow of current [7].
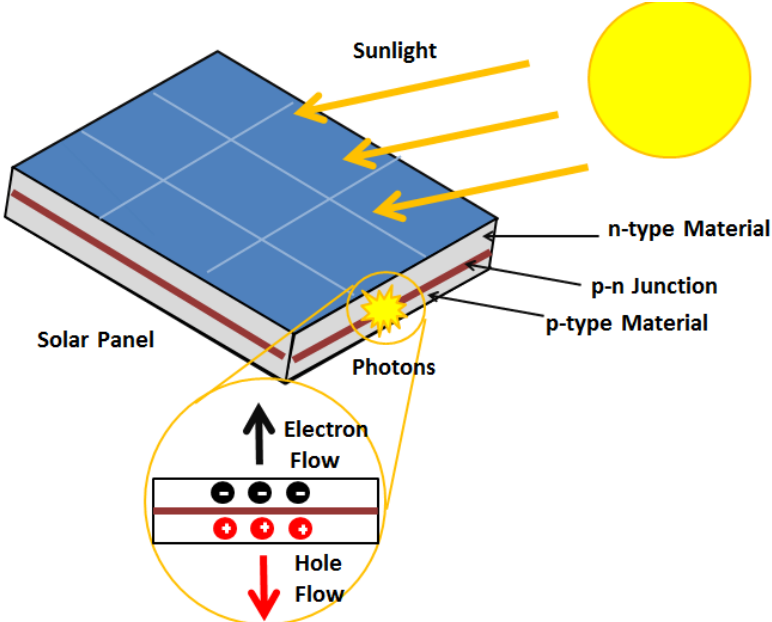
Figure 2.1 – Visual representation of photovoltaic effect [7]

2.1.2 Photovoltaic panel configuration

A brief explanation of the PV panel configuration described and shown for a more straightforward analysis in the following chapters. In Figure 2.2, it shows the necessary structures in which a PV panel can be divided, from the smallest to the panel itself. The simplest element is the cell, which enables the photovoltaic effect. Followed by a module, which consists of cells in series with a bypass diode in parallel. Finally, an array to link all the modules. Bypass diodes help in when an open circuit is detected, or current values determined to be below a certain threshold, allowing the current to flow through the diode. If there were not a diode, there would be no power in case there was no current, or it would produce a low value, as the modules are all connected in series. The array is encapsulated in a hard case with a select glass mirror in front and ethylene-vinyl acetate between the glass and the case to

protect the cells.
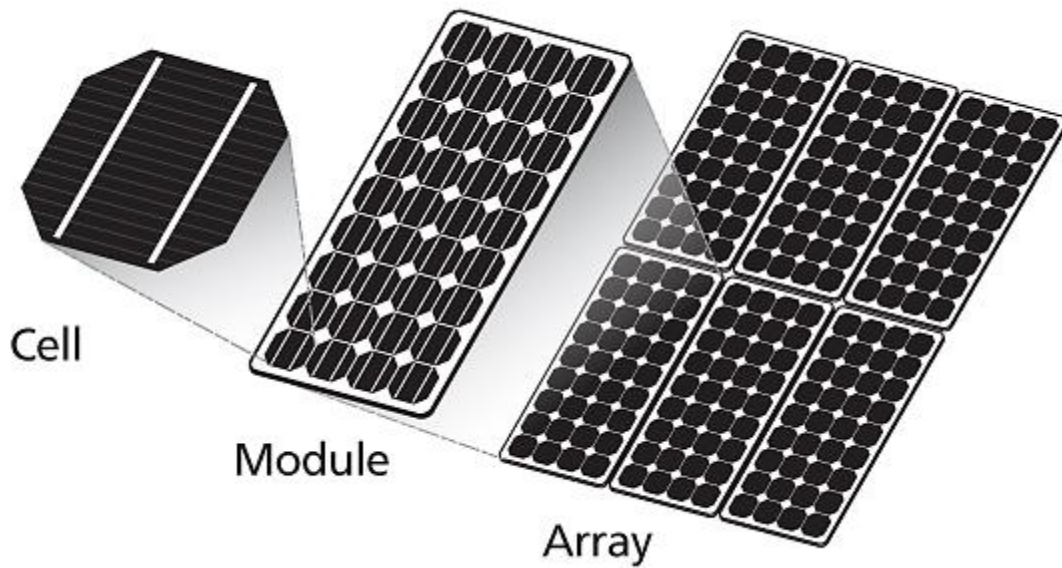
Figure 2.2 - Representation of the basic structures in a photovoltaic panel [8]

Panels have the same configuration; however, they can vary in the number of elements, the number of cells in series can change depending on the panel. In the laboratory the panel has 72 cells in series, meaning the module, or substring has a third of that value, 24 cells.

2.1.3 Inverter configuration

An inverter is required in most systems, to convert DC to AC. For different configurations of PV systems, and considering their power and their application, a specific type of inverter may be used to maximize the power/cost ratio. Commonly types of inverters and their connection to the system are presented for better representation of the problem to the reader. There are three typical inverters:

- Central inverter: connects to the end of the module string of a PV system, displaying the power produced by the whole system. In larger systems is a cheaper option even though a less accurate compared to the rest.
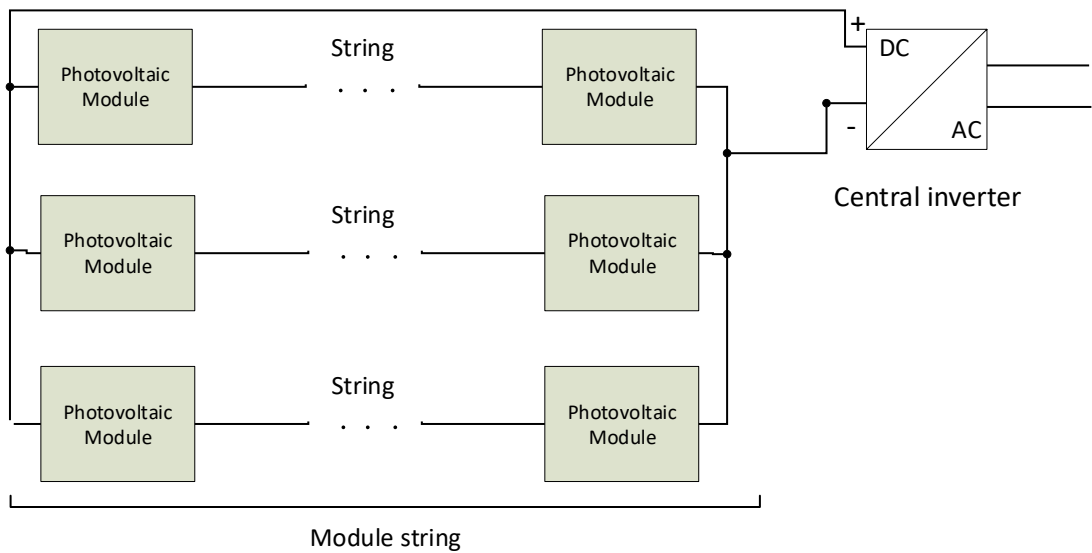
Figure 2.3 – Connection of PV modules to a central inverter [9]

- String inverter: These connect, each to a string of PV modules, showing the power produced by a string.
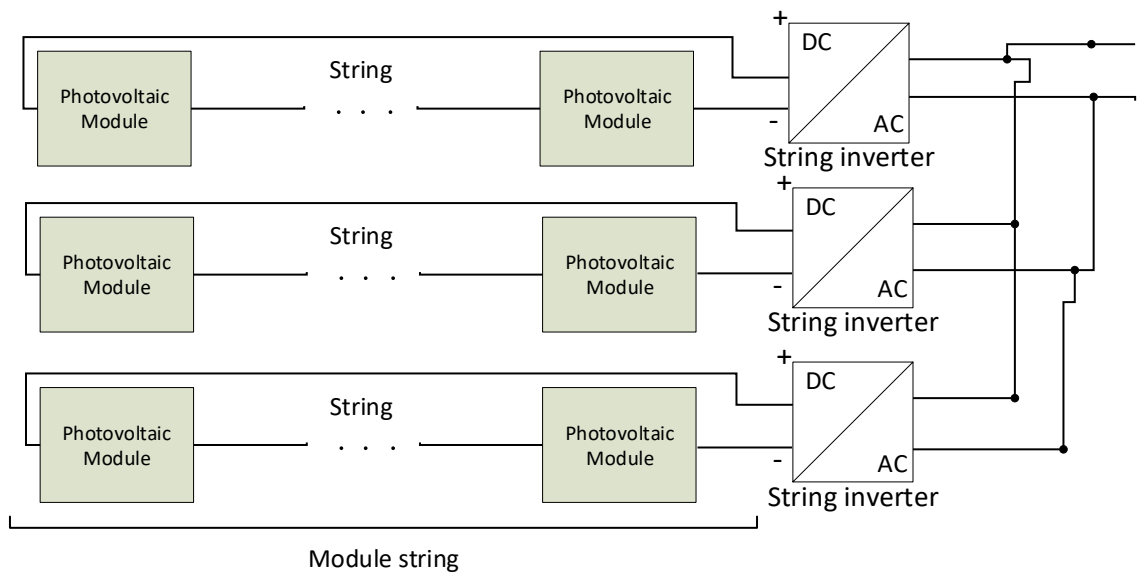
Figure 2.4 - Connection of PV modules using string inverters [9]

- Micro inverter: This type of inverter connects to a single PV module, a higher level of detection however it also has the highest cost, and it increases with the size of the system, more panels translates into more cost.
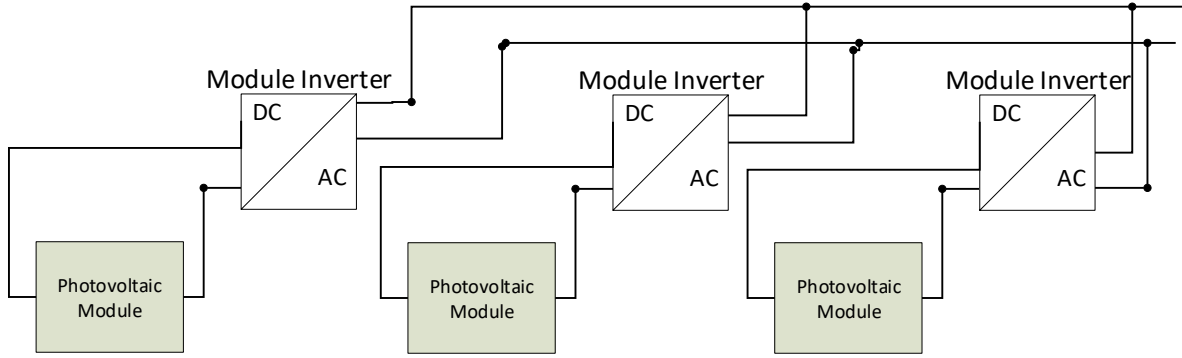
Figure 2.5 - Connection of PV modules using micro-inverters [9]

2.1.4 One diode and five parameter model for solar cell

Representing a solar panel in a simulation environment requires a model for approximation of the characteristics observed in a real environment. Several models have been developed and used over the years [10]. One of the most used is the one diode and five parameter model, due to a balanced ratio between precision and computation cost. Its name is self-explanatory it possesses a single diode, and it has five unknown parameters to be calculated, those being I_{ph}, I_s, n, R_s, R_{sh}. The figure below, Figure 2.6, shows the model, where it can be seen the photodiode current (I_{ph}), which simulates the current obtained from the incident solar irradiance in the cell. The diode simulates the behaviour of the pnp junction. Diode's characteristic equation (1):

$$I_D = I_s \left(\exp\left(\frac{V_D}{n \cdot V_t}\right) - 1 \right) \tag{1}$$

where V_D is the voltage across the diode, V_t is the junction thermal voltage, T is the temperature, k is the Boltzmann constant ($8.617332478*e^{-5}$ eV/K), q is the electron's charge ($1.602*10^{-19}$ C)

$$V_t = \frac{k \cdot T}{q} \tag{2}$$

From the equation of the diode appear two parameters, the diode ideality factor (n) and the saturation current (I_s). In parallel with it is a shunt resistance (R_{sh}). It represents the shunt path for the current flow bypassing the active solar cell, defines the recombination of electrons and holes. A series resistance (R_s) defines the losses in the connections between the metal contact and the silicon, the top and rear metal contacts.
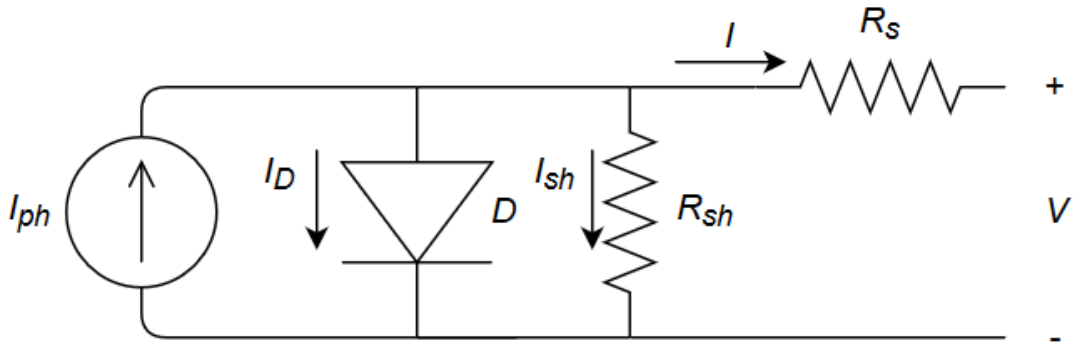
Figure 2.6 - One diode and five parameter model [10]

Defining this equivalent system is the following equation:

$$I = I_{ph} - I_s \left(\exp \left(\frac{V - R_s \cdot I}{n \cdot N_s \cdot V_t} \right) - 1 \right) - \frac{V + R_s \cdot I}{R_{sh}} \tag{3}$$

where N_s is the number of cells in series.

Determining the five unknown parameters (I_{ph}, I_s, n, R_s, R_{sh}) requires knowledge of three major points of the I-V curve, the short circuit, where the current is at a maximum and voltage is zero, the open-circuit voltage, maximum voltage value and current at zero and the maximum power point, defining the correlation between current and voltage that gives the maximum value of power, giving three equations. The remaining two equations are obtained from the derivative of the short circuit condition in order to the voltage (6) and from the derivative of the maximum power point condition in order to the voltage (8). Now having five equations is possible to discover the five unknown parameters:

$$Short\ circuit\ (0, I_{sc}) \rightarrow I_{sc} = I_{ph} - I_s \left(\exp \left(\frac{R_s \cdot I_{sc}}{n \cdot N_s \cdot V_t} \right) - 1 \right) - \frac{R_s \cdot I_{sc}}{R_{sh}} \tag{4}$$

$$Open\ circuit\ (V_{oc}, 0) \rightarrow 0 = I_{ph} - I_s \left(\exp \left(\frac{V_{oc}}{n \cdot N_s \cdot V_t} \right) - 1 \right) - \frac{V_{oc}}{R_{sh}} \tag{5}$$

$$\frac{dI}{dV} \Big|_{\substack{V=0 \\ I=I_{sc}}} = -\frac{1}{R_{sh}} \tag{6}$$

$$Maximum\ power\ point\ \left(V_{mpp}, I_{mpp} \right) \rightarrow$$
$$I_{mpp} = I_{ph} - I_s \left(\exp \left(\frac{V_{mpp} + R_s \cdot I_{mpp}}{n \cdot N_s \cdot V_t} \right) - 1 \right) - \frac{V_{mpp} + R_s \cdot I_{mpp}}{R_{sh}} \tag{7}$$

$$\frac{dP}{dV} \Big|_{\substack{V=V_{mpp} \\ I=I_{mpp}}} = 0 \tag{8}$$

Solving this system of equations analytically can prove to be difficult, preferably there is a better way to solve them, that is, using an iteration method. As stated previously an I-V curve should be presented to calculate these parameters as it will give the short circuit value and the open circuit one. From equations (6,7,8), the parameters (R_s, R_{sh}, n) can be retrieved and after from equations (1,3), I_s, and I_{ph} can be obtained. These require iterations of the equations and to converge in a fast and reliable fashion proper initial values should be chosen.

2.1.5 I-V curve

When talking about solar cells, a usual way of finding out their characteristics and their behaviour is with an analysis of the I-V curve. A typical I-V module measurement system consists of a light source, to activate the cell's capability of transforming light irradiation in electricity, an external load that is forced to vary from zero to a high value, a temperature gauge and an acquisition system to retrieve the values of voltage and current. From this curve, three significant points are retrieved and identified, the short circuit current (I_{sc}), the open-circuit voltage (V_{oc}) and the maximum power point (V_{mpp}, I_{mpp}). The short circuit current is the current that passes through the cells when the voltage is close to zero, and the open-circuit voltage is the maximum voltage from a PV module when the current is zero. To obtain an I-V curve an external load should be put to zero, and from there the short circuit current is known, then, with the variation of the external load to a high value, the curve will be drawn until it reaches the open-circuit voltage. An example of an I-V curve is shown below in Figure 2.7.
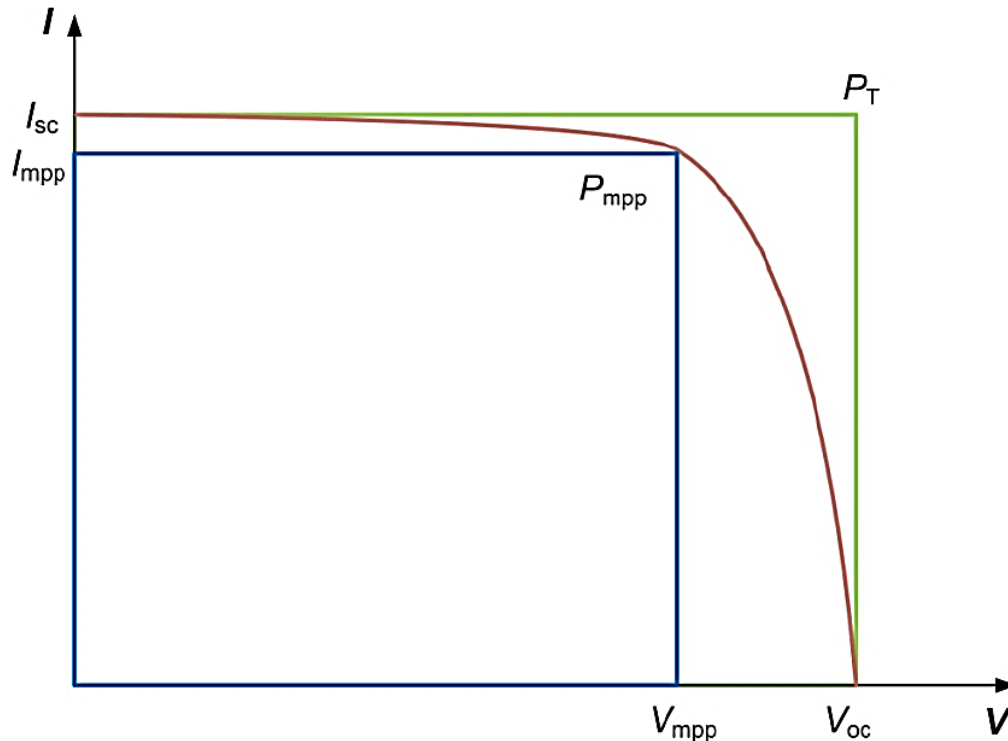
Figure 2.7 - I-V curve with the most important parameters: I_{sc}, I_{mpp}, V_{oc}, V_{mpp}, P_T, P_{mpp} [11]

Going forward, I-V curves will be a staple in reaching conclusions and explaining processes, so now

a brief explanation of the curve itself will be given.

In the curve, there are two values that each defines a slope, one from the zero voltage to the maximum power point and another from the maximum power point to the open-circuit voltage value. The first one changes with the shunt resistance (R_{sh}), an increase in this resistance brings the slope to a more horizontal configuration. The second slope changes with the series resistance (R_s), an increase in this value, diminishes the slope thus reducing the maximum power. In Figure 2.8, a demonstration of what happens when variating both resistances is shown.
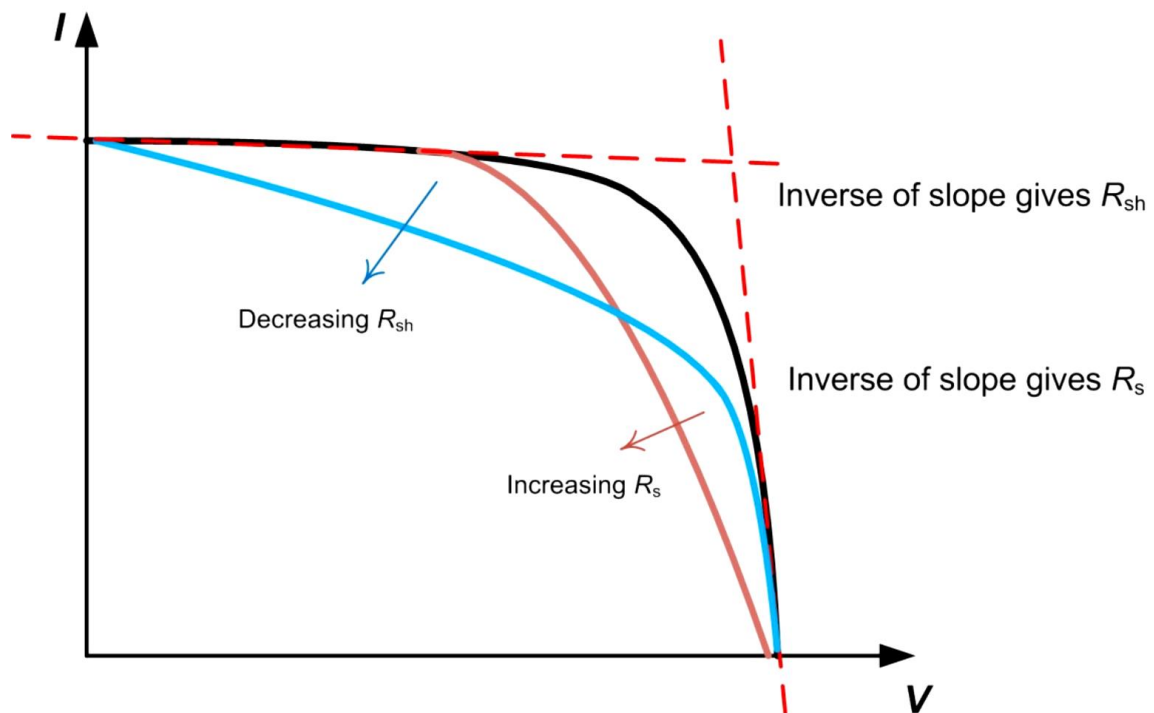
Figure 2.8 - Influence of R_s and R_{sh} on an I-V curve [11]

The utilization of series resistance and shunt resistance is only valid if the cells have the same manufacturing process and parameters [11].

2.2 State of the art

2.2.1 Overview

A PV system may present malfunctions over the course of its lifespan. These can occur anytime and are divided as electrical failures or mechanical failures. Faults of that kind will cause a reduction of the output power resulting in economic losses because of the lower power generated compared to the expected or if needed the replacing of the equipment. Other reducing power problems may be temporary and won't damage the structure of the system, however if not timely and adequately attended may

originate major malfunctions. All can damage the system over the course of time, preventing the system of achieving its maximum performance and to be reliable as an energy source of electricity. Other energy sources such as coal have much higher reliability as when energy is requested the coal power plants can meet the demand with very fast response time. Solar PV only produce energy during daytime, they need solar irradiance to be at a level only reachable at daytime and with solar exposure, allowing the transformation of solar energy to electricity by the solar cells. Thus, they need to be available to collect the energy from the sun and inject it directly to the load or to store it in batteries. For all reasons stated reliability and performance need to be improved or maintained at the highest, a better maintenance and fault diagnosis needs to be implemented for it to happen. Reducing the economic losses will be a priority thus monitoring PV systems with a fault diagnosis method can help to track faults, localize them and diminish their repercussions in the electrical parameters. Gaining knowledge of faults would help maintenance in tracking the malfunction and access the best option to resolve it. For difficult places to reach, panels installed in the city are usually located in the rooftops, the maintenance crew could be called only when needed resulting in a lower economical stress and a better scheduled maintenance. Other adversities resulting in a permanent damage to the system will tell the breaking point where it is necessary the replacement of the damaged equipment.

A brief history of monitoring in photovoltaic systems is described to understand the evolution of monitoring, investigation errors made, and breakthroughs discovered.

In this chapter will be revised studies mainly using system's output power considering weather conditions, observing the voltage and current being produced by the system, to detect if the system is running at its standard efficiency. It is also essential to collect and analyse each system data, so abnormalities can be studied and recognized earlier. Allowing to construct a proper database for different faults caused by various factors, improving the detection time of a flaw and the probability of the type of fault. In panels developed in the next years, with these databases, they can be improved to be more resilient to these faults studied and documented.

2.2.2 Definition of fault

Firstly, a look at the known faults and their complications is analysed thoroughly to understand which constitute a real problem to PV systems and which can be neglected. A definition needed to start the analysis is to define what is considered a fault in a PV system, it is an effect that lowers the power delivered by the module which cannot be reversed by normal operation. Autor's differ in the percentage of reduction in the output power to be considered a fault, which can be related to the errors on the measuring devices utilized. The definition of power loss failure is, according to [11], if the power module measured P_m, based on IEC 60904, plus the uncertainty of the measurement ΔP_m, which depends on the measurement error of the equipment used, is lower than the power on the module label P_l minus the tolerance stated on the label ΔP_l, shown in equation 9, a failure exists.

$$P_m + \Delta P_m < P_l - \Delta P_l \tag{9}$$

There are occurrences where the output power decreases temporarily, as stated before it does not

qualify as a fault, it is seen as a failure of planning or maintenance. Still it should be detected to allow a corrective planning. Also, the modules can have malfunctions in their 'early life', during their first hours of functioning, due to manufacturing mistakes. In some reports such as [11] a defect is considered to be a broader definition, including the definition of failure and also temporary power losses.

2.2.3 Monitoring brief history

Monitoring started with day-by-day observations of the power output produced, where comparisons between days were made. It proved to be a not very reliable way of determining if the system was performing well. It did not consider the temperature and irradiance and from day to day they can change very significantly. After came the simple measurements of common electrical parameters, it was straightforward as it only compared the difference between measured parameters and predictions from a base model [12]. This method proved not to be enough, unexplained power losses needed to be better studied and understood. Models were developed to estimate the output power given the parameters of the PV module (I_{sc}, V_{oc}, R_s, R_{sh}, I_s), called one diode and five parameter model, this is considered to be the most used model given its relation between fast computation and overall accuracy [13]. One of the many unknowns was when should the measurements be taken and how should the comparison with the known model be made, i.e. if it were to compare a PV panel under total shading conditions with the model it would imply the breakdown of the panel and substitution of it. Now can be understood that a monitoring system should have time intervals to observe the output of the system and determine the cause of the problem. Many solutions have been proposed, aiming in a real time prediction of the system's output power, based on weather conditions, and comparing the model's output power with the actual one measured from the system [14]. The problem resides in determining what caused the difference between the calculated by the model and the measurements. A suitable monitoring system should answer the following questions: what should be measured, how to measure it and how to handle those measurements. The first question relates to the system overall configuration and variables to store. About its configuration, it should be made clear the number of panels in a string as string inverters are the most common. In this thesis it was chosen the string inverter due to it being the most used in systems.

The overall performance of the whole system depends on the performance of each subsystem, where it reaches the most basic structure the cell. Here dwells on the overall accuracy of maintenance & diagnosis and its expense, increasing the first also increases the latter. An initial classification was created for rough levels of M&D electrical techniques based on "level of granularity". Lowest LoG is a more rustically monitoring, where the system is considered equal, retrieving only measurements in the inverter and higher LoG start to have more specific and detailed monitoring [12]. The simplest trends in monitoring are based in checking only the instantaneous power generated by the PV system, at either the DC side or the AC side. The power is measured and converted into energy produced by the system. To understand how the system is performing, a ratio was developed, called Performance ratio, PR. It correlates the measured instantaneous power, P_i, and the nominal power, P_{nom}. Also, to correct the ratio it is multiplied the standard test condition irradiance, G_{STC}, over the instantaneous irradiance, G_i[12].

$$PR = \frac{P_i}{P_{nom}} \frac{G_{STC}}{G_i} \tag{10}$$

The ratio gives an estimate of how the system is performing however it fails to consider the temperature, a factor decisive in the output power. Having that in mind a new ratio was proposed to support temperature:

$$PR(T) = \frac{P_i}{P_{nom} + \beta \cdot \Delta T} \frac{G_{STC}}{G_i} \tag{11}$$

A coefficient for the power generated (β) and an increment with respect to 25ºC (ΔT) to the formula. Even with these alterations it was not enough to produce accurate approximations, for the ratio used P_{nom} which does not consider several factors leading to a deviation from the actual performance of the system. So, a new model was proposed to take into account the difference in nominal power, changing P_{nom} to P_{ac}.

$$P_{ac} = G_i(a_1 + a_2 G_i + a_3 \log(G_i))\big(1 + a_4(\Delta T)\big) \tag{12}$$

Where a_1, a_2, a_3 and a_4 are fitting parameters, calculated for each installation [12].

Along the years there has been an increase in PV system's average *PR* values, demonstrated by studies [15], represented in Table 2.1.

Table 2.1 - Evolution of PR from the 80s until the beginning of the 21st century [15]

Installed	Location	Range of *PR*	Avg. *PR*
1980s	Worldwide	0.50 – 0.75	Individual estimates
1990s	Worldwide	0.25 – 0.90	0.66
1990s	Worldwide	0.50 – 0.85	0.65 – 0.70
1990s	Germany	0.38 – 0.88	0.67
2000s	France	0.52 – 0.96	0.76
2000s	Belgium	0.52 – 0.93	0.78
2000s	Taiwan	<0.3 – >0.9	0.74
2000s	Germany	0.70 – 0.90	0.84

The early PV systems presented large variations between the actual values and the nominal provided by the manufacturer's datasheet. In addition, incorrect installation, ineffective or non-existing MPPT tracking. With the adaptation of monitoring systems, important data could be collected and used to comprehend the losses and common problems of PV systems. This understanding helped improving the maintenance of PV panels. Well maintained systems demonstrated higher PR values.

Monitoring of electrical parameters required storage and analysis of large amounts of data. This motivated the adaptation of advanced data processing techniques such as artificial intelligence and big data to better extract useful information [12]. Artificial intelligence and data mining can be split into three categories: signal processing, classification and inference. Signal processing extracts characteristics of the signals can be given a determined state of health. For this technique normally it is applied a wavelet transform or a fast fourier transform. The classification is done by artificial intelligence, from the training dataset the algorithm can learn and draw knowledge from it. As the database grows larger the learning algorithm can learn the behaviour and nuances of the overall system [12]. Papers and reports enter this field of expertise and try using different methods of AI [16][17][18][19].

2.2.4 Faults in photovoltaic panels

Faults in photovoltaics can be classified, in a first approximation, into two types: irreversible, caused by electrical or mechanical problems. Enumerating some of these are short circuits, open circuits and aging of the panels. Other type of faults are the reversible ones, which consist in temporary power losses mainly caused by shading or soiling [20]. Another used approach to categorize faults is dividing them in lifetime wise of the panels. Three categories: infant-failures, midlife-failures and wear-out-failures, presented in Figure 2.9. As the name indicates the infant ones occur in the beginning, after installation, these are more prone to happen. Midlife failures are not so usual compared to infant the latter, the end if life failures [11].
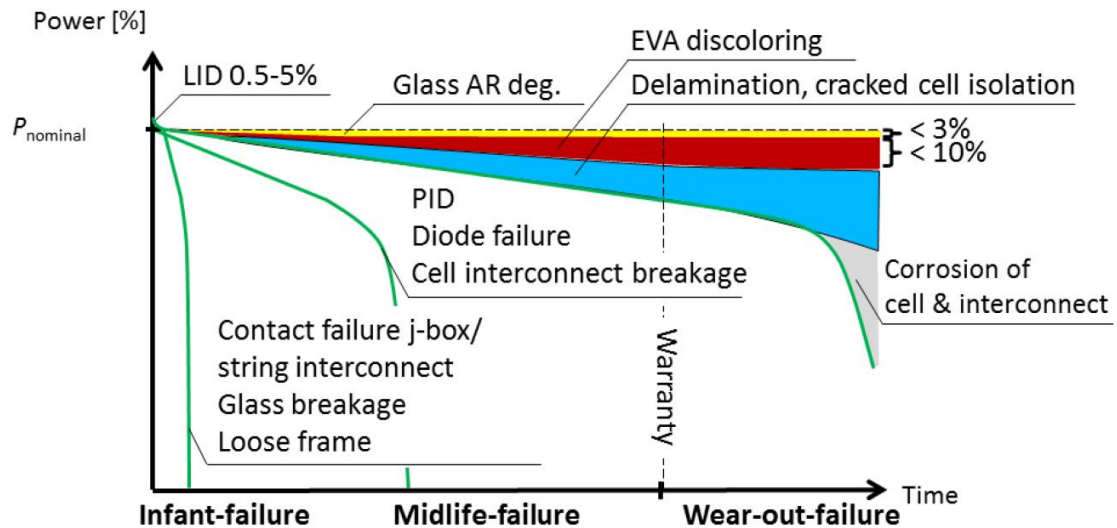
Figure 2.9 - Typical failure scenarios for wafer-based crystalline photovoltaic modules [11]

Studies have been made, to detect and identify faults and their respective location in photovoltaic systems. These are three irreversible different categories of faults considered in reports and papers:

- Ground faults which occur when an inadvertent contact between a conductor and the ground or equipment frame is established. Frequently the result of insulation breakdown.
- Line-to-line faults occurring when two lines at different potential establish a connection.
- Open circuit faults happen when the electric connection suffers an opening, disrupting the electric current within that path.

Ground faults are easier to detect using protection devices and for that reason they will not be addressed in this thesis. The focus will be in mismatch faults. Now that some categorization has been presented, there will be a more detailed explanation of the most common faults that can happen in a photovoltaic system and what the studies say about their effect in the electrical parameters and the performance of the system itself. First the flaws caused by external sources such as: glass breakage, usually caused by clamping or physical impact, causes loss of performance due to bad enclosure as oxygen and water vapor enter through the breaches corroding the cells and the electrical circuit. Also, the damage done to the glass will diminish the irradiance absorption from the cells. Likewise transport and installation if not properly done can cause this irregularity and others. Lightning strike can interfere with bypass diodes function and provoke subsequent safety failures. Not all power losses behave in the same manner and as some are critical to correct others are not so in the immediate time. A definition of power loss categories is presented in Table 2.2. These different categories explore the various variations seen in the power curve over time.

Table 2.2 - Categories of power loss [11]

Power loss category	Description
A	Power loss below detection limit <3%
B	Exponential-shaped power loss degradation over time
C	Linear-shaped power loss degradation over time
D	Power loss degradation saturates over time
E	Degradation in steps over time
F	Miscellaneous degradation types over time

Many factors impact the behaviour of the system, in Table 2.3 the ones more relevant for power loss calculations are presented.

Table 2.3 - List of possible dependencies of power losses [11]

Representative symbol	Power loss factor
T	Temperature
G	Irradiance
I	Current
V	Voltage
H	Humidity
M	Mechanical load
U	UV radiation
TC	Thermal cycling
S	Shading

2.2.5 Artificial neural networks (ANN)

Human brains can process a huge data of information, highly complex and non-linear. They can think differently from a computer and can recognize patterns immediately from previous experiences, such as recalling a person's voice or face. In the Figure 2.10 below is represented the mathematical description of how a basic neuron in an ANN works.
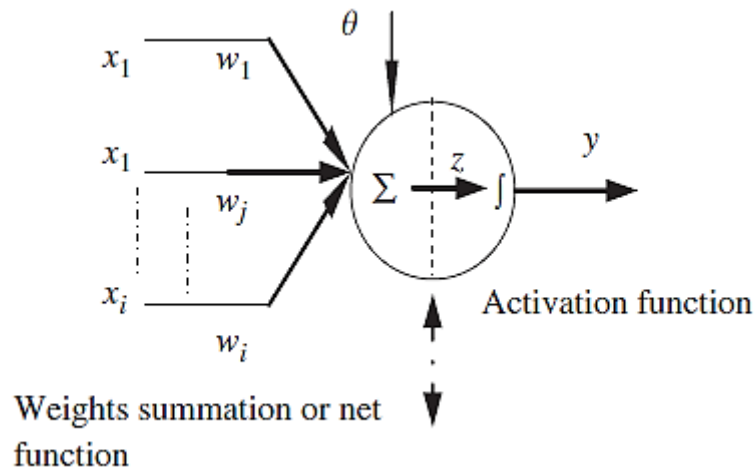
Figure 2.10 - The basic neuron [21]

Artificial neural networks try to mimic this process. An ANN is a machine or algorithm designed to model the way the brain performs, they are not programmed to execute specific tasks instead they are trained with data sets until patterns are constructed from known inputs and outputs [21]. They can be implemented using electronic components or digital programming. It consists primarily in these concepts:

➢ A processor with several units of processing, interconnected, which has the capacity of storing experimental or simulated knowledge and make us of it when needed. It presents several stages:

- Knowledge: it must be taught, using data known and tested, it is given to the neural network the inputs and the known outputs so it can extract a pattern from them. This is described as the learning process.
- Connections between the neurons, named weights, and bias are utilized to gather the acquired knowledge. These will be optimized through the learning process to be able to reproduce the pattern in the most suitable way possible.

This technique is mainly used to recognize patterns, classify data and forecast future events. Its behaviour is defined by the individual elements and the way they are connected and by the weight and bias of their connection. How do they work?

- Inputs are chosen and given to the artificial neural network.
- Input layers are equal to the number of inputs given.
- The number of hidden layers is chosen. In theory it is $\frac{2}{3} Input\ layers + \frac{1}{3} Output\ layers = Hidden\ layers$ however in practice it should be tested for several until the best solution is found.
- Output layer comes last and defines the number of outputs wanted.

A visual representation of an ANN is shown in Figure 2.11.
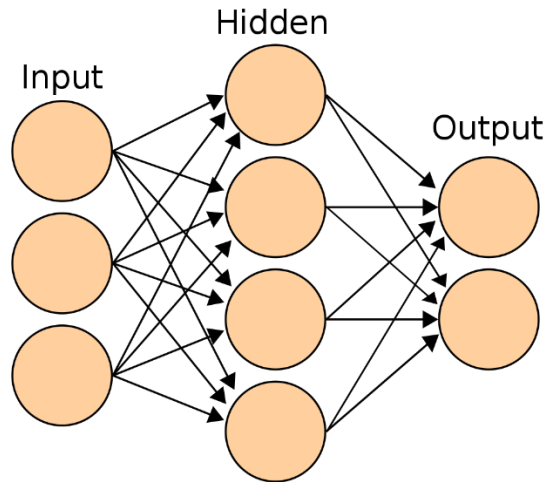
Figure 2.11 - Example of an artificial neural network [22]

A neural network needs a learning process, where data known is given to it so that it can assimilate the pattern. First the inputs must be chosen, this will give the number of input nodes in the input layer. After the number of hidden layers is decided, usually only one hidden layer is applied. The increase of layers will augment the complexity of the ANN and its computational cost. To finalize the process the number of outputs is decided, and for each one an output node is added. Initiating the learning process random or pre-determined weights, in the connections between the input nodes and the hidden layer, are fabricated. Each connection will have its weight and it will multiply that number by the passing value. On the hidden layer an activation function exists to approximate the values of an output, usually sigmoid or hyperbolic tangent. A bias is also applied in each hidden layer node and in the output nodes. Again, exist weights amid the hidden layer and the output nodes connections. An output for each node is reached and compared to the actual output known. Subsequently there is a function that will revise the weights and biases for a better approximation to the real value, usually the backpropagation function. In the next iteration with other values the weights and biases will be different having been changed by the backpropagation function. It computes the input $p_j(t)$ to the neuron j from the outputs $o_i(t)$ of predecessor neurons with the sum of the bias ω_{oj} [23].

$$p_j(t) = \sum_i (o_i(t) \cdot \omega_{ij}) + \omega_{oj} \tag{13}$$

It can be described as mathematical functions, $f(x)$ that can be viewed as a composition of other functions $g_i(x)$, and this also can be decomposed into other functions. Normally, is used the nonlinear weighted sum,

$$f(x) = K\left(\sum_i \omega_i g_i(x)\right) \tag{14}$$

where K serves as the activation function. It provides a smooth transition as input values change so do the output, however they do not suffer sudden impactful changes, meaning small input changes translate to small output changes [23].

Deep neural network is one specific type where the methods for the learning process can be several such as stated in [24]:

- Supervised: trained to produce desired outputs in response to sample inputs; suited for modelling, controlling dynamic systems, classifying noisy data and predicting future events. This method subdivides in:
    - Classifications: learns to classify new observations from examples labelled data.
    - Regression: describe the relation between a response variable (output) and one or more predictor variables (input).
    - Pattern recognition: classifying input data into objects or classes based on key features, using either supervised or unsupervised classification.
- Unsupervised: trained by letting the neural network continually adjust itself to new inputs. Used to draw inferences from data sets consisting of input data without labelled responses, used to discover natural distributions, categories, and category relationships within data.
    - Clustering: used for exploratory data analysis to find hidden patterns or groupings in data.

ANN can solve these types of problems however they can take a long time or too much computing power, to solve this it is better to pre-process the data, focusing on the more relevant attributes in the input to achieve the desired output. The inputs should not contain too much information. Forms of reducing time and computing power to achieve convergence in the solution:

- Reduce the dimension of input vectors using component analysis.
- Perform regression analysis between the network response and the corresponding targets.
- Scale inputs and targets so they fall in the range [-1,1].
- Normalize the mean and standard deviation of the training data set.
- Use automated data pre-processing and data division when creating networks.

With improvements the network's ability to prevent overfitting, problem which occurs when a network has memorized the training set however has not learned to generalize to new inputs, improves. Knowing when a neural network is overfitting is not an easy task, the first step to avoid this should be dividing the train set, the validation set and the test set according to the size of the dataset being used, as demonstrated by Figure 2.12.

Small dataset

60%	20%	20%

Big dataset

98% 1% 1%
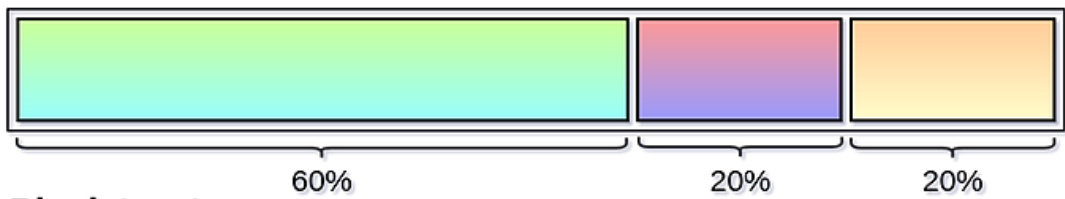
☐ Train set ☐ Dev set ☐ Test set
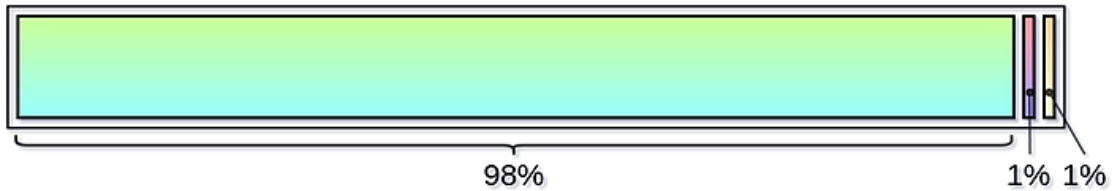
Figure 2.12 - Recommended method of dividing a data set [25]

Usually overfitting produces a small error on the training data set, the problem starts when new data is presented to the ANN, not knowing how to generalize it will produce a much larger error.

2.2.6 Artificial neural networks in fault detection/diagnosis

Papers have already started to introduce artificial neural networks to fault prediction methods in photovoltaic systems. In [26] the authors explore the utilization of this tool to build a detection network. Evaluation of the ANN's performance is done using the mean square error (MSE). Their system is composed by a photovoltaic panel connected to a DC/DC boost converter with P&O algorithm for MPPT control, presented in Figure 2.13.
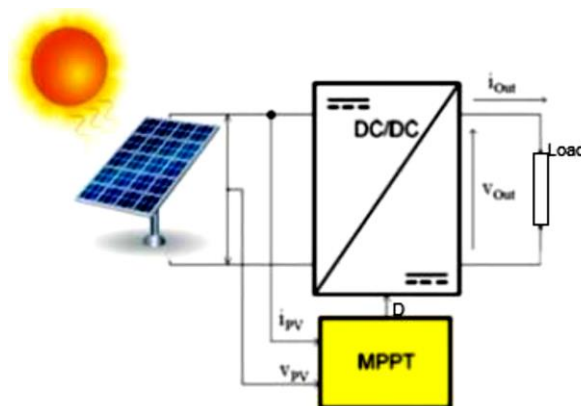
Figure 2.13 - System design [27]

In this study, five faults were tested, demonstrated in Table 2.4. It was subdivided in normal operation, one inverse module where the current passes through the bypass diode, two inversed modules, partial shadow in two or three modules and total shadow effect.

Table 2.4 - Classification of PV array faults [26]

Category	Fault type	Symbol	Code
1	Normal operation	H	[1;0;0]
2	Inversed module: one module	F1	[0;1;0]
3	Inversed modules: two modules	F2	[1;1;0]
4	Shading 1: two modules shadowed	F3	[1;0;1]
5	Shading 2: three modules shadowed	F4	[0;1;1]
6	Shading 3: intense shading effect	F5	[1;1;1]

To construct the artificial neural network, they used a Matlab based model and configured the inputs as voltage and power of the PV system, the targets were set as the codes shown in Table 2.4. Data set for this problem was split into two subsets, 70% was used to train the gradient and to readjust the bias and weights. The other 30% were samples to validate the model. MSE allowed for a more accurate test of the model as it utilized as an error the difference between the targets and the outputs obtained in the training process. A normal measure to apply to these types of networks.

Over in [16] another type of fault detection using ANN is proposed. First simulations were performed based on the system to identify normal operation and define a threshold limit for it. A study of the I-V curve, of the characteristics of the system was carried out. In Table 2.5 is shown a list of faults recognized by the authors as the more probable to occur and more prone to higher loss of power.

Table 2.5 - Types of faults occurring in a PV array [16]

Type of fault	Name	Symbol
Module	Short circuit fault in any bypass diode or (cell or module)	F1
	Inversed bypass diode fault or (cell or module)	F2
	Shunted bypass diode or (cell or module)	F3
	Open circuit fault in any cell or (module)	F4
Connection fault	Connection resistance between PV modules	F5
Partial shadow fault	Shadow effect in the modules with normal operation of different components of PV strings	F6
Shadow effect with faulty bypass diode	Shadow effect in a group of cells equipped by a faulted bypass diode open	F7
Shadow effect with connection fault	Shadow effect in a group of module connected by a connected resistance	F8

After simulations and already an established set of faults, starts the first part of the schematic designed by the authors. Values are compared with simulations to detect if the threshold limits, previously stated, are in check or not. Then it leads, based on the previous identification, to an attribute's selection. This was defined in the simulations, where from the study of the characteristics of the system a relation could be made with the electric parameters. The attributes can relate to the number of faults and their type of flaw. Moving further, knowing the attributes, the authors differentiated faults with two algorithms, number one isolating the faults when they have different combinations. Here is where the simulated and measured are calculated, and their relative difference is compared with threshold values. All of them are determined by the measurement noise and the model's uncertainty. Then the second algorithm distinguishes faults that have the same attributes. It uses an ANN to choose which fault is affecting the system [16].

Chapter 3

Model interpretation

This chapter focusses on the model utilized to create a database for faults and standard behaviour. Here it is described how it was developed and a detailed explanation why it was chosen for this purpose. Also, as it is a new model, an experimental procedure was done to conclude if it managed to follow characteristic I-V curves and could simulate mismatch faults.

3.1 Model explained

The model developed by Eduardo Sarquis [27], was chosen to create a database with standard behaviour and abnormal behaviours. This model uses the same methodology of the one diode and five parameter model to represent the equivalent solar cell circuit. However, it has a different view, instead of looking at the system as a string of photovoltaic modules, goes into more depth and gives a more detailed approach, by considering each cell. The basic structure in a PV panel, the solar cell, is taken as the most important, a panel is now seen as a set of cells with a predetermined configuration. For example, the ones used in the energy laboratory in IST are constituted by three substrings, each with twenty-four cells. Every cell will have its parameters, and the aggregate of all will constitute the equivalent circuit and therefore the panel. A more detailed explanation will be given now, where all the pros and cons of using the model will be elucidated.

It has the same fundamental equations as the one diode and five parameter model, the ones stated in chapter 2.1.4. The difference is that usually, that model is used to calculate those unknown parameters, and, in this case, those values are from CEC. CEC means the California energy commission. It is a known and respected institute, that test various types of photovoltaic panels. They went through rigorous and precise experimental tests, leading to the obtention of the five parameters for each of those panels. An optimized variation of the one diode and five parameter model is used by this institution, which will be presented below [28].

The presented equation (15) below is equal to the model shown in chapter 2.1.4, equation (3).

$$I = I_{ph} - I_s \left(e^{\frac{V + I \cdot R_s}{n \cdot V_t}} - 1 \right) - \frac{V + I \cdot R_s}{R_{sh}} \tag{15}$$

However, CEC uses the sixth parameter to adapt the temperature coefficient of the short circuit (μ_{Isc}) and open-circuit voltage (β_{Voc}). The reference values ($\alpha_{sc,ref}$) and ($\beta_{oc,ref}$) are obtained from the panel's datasheet, provided by the manufacturer. From this new parameter, called *adjust*, the new values are calculated using the next two equations (16,17):

$$\alpha_{Isc} = \alpha_{sc,ref} \left(1 - \frac{adjust}{100} \right) \tag{16}$$

$$\beta_{Voc} = \beta_{oc,ref} \left(1 + \frac{adjust}{100} \right) \tag{17}$$

Searching their website, at [29], gives access to a database of photovoltaic panels, where they were tested, allowing the gathering of their parameters ($R_{s,ref}$, $R_{sh,ref}$, n, I_s, I_{ph}, *adjust*), at STC [30]. The

reference values are given at standard test conditions, a normalized reference for all panels. Values these refer to the temperature at 25°C and irradiance at 1000 W/m². As known the panel has a usual configuration being regularly constituted by three substrings, each with a bypass diode, and with twenty or twenty-four cells in each substring. In this model, a description of the configuration is given in the form of a matrix. It has three fields [i, j, k], i represents the number of strings in parallel, all with the same number of panels, j represents, simultaneously, the number of substrings and the number of bypass diodes, and k symbolizes the number of cells in each substring. As an example, a 3x3 panel configuration where each panel has seventy-two cells would be represented in a matrix composition like this [3, 3, 24]. Meaning all the variables will have a matrix of their own to represent each cell in the system. To commence the model receives as input all the variables below:

- Input:
  - V_{ref} – voltage determined by the number of panels in the system.
  - G – irradiance values for each cell.
  - T – temperature values for each cell.
  - R_s – a coefficient matrix used to multiply the $R_{s,ref}$ values, with standard behaviour each value of the matrix will be equal to one.
  - R_{sh} – a coefficient matrix used to multiply the $R_{sh,ref}$ values, with standard behaviour, each value of the matrix will be equal to one.
  - $I_{ph,cell}$ – a coefficient matrix used to multiply the I_{ph}, with standard behaviour, each value of the matrix will be equal to one.
  - CellParam – a vector with all the CEC parameters, in order: I_{ph}, I_s, R_{sh}, R_s, n, $α_{sc}$, adjust.
  - Inicond – represents initial conditions, can be optional to start, storing results in the output RES from previous iterations.

A brief illustration of all constants used over the calculations:

Constants:

- Boltzmann constant → $KBev$ = 8.617332478e^{-5} [eV/K] or KB = 1.381*10^{-23} [J/K]
- Electron charge → q = 1.602*10^{-19} [C]
- Cell material bandgap energy → Eg_{ref} = 1.121 [eV]
- Cell material bandgap correction → dEg/dt = -0.00002677
- $α_{Isc}$ = $α_{sc}$*(1-adjust/100)
- Cell temperature = T+273.15

The current is the main output and this variable changes significantly with weather conditions, primarily with temperature and irradiance, in the following equations (18,19,20,21,22). According to [28], these variations are calculated to provide a more accurate calculation of the output current. Next equations (18,19) retrieve the value of I_s. In (18), a correction to the cell material bandgap energy based on the temperature is applied. In (19) that correction is used in conjunction with the temperature of the cell to obtain the saturation current.

$$\begin{cases} E_g = E_{g,ref}\left(1 + \dfrac{dE_g}{dt}\left(T_{cell_K} - 298\right)\right) & (18) \\[4mm] I_s = I_{s,ref}\left(\left(\dfrac{T_{cell_K}}{298}\right)^3 e^{\left(\frac{E_{g,ref}}{KBev\cdot298} - \frac{E_g}{KBev\cdot T_{cell_K}}\right)}\right) & (19) \end{cases}$$

Another correction based on temperature and irradiance values is made in the next two equations, now interfering with I_L, the cell's output current:

$$\begin{cases} I_{ph} = I_{ph,ref} \cdot I_{ph_{cell}} + a_{Isc}\left(T_{cell_K} - 298\right) & (20) \\[3mm] I_{ph} = I_{ph} \cdot \dfrac{G}{1000} & (21) \end{cases}$$

The series resistance is not influenced by any of these factors, at least significantly to be noted; on the other hand, the shunt resistance is by irradiance values. A correction is applied with the equation below (22):

$$R_{sh} = R_{sh} \cdot \frac{1000}{G} \qquad (22)$$

Configuring the bypass diode requires parameters, these are:

- $T_D = 25^oC \rightarrow$ Bypass diode temperature.
- $I_{sd} = I_{s,ref} \rightarrow$ Bypass diode current.

Having all the corrected values, according to the selected temperature and irradiance, lets the known method newton-raphson be employed to acquire the output current. It is an iterative approach where an initial guess is given to the output and iteratively calculates another output until the error between the two last value is less than the desired.

Finally, the model is complete, and it gives the outputs, these are:

- ✓ I_{out} – the output current.
- ✓ RES – a vector with all calculated variables.
- ✓ Flag – numerical convergence alerts.
- ✓ Y – residual error from the system of equations.

The model can express mismatch errors in the I-V curve. It proves to be extremely helpful to simulate specific behaviours, producing very reliable sources of knowledge to determine abnormal operation.

Hence the utilization of said model; however, it also brings a high computational cost. It needs to calculate for each cell the output current instead of only for one panel. In the next subchapter, an experimental procedure will be explained to demonstrate the capabilities of the model and compare it with the results obtained from that experiment.

3.2 Experimental procedure – Testing the model

In the energy laboratory, an experimental procedure for analysis of the I-V characteristic curve of healthy and damaged solar cells in PV panels was carried out. There were two solar monocrystalline PV panels, both with the same manufacturer's datasheet. They were previously set in the rooftop of the north tower in IST, and one of them suffered damage from high-velocity wind caused by a storm. From this event that panel had broken glass and had one broken cell, these conclusions were drawn by visual observation. Figure 3.1 shows a side by side comparison of the two panels tested in the laboratory.

Figure 3.1 - The two panels tested in the laboratory: left side (broken glass and a broken cell), right side (healthy panel)

Behind the solar panels, openings were made to access the cells individually. For that, and to have more immediate, more effortless and more stable readings; small claws were welded to allow links for wires, shown in Figure 3.3. Before mounting the setup, the manufacturer's datasheet, seen in Table 3.1, was observed for better comprehension of the limits and boundaries of the equipment.

Manufacturer's Datasheet	Suntech
Model	STP190S-24/Ad+
Rated Maximum Power (P_{max})	190 W
Output Tolerance	0/+5%
Current at P_{max} (I_{mp})	5.20 A
Voltage at P_{max} (V_{mp})	36.6 V
Short-Circuit Current (I_{SC})	5.62 A
Open-Circuit Voltage (V_{OC})	45.2 V
Nominal Operating Cell Temp. (T_{NOCT})	45ºC ± 2ºC
Weight	15.5 Kg
Dimension	1580mm×808mm×35mm
Maximum System Voltage	1000 V
Maximum Series Fuse Rating	15 A
Cell Technology	Mono-Si
Application Class	A

All technical data at standard test condition

AM = 1.5 $E = 1000 \ W/m^2$ $T_C=25ºC$

After acknowledging the significant aspects of the datasheet, the setup for the experiment was constructed. A configuration of the experimental setup is present in Figure 3.2. It is composed by the solar cell on the panel, connected through wires, in parallel to the voltage sensor, and to the current sensor and the variable resistance, these last two connected in series.
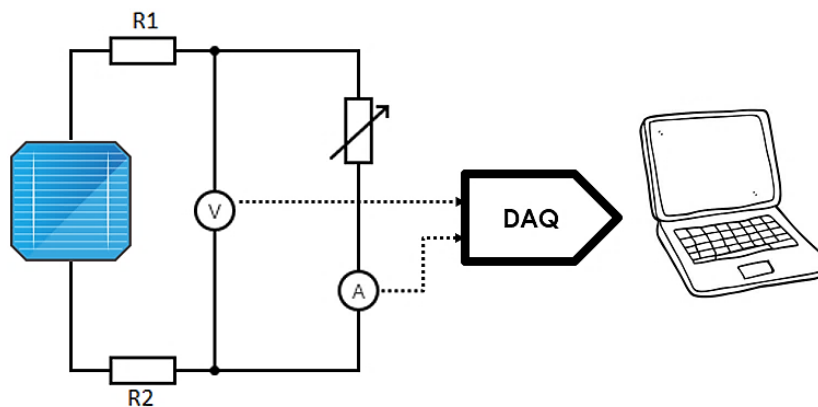
Figure 3.2 - Illustration of the setup in the energy laboratory

In Figure 3.3 and Figure 3.4, the actual setup and its configuration are demonstrated.
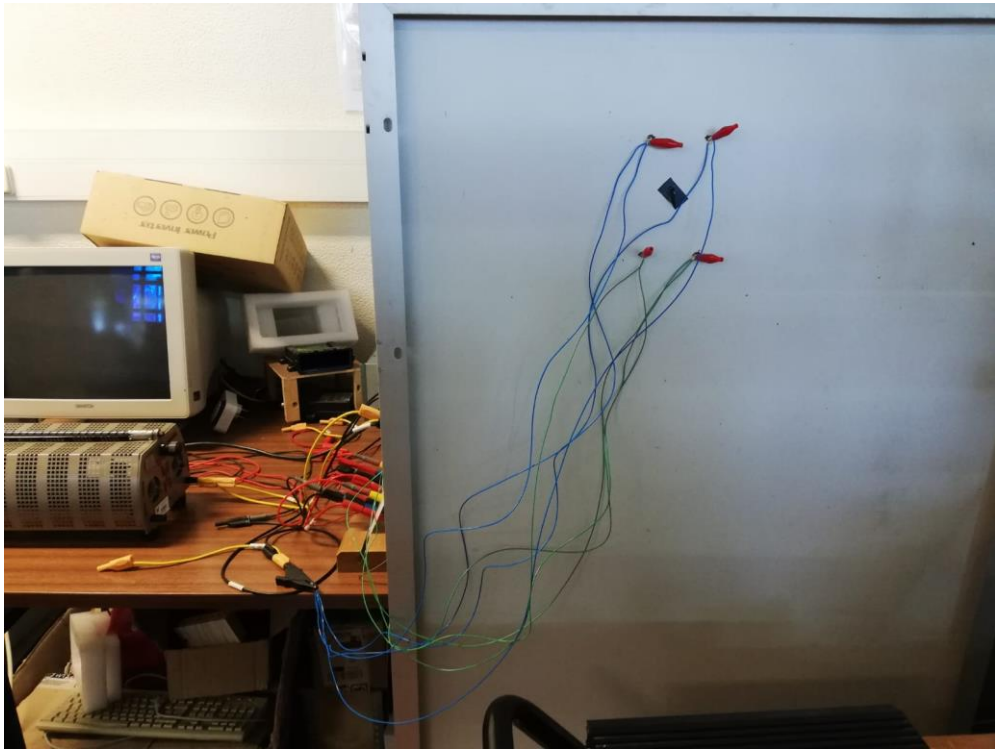
Figure 3.3 - Connection of the cell in a panel to the I-V curve tracer system
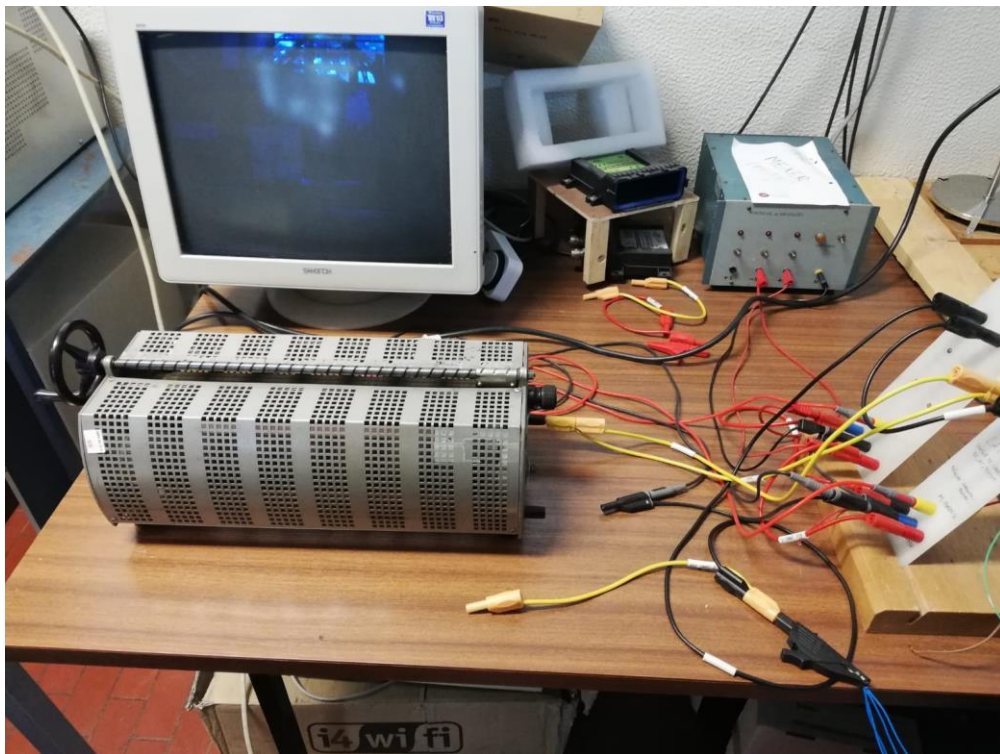
Figure 3.4 – Overall connections of the system

The sensors are then linked to a DAQ to obtain the measurements and transfer them to the PC. Reproducing the effect of the solar light is a 500 W lamp, fixed in an angle of incidence of approximately 90 degrees. It was a necessity to maintain the irradiance at maximum value and with the least refracted and reflected light in the area lit. Rotation of the lamp and the movement back and forth, to simulate different levels of irradiance, initially caused errors in measurements, so a custom design track was built to reduce those errors to a minimum. As it can be seen in Figure 3.5, it only allows the movement of the lamp in a single direction. Also simplified the measurements of irradiance, as the measurements were marked in the track, representing the various irradiance levels related to the distance to the solar cells.

Figure 3.5 – Custom design track

To accomplish this objective, first, there had to be a consideration for the maximum current and maximum voltage that the solar cells could reach. All this to be possible to draw the I-V curve in all its points. However, when the resistive load is applied, it was not possible to obtain the short circuit current (I_{sc}), but a very close value. It was due to the resistance applied by the wires. These will always apply resistance which in this case cannot be neglected. To minimize this effect two cables were connected in parallel in each solar cell terminal, instead of only one. A reduction of almost 50% was accomplished by using this method. Measurements for the wires' resistance are presented in appendix A. Another effect to be considered, the input/output variation applied by both the voltage and the current sensor (hall sensors). Linear regression was done, to fix the deviation caused by this error, seen in Figure 3.6 and corrected in the final curve, one for the voltage sensor and another for the current sensor.
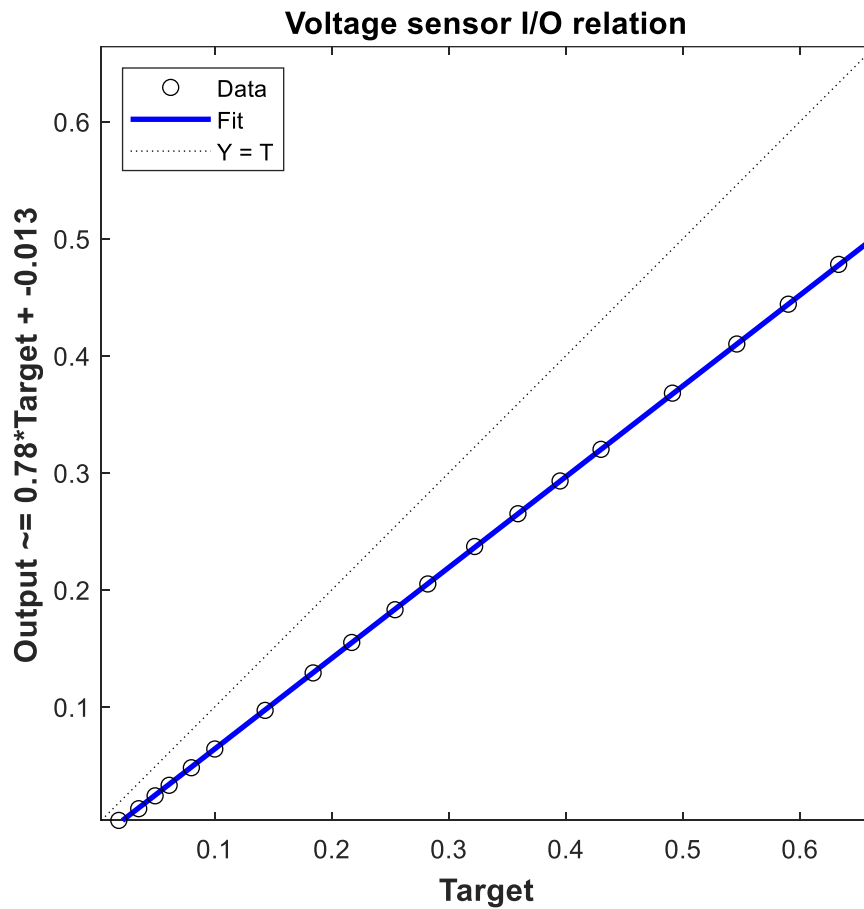
Figure 3.6 – Hall sensor correction

To the DAQ results, the *V* and *I*, respectively, voltage and current, the obtained sensor's input/output corrections below are applied:

Voltage sensor → $0.78 \cdot V - 0.013$

Current sensor → $1.88 \cdot I - 0.027$

In Figure 3.7 and Figure 3.8 can be seen the juxtaposition between an I-V curve with the hall sensor corrections, in voltage and current measurements, in blue, and an I-V curve without corrections, in orange. With higher irradiance is possible to see the limitations of this setup as it does not trace the curve to the point of the short circuit. However, the objective is still maintained, as the curve is perceptible, in Figure 3.7. In the lower irradiance curve, Figure 3.8, the curve is much more defined.
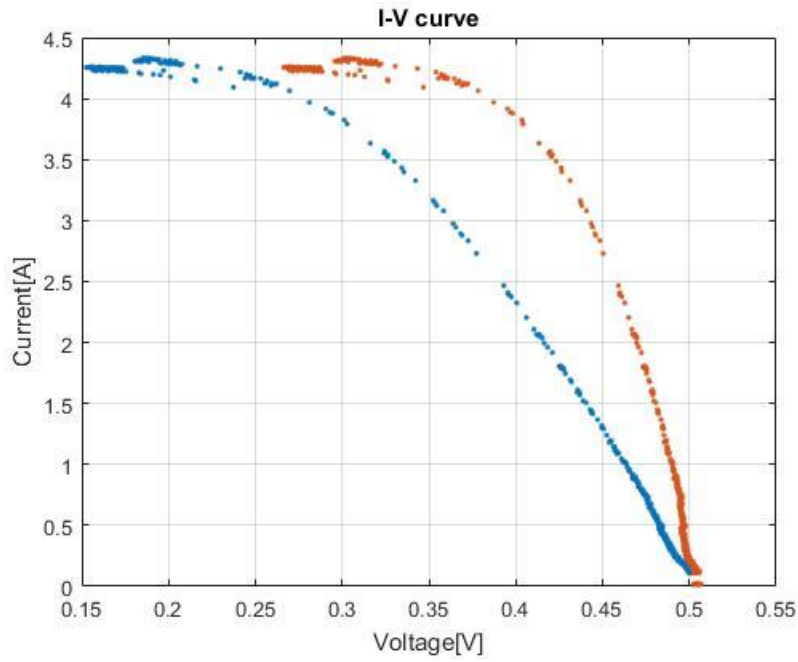
Figure 3.7 - Experimental result of an I-V curve with an irradiance of 1000 W/m²
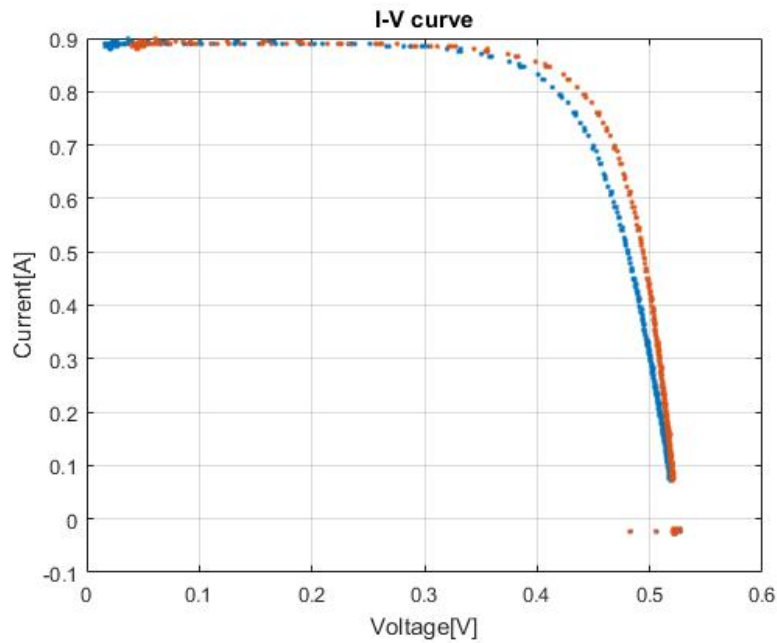
Figure 3.8 - Experimental result of an I-V curve with an irradiance of 230 W/m²

Continuing the corrections' path, the resistance of the wires was calculated for each irradiance level. Even with the corrections, it is not possible to reach the short circuit current in higher irradiance levels, and it becomes worse due to the still high levels of resistance. The relation between the voltage, current and resistance is known as ohm law (23).

$$V = R \cdot I \tag{23}$$

In the lower irradiance levels, the current presents lower values and the resistance in the wires is interferes less in the I-V curve. However, when increasing the current, the resistance will affect more and prevent the tracing of lower voltage points, figures Figure 3.13, Figure 3.14 and Figure 3.15 illustrate this side effect ideally. For the maximum efficiency and to minimize errors, an experimental method was outlined and followed precisely in every measurement made, all the steps can be seen in Figure 3.9.

```
┌─────────────────┐
│   Measuring     │
│    Routine      │
└────────┬────────┘
         │
         ▼
┌─────────────────────────────────────────────────────┐
│ Position the lamp, according to the requested        │
│ irradiance:                                          │
│ a. G=1000, 800, 600, 400, 230 [W/m^2]                │
└────────────────────────┬────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────────┐
│ Turn on the lamp and let the temperature stabilize,  │
│ verifying the multimeter is not on HOLD:             │
│ a. Time needed for temperature to stabilize ~10 min  │
│ b. Register temperature                              │
└────────────────────────┬────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────────┐
│ Obtain the data to calculate corrections in the I-V  │
│ curve:                                               │
│ a. Apply short circuit in the load                   │
│ b. Measure V_cc, voltage in the terminals of the     │
│ voltage sensor                                       │
│ c. Measure V_R1, voltage between the voltage sensor  │
│ and the positive terminal of the cell                │
│ d. Measure V_R2, voltage between the voltage sensot  │
│ and the negative terminal of the cell                │
│ e. Measure I_cc, current flowing in the cell's       │
│ wires, for both terminals                            │
└────────────────────────┬────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────────┐
│                                                      │
│ Prepare DAQ for data acquisition                     │
│                                                      │
└────────────────────────┬────────────────────────────┘
                         │
                         ▼
┌─────────────────────────────────────────────────────┐
│ Initiate process to obtain I-V curve:                │
│ a. Begin with open circuit                           │
│ b. Connect the circuit and move the variable         │
│ resistance from maximum value to minimum value       │
│ c. End with short circuit                            │
│ d. Repeat this process at least 10 times             │
└─────────────────────────────────────────────────────┘
```

Figure 3.9 – Experimental procedure

Proceeding with the experiment, other types of corrections for measured data had to be delimited. These are all stated in Figure 3.10.
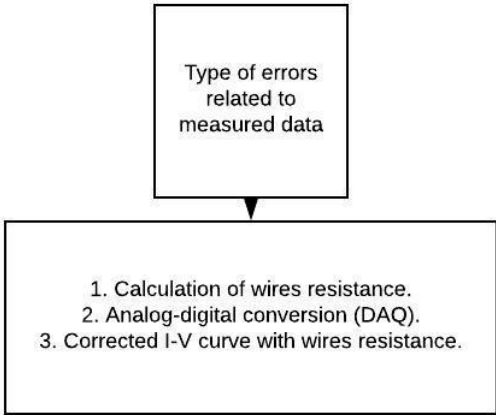
Figure 3.10 – Corrections for measured data

Moreover, the own equipment's error window had to be accounted for. This is shown in Figure 3.11.
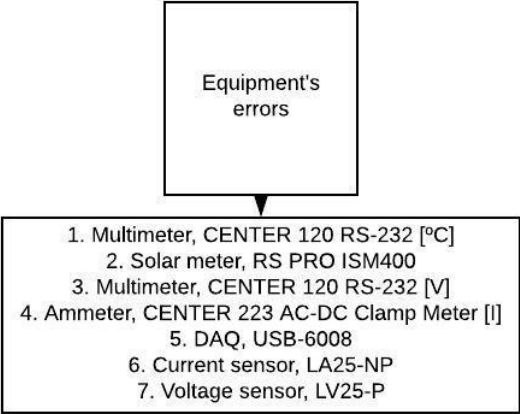
Figure 3.11 – Errors due to equipment

Based on the previous enumerations of errors a diagram was constructed for easier visual comprehension, it can be observed in Figure 3.12. Labels, in the figure below, were designated according to the error they represent. Label 1 is an error caused by measuring errors that can be eliminated by doing at least ten measurements and trace an average value of them. Label 2 is the equipment's error, stated in their datasheet, which is available in the annex. Label 3 is the equations' error, which includes the wire resistance and the sensors corrections.
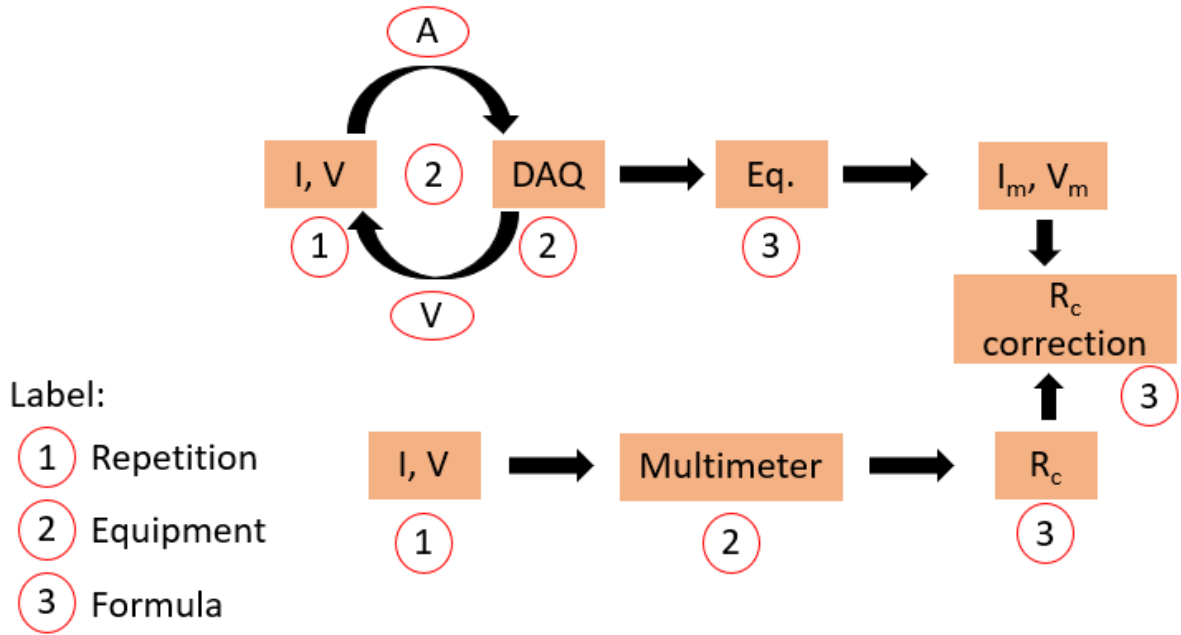
Figure 3.12 - Diagram representing all error steps

All the procedure of measuring contained errors in the equipment and with the several steps, a propagation of the error had to be calculated. According to [31], the ways to calculate error propagation depend on their combination, when an addition is applied using a corrected value equation (25). Equation (24) serves as an example of corrected variables being summed or subtracted.

$$Q = a + b + \cdots + c - (x + y + \cdots + z) \tag{24}$$

$$\delta Q = \sqrt{(\delta a)^2 + (\delta b)^2 + \cdots + (\delta c)^2 + (\delta x)^2 + \cdots + (\delta z)^2} \tag{25}$$

In another case, when the corrected value is used in a multiplication or a division, a different error propagation equation (27) should be applied to calculate the proper error. In equation (26) an example of corrected variables being multiplied and divided is given to demonstrate the process used in the calculation.

$$Q = \frac{a \cdot b \cdots c}{x \cdot y \cdots z} \tag{26}$$

$$\frac{\delta Q}{|Q|} = \sqrt{\left(\frac{\delta a}{a}\right)^2 + \left(\frac{\delta b}{b}\right)^2 + \cdots + \left(\frac{\delta c}{c}\right)^2 + \left(\frac{\delta x}{x}\right)^2 + \cdots + \left(\frac{\delta z}{z}\right)^2} \tag{27}$$

The work started with the acquisition of the I-V curve from the DAQ, where it was performed a total of ten trials for each level of irradiance for two different cells, one in the heavily damaged panel and one in the healthy one. It was mainly to test the model with the actual cell behaviour. However, another

comparison between a healthy cell, a broken glass cell and a broken cell was made to understand the behaviour and comprehend what happened to the characteristic curve in these situations. Below in Figure 3.13, Figure 3.14, Figure 3.15, respectively broken cell, broken glass and healthy panel cell, are represented the I-V curves obtained for each radiation (230, 400, 600, 800, 1000 W/m²), where each colour is bond to a radiation value. It is shown all the ten tests in each radiation, creating a cloud of results for all I-V curves. The curves follow the usual characteristic I-V shape. However, it can be seen that the current is the highest in the healthy panel. Then comes the broken glass cell and finally with almost a 50% reduction in current the broken cell. This effect is more pronounced in higher irradiance levels.
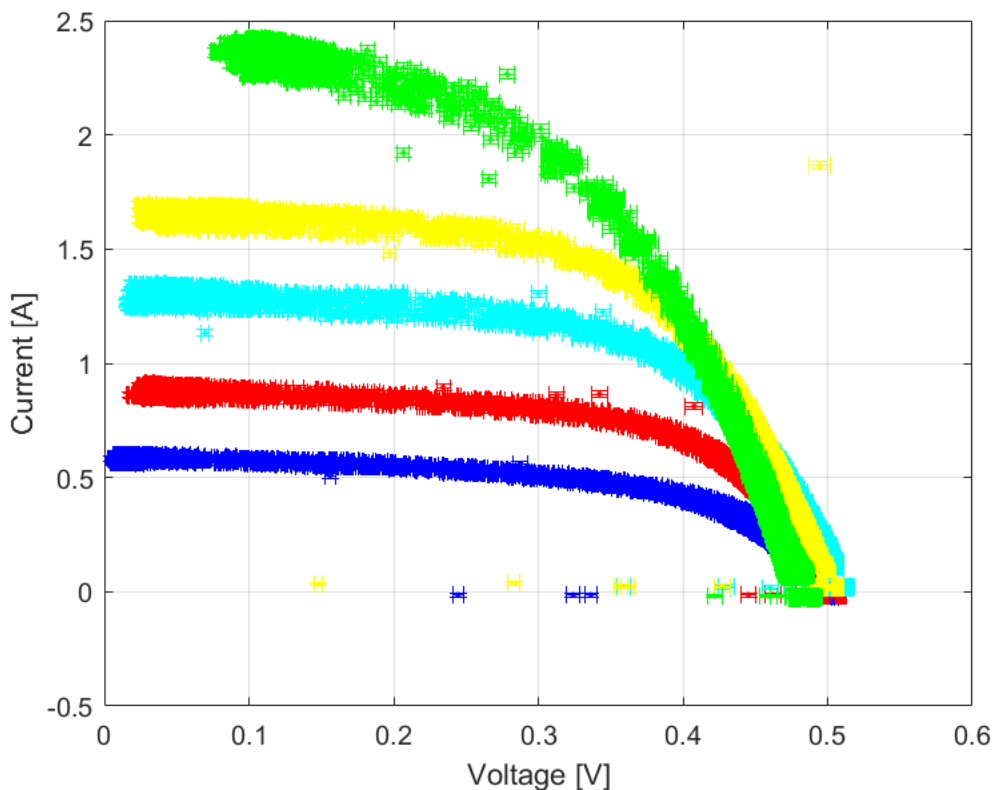
Figure 3.13 - Broken cell I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m²]

Initially, it was supposed to implement broken cells and broken glass as faults, however, these were not. It was tested and confirmed that broken glass changes the variation of received irradiance, due to the breakage of the glass, yet it remains difficult to characterize this deterioration given the inconsistency of the glass fractures. In the broken cell case is even more complicated. It requires the place where it got cracked, with high precision as it can induce several errors depending on the location. It is also problematic to know if the cell is damaged and not only the glass protecting it. For these reasons, those faults were not pushed in the variants to analyse.
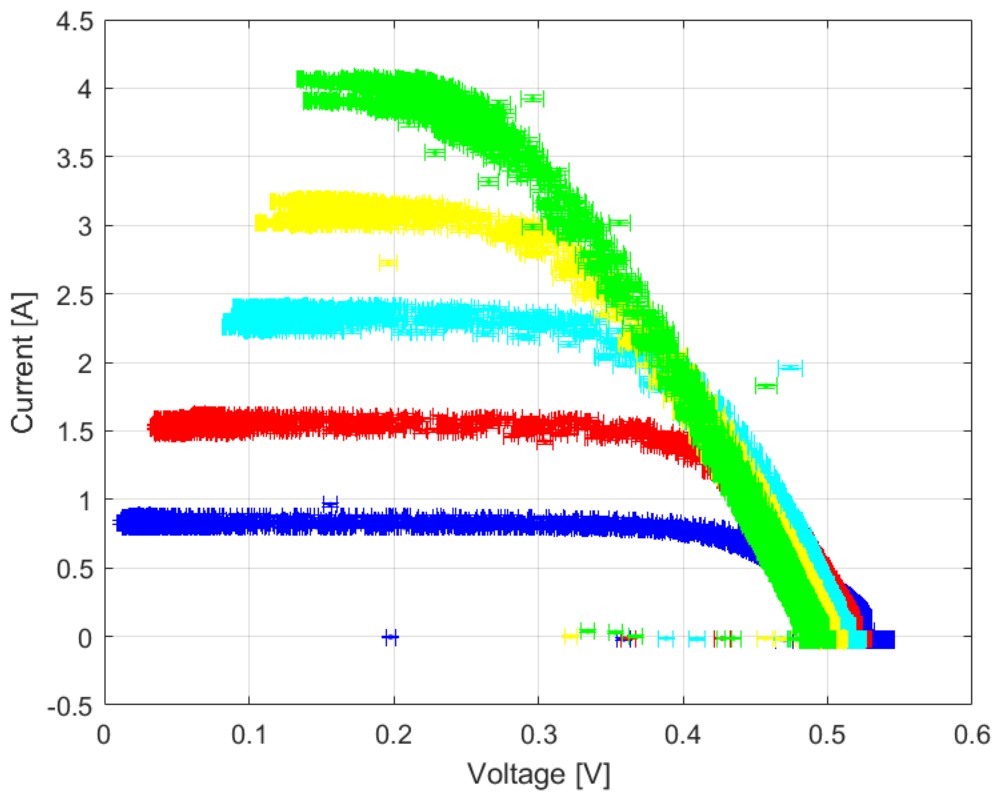
Figure 3.14 – Broken glass I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m²]

Going further, the plots referring to the experimental procedure will be using the healthy panel.

Now to test the model with the experimental results I-V curves with the same characteristics, temperature and irradiance, a simulation will be produced by the model and compared in Figure 3.16.

In a first juxtaposition, there is a visible difference in the current and voltage of the model compared with the experimental, being the latter smaller. It was detected a variation in the temperature read in the multimeter and the actual cell temperature, this was due to the location of the measurement. As it cannot be read in the front of the panel because it would block the solar irradiation. So, it is read behind the panel, and a constant temperature is applied to that value, previously calculated as an increase of 25°C. As for the irradiance, a ratio is multiplied to the value measured by the solar meter, more specifically, 0.72. This lower irradiation caught in the experimental result is caused by a lower spectral power range of the 500 W lamp. The model is prepared for the sun's irradiation which has higher spectral power. Hence the use of a ratio to fit the current, this will be justified, experimentally, in Figure 3.20 where the whole panel was tested outdoor. Applying both amendments to get a good approximation, obtained, as shown in Figure 3.17.
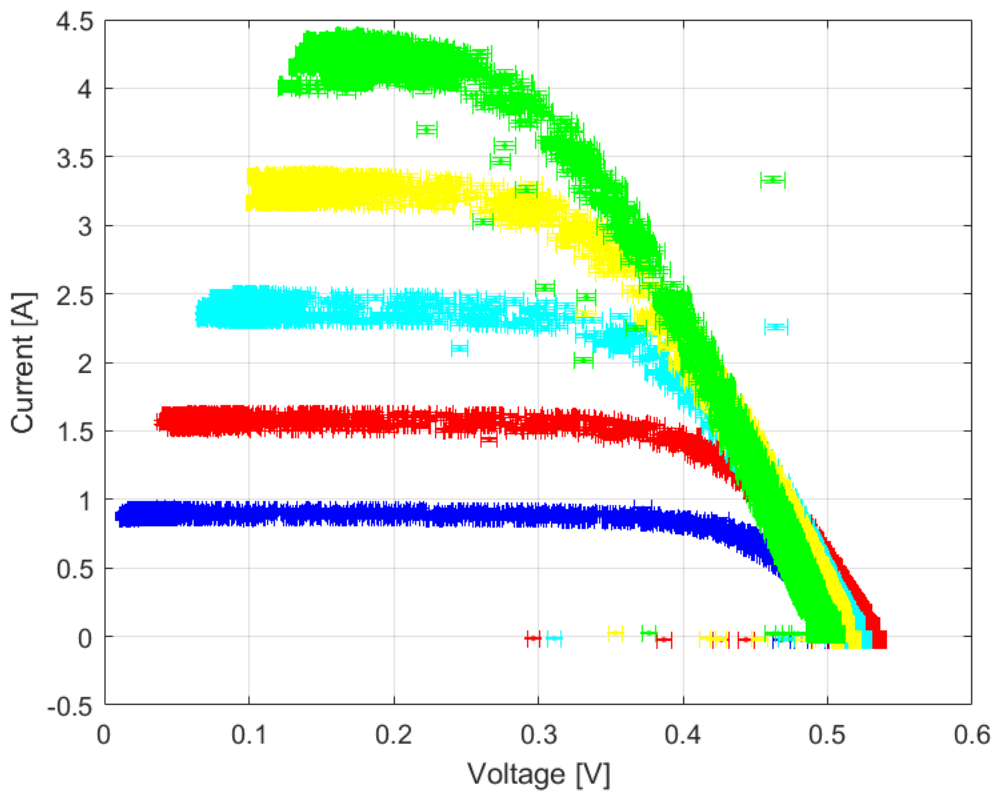
Figure 3.15 – Healthy panel I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m²]
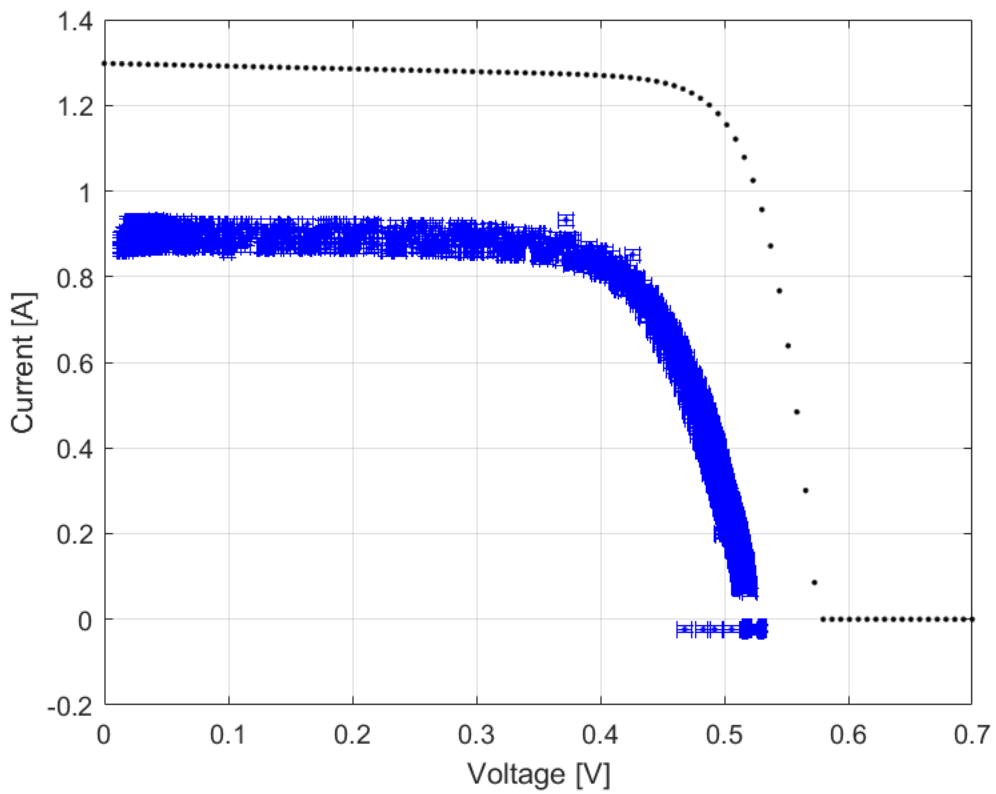
Figure 3.16 - Model and experimental results, in blue experimental and black model
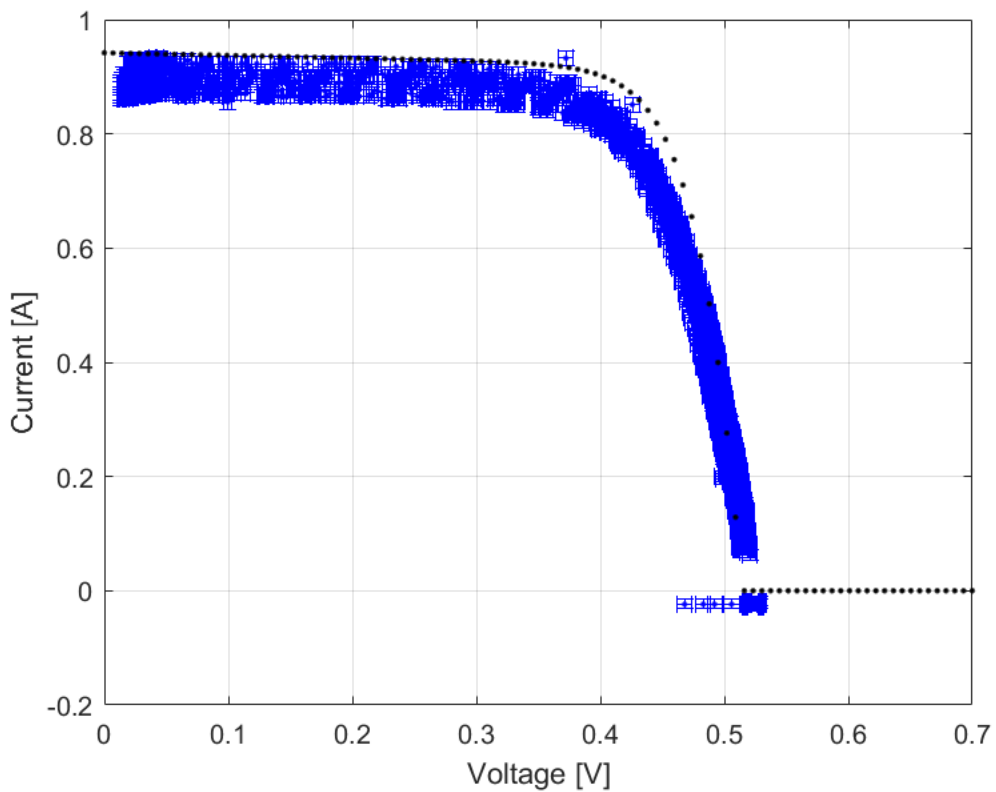
Figure 3.17 - Model, with ratio (0.72) multiplied by the irradiance, in black and experimental results in blue

The basic unit defined in the model was tested and produced similar results to the experimental, proceeding with an experiment on the whole panel. The same experiment as the one done for the cell was applied to the whole panel. However, it was done outdoors. In Figure 3.18, the outdoor experimental setup is shown. Now the cells were obtaining light irradiance from the sun, and the levels of irradiance were not possible to control, contrary to the lamp.

Figure 3.18 – Outdoor experimental procedure

For this test, a broad range of faults was tested, and all are present in Table 3.2. Standard behaviour was tested, and then the significant faults were drawn to be tested. Those were short circuit, shading and open circuit. However, the open circuit could damage the panel, and it was dropped. Subtypes of faults had to be delimited. For short circuits, it was necessary to test only one short-circuited cell, to comprehend the behaviour and what would imply in other short circuits. Twenty-Four cells and forty-eight cells short circuits were tried out to simulate a substring or two substrings short circuits, respectively. For shading, a more extensive range of tests had to be done. First, one cell was shaded with several percentages of shade in each test. That was accomplished by blocking, with an opaque object, a percentage of the cell equal to the percentage of shading. Concluding that test then two cells were shaded, one in each of two substrings, and the same levels of percentage were applied. Finally, the same was done for six cells, two in each substring. For every trial, there were done five curve tracings, to reduce measuring errors. Plot results of the procedure can be seen below in Figure 3.19.

Table 3.2 - Behaviours tested outdoor

Behaviour:	Standard	Short Circuit	Shading
Subtypes of behaviour	Standard operation	One cell	**One cell-shaded (in one substring)**
			25%, 50%, 75%, 100%
		Twenty-four cells	**Two cells shaded (one in each substring)**
			25%, 50%, 75%, 100%
		Forty-eight cells	**Six cells shaded (two in each substring)**
			25%, 50%, 75%, 100%

The I-V curves may change accordingly to the temperature and irradiances in the moment of the extraction of results, for every graph a table with the respective irradiances and temperatures will be below it to verify the variations of the curve with the actual variables.
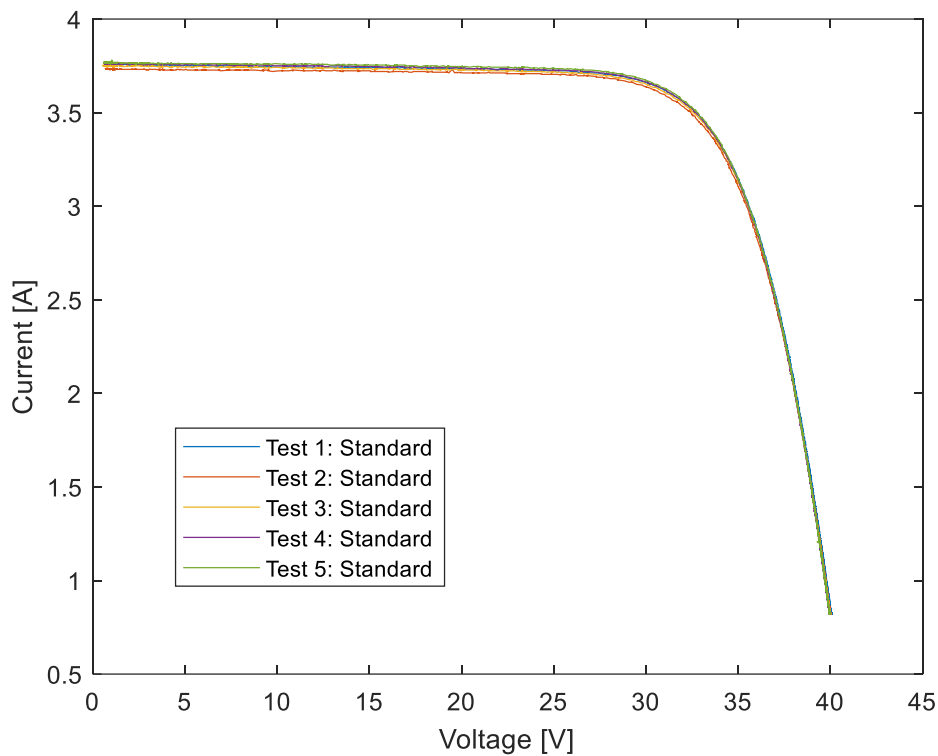
Figure 3.19 - Healthy panel in standard operation

Table 3.3 - Irradiances and temperatures related to figure 3.19

Tests	Test 1	Test 2	Test 3	Test 4	Test 5
G [W/m^2]	641	636	638,5	641,5	643
T [$^\circ$C]	42	42	43	42,5	42

The values, in Table 3.3 from column Test 1, were inserted into the model and a graph was plotted to envision a model comparison with the actual outdoor experiment, the first test can be seen in Figure 3.20. A relative distance is immediately noticed; however, it should be remembered that the measurements of irradiances vary a lot in a brief period. The solar meter has a large gap of failure, and the irradiance measurement was done with visual observation of the solar meter. It can also lead to a small error, allied to the back measurement of the cell's temperature can induce a variation, as indicated in the laboratory essays. However, it was seen that the same constant could not be applied. A trial and error method was tested in the model, and for the mentioned plot, the best approximation was defined to be plus 9°C to the measured temperature. In Figure 3.21, a plot of the referenced work is shown. All statements conclude that measurement errors can justify these small distances in the graph.
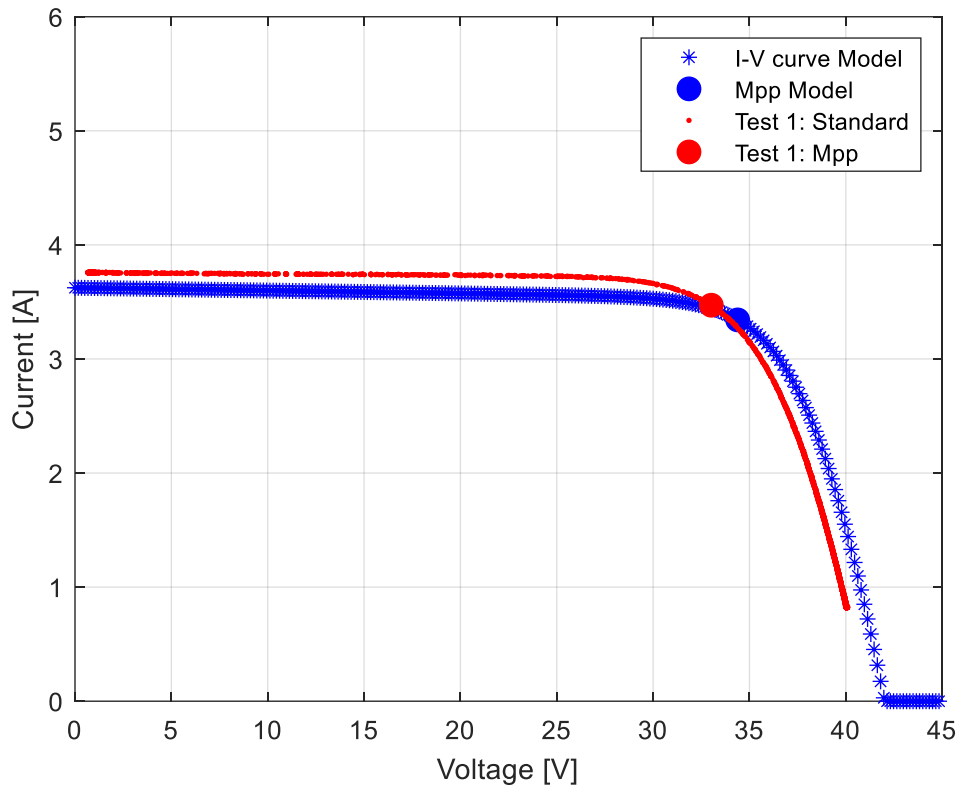
Figure 3.20 - Model vs Outdoor experiment comparison, without temperature adjustments
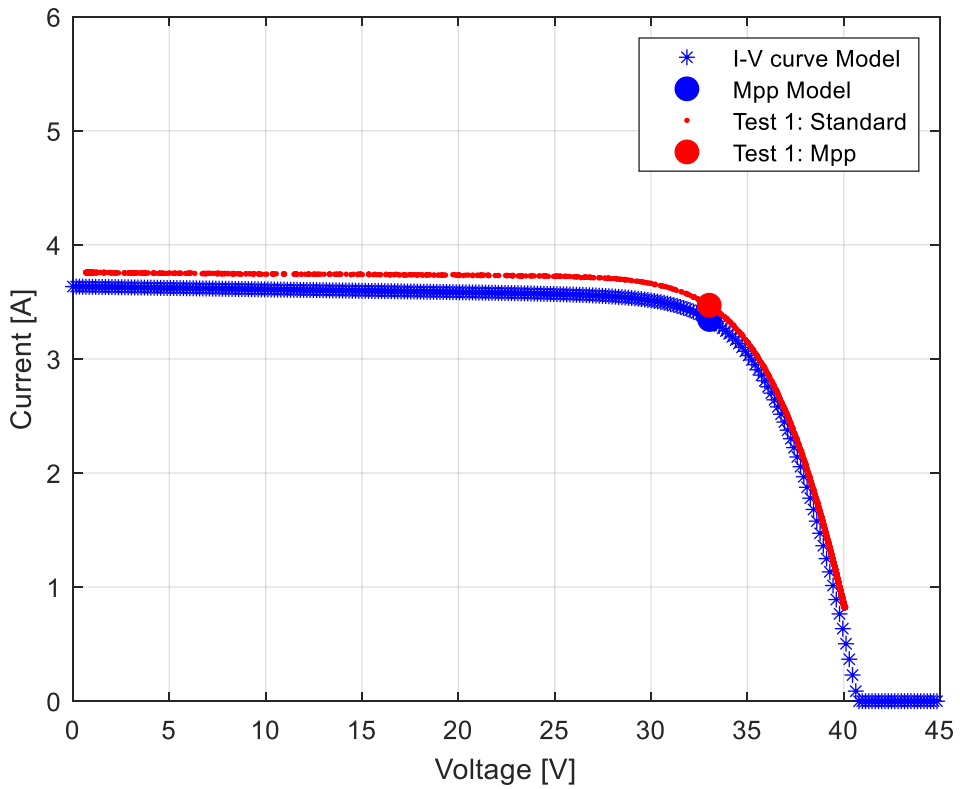
Figure 3.21 - Model vs Outdoor experiment comparison, with temperature adjustments

Chapter 4

Database construction

In this chapter, a meticulous explanation of the database creation process is done. Faults used in the database are specified, and the reasons behind their choices are justified.

4.1 Overview

To create a database, an understanding of the faults produced by the model was necessary. For this, it was decided to plot specific abnormal behaviours and a standard one to compare with each other and comprehend them hence choosing the best way to form a database. Table 4.1 enunciates all the chosen modes of operation. A typical mode, a short circuit of one cell, a short circuit of a substring, a short circuit of two substrings and all three short-circuited substrings. These form a general perspective of a panel and carry general options that can happen. For shading, broader subtypes were simulated. It was due to having more factors that could influence this type of power loss. Such as the number of cells being shadowed, their location and the percentage of shading being applied.

4.2 Fault creation

The model proposed in chapter 3 can effectively obtain results for mismatches in cells in a panel. The first approach for a database was standard behaviour, this term should not be mistaken with standard test conditions, the temperature and irradiance change in each test, however the system has no faults. To obtain results, a user introduces the topology of a system, the irradiance and temperature. Then the result will be an I-V curve with points equal to the number of substrings in the system times one hundred ($j \cdot 100$), equally spaced. It was essential to maintain the same step between the points in the curve and to keep the step small because in the process of choosing the maximum power point a higher step could cause deviations, producing a less viable database. For the modes enumerated above Table 4.1 is presented for a more transparent comprehension of them. These subtypes were chosen for different reasons. Starting with the short circuit denomination: a one-cell short circuit was especially to characterize the primary unit and the minimum decrease that could happen. The substring and the two substring short circuits were simulated to represent one third and two-thirds of this designated fault pattern. In shading operation, a little shading, minus 50 W/m², and a more substantial gap shading, minus 500 W/m², these numbers refer to the minimum error of the solar meter and half of the standard test condition irradiance, respectively.

Table 4.1 - Types and subtypes of operation simulated using the model

Types of behaviour	Standard	Short Circuit	Shading → - 50 W/m²	Shading → - 500 W/m²
Subtypes of behaviour	Standard	1 Cell → (SC1)	1 Cell → (SH1-G50)	1 Cell → (SH1-G500)
		24 Cells → (SC24)	2 Cells, one in each substring → (SH11-G50)	2 Cells, one in each substring → (SH11-G500)
		48 Cells → (SC48)	3 Cells, one in each substring → (SH111-G50)	3 Cells, one in each substring → (SH111-G500)
		72 Cells → (SC72)		

Obtaining a standard operation mode was simple. It consisted of running the model, and the only variables to choose were the topology, this is the overall system configuration, the temperature of the cells, and the irradiance being absorbed by the panels. For other modes of operation, it would imply changes to the matrices created by the model. To create short circuit faults the cells need to be in short circuit, to imitate this effect in the simulation model the R_s and the R_{sh} need to be zero. In this simulation, these values cannot be zero, so a low value is given to them to provoke the short circuit. In Figure 4.1, an equivalent circuit of a short-circuited cell can be observed. It provides an example of merely one cell. In the whole system, depending on the number of short circuits, there will be an equal number of cells with the same parameters as the one shown in Figure 4.1. In this type of fault, it does not matter the position of the cell and the consequence in the output will be a lower voltage value. If a lower voltage is detected, it is characterized as a short circuit of several cells, which depends on the decrease in voltage.
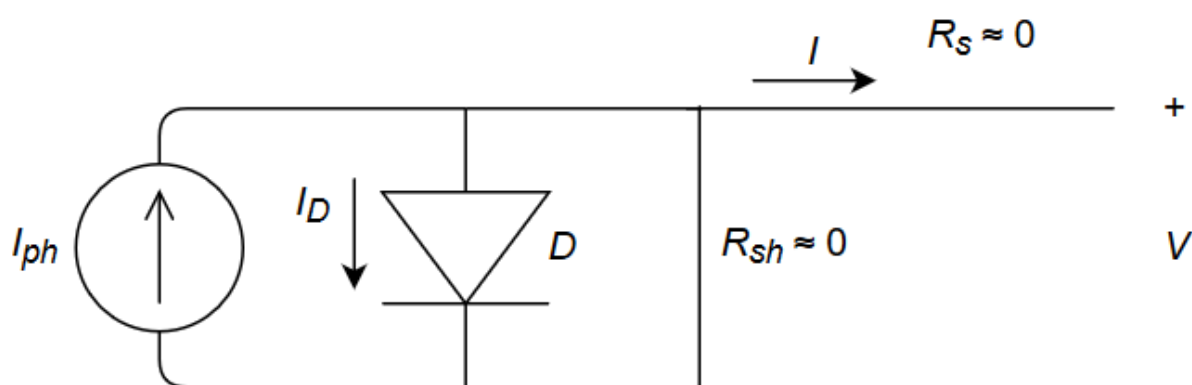
Figure 4.1 - Equivalent circuit of a cell under the effect of a short circuit

For a shading operation, when part of the panel is receiving less irradiance, the matrix of the irradiance must be changed in the position of the cells which are under that effect. In this case, the position of the cells being affected does matter because when a cell receives lower irradiance, it will have less current passing through it, changing the current, in that substring, to the value of the lowest cell's current. Causing mismatch and in the long run, will damage the cell and in consequence, the panel. Having that in mind, when simulating shade in cells, only a cell in a substring should be tested. Another factor to notice is the percentage of shade applied to the cell. Lower percentages indicate lower irradiance drops. In contrary, higher percentages of shade present high irradiance drops. However, when the difference between currents in a substring is too high, the bypass diode starts conducting, and the current passes through it instead of the cells in parallel. This is a mechanism to protect cells from major mismatch faults. However, when the bypass diode is damaged, the substring will suffer damage faster and become unused. So, giving input to the script should have in attention the position and the percentage of the shade, in Figure 4.2 is a representation of a shaded cell.
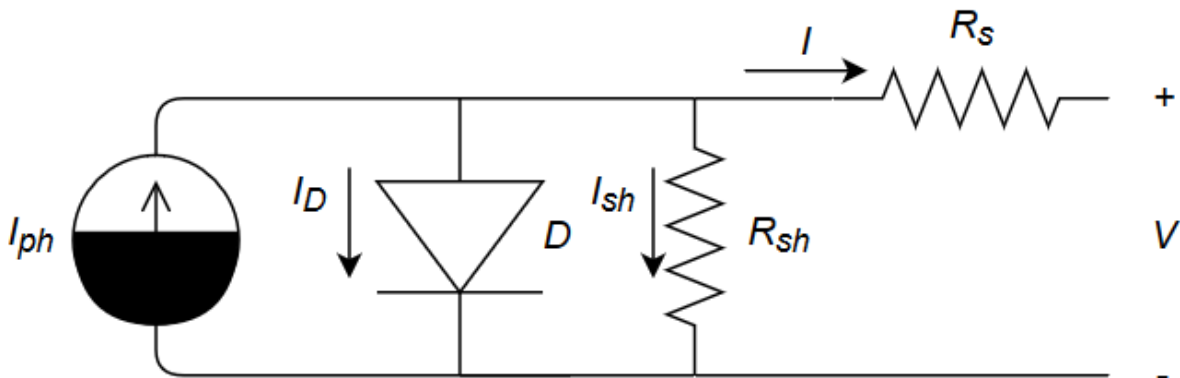
Figure 4.2 - Equivalent circuit of a cell under the effect of a shade

A scheme presented in Figure 4.3 shows a simplified description of the necessary variables to change and the inputs to insert to create each behaviour. In conjunction with this, there are also figures to complement with a visual aid of how the panel would look like in a cell level under these modes of operation.
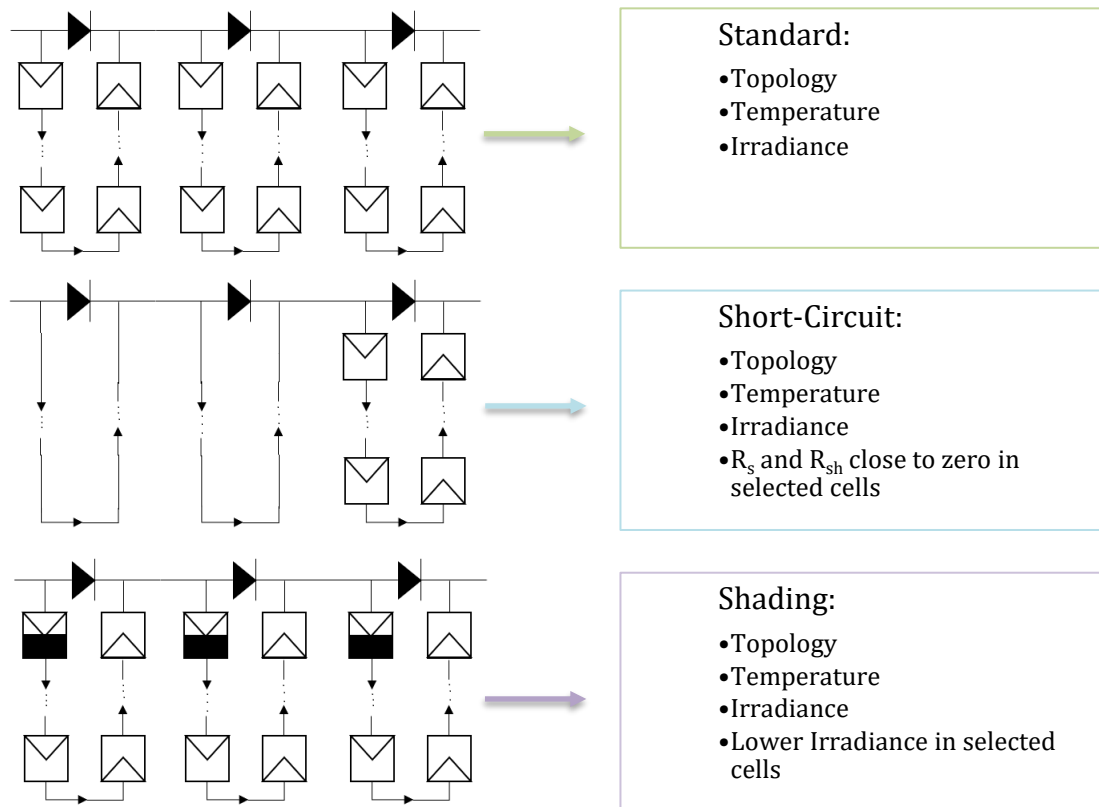
Figure 4.3 - Methods applied to simulate chosen faults and visual representation of said faults

4.3 Plot observation and comparison

From Table 4.1, a first simulation, based on the described modes, was made to be able to plot a combination of different behaviours. Making visible the significant similarities and differences between abnormal and standard behaviours. Every type and subtype were compared with each other as a way of showing to the reader what can and what cannot be distinguished. In Figure 4.4, a plot, of standard and short circuit can be observed, with a maximum power point for each curve. For this plot there is a clear distinction between the short circuit and standard behaviour, except in the low short circuit case. The loss of one cell's voltage will not cause enough drop of voltage to guarantee a clear separation of these two behaviours. There is an unmistakable difference between all the short circuits. All other plots are available in the appendix B. In these graphs a zoom is applied for difficult points to observe and in their legends is all the essential information.
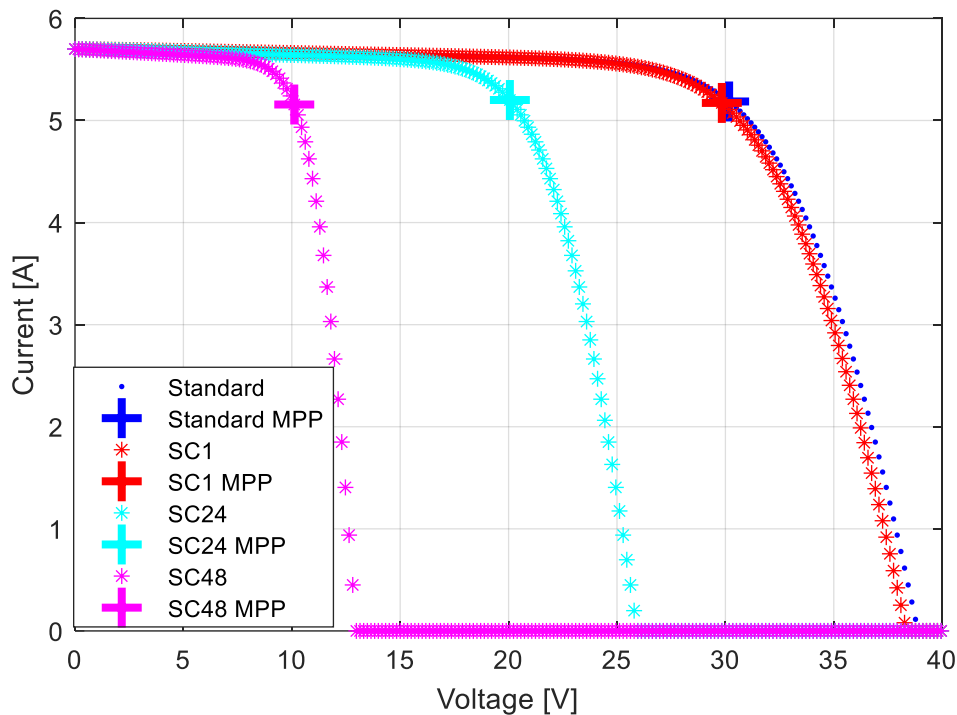
Figure 4.4 - Standard and Short circuit, both with an irradiance of 1000 W/m² and temperature of 70ºC

4.4 Database structure

Proceeding with the database structure, parameters were set to define boundaries, the temperature was set from zero to ninety degrees Celsius {0,90} and irradiance from two hundred to a thousand watts per square meter {200,1000}. These values were chosen based on the average ambient temperature plus the irradiance in contact with the panel, which increases their temperature. For irradiance, the values were based on the lowest meaningful to the standard test irradiance value. The temperature, in a cycle, would change from zero to ninety in a two Celsius step, this was adopted because the datasheet of the panel had the minimum error defined equal to two. Again, the irradiance step was set to fifty watts per square meter as it was the solar meter established error. This process of construction of the database proved to be inefficient when introduced in the artificial neural network. It overfitted since the data collected was simulated statically, only changing the topology or the CEC parameters of the panel, affecting the results. So, the ANN could imitate the function simulating these results, causing a deficient error percentage. However, it will not be a real demonstration of the pattern capabilities of this tool and provides inaccurate output. Resolving this upset a whole new logic was applied to the construction of the database, replacing a static structure by a random. In place of doing cycles with temperature and irradiance, these are chosen from a random discrete uniform distribution with a minimum and maximum value defined. To increase the randomness of the whole process, the mode of operation is also randomly chosen with the same configuration, only changing the minimum and maximum. From {0,3} where zero to one defines the standard operation, one to two defines short circuit operation and from two to three

defines shading operation. Inside the short circuit mode, also a random distribution is applied to the number of short circuits. In shading, two random operations must be applied, the number of cells being affected and the percentage of shadow affecting them.

Having a defined structure with all variables used stored was very important. It allowed to define and identify specific subtypes of faults; in this matter, it is presented a vector of the final structure containing all the parameters inputted and the outcome variables. As mentioned before the first values are the voltage and then the current from the I-V curve tracing, necessary to observe if anything appears to be out of order. Subsequently come the temperature, the irradiance, the voltage of maximum power, the current of the maximum power and the maximum power. Later the subtypes specific variables appear, the number of short circuits (Nsc), the number of shaded cells (Nsh) and the percentage of shade (Psh). The final three columns represent the output code, defining the behaviour. Over in Table 4.2, an example of a row of the matrix, for one panel, is shown, it is represented as a column since it provides a better portrayal.

Table 4.2 - Example of a row in a database, with the number of components below their respective names

V (300)
I (300)
T (1)
G (1)
V_{mpp} (1)
I_{mpp} (1)
P_{mpp} (1)
Nsc (1)
Nsh (1)
Psh (1)
OC (3)

After concluding the random database had better results in artificial neural network training as it had less issues with overfit, however another issue arose from the same theme. The number of entries in the system could cause overfitting. So, to understand how many entries the database should have the maximum number of possibilities were thought out, considering only one panel. In Table 4.3, it is presented all the possible cases.

Table 4.3 - Total combinations possible using the random database

Standard	Short-Circuit	Shading
Variables: T, G	Variables: T, G, Number of Short circuits (Nsc)	Variables: T, G, Number of shaded cells (Nsh), percentage of shading (Psh)
T = [0,90]; G = [200,1000]	T = [0,90]; G = [200,1000]; Nsc = [1,72]	T = [0,90]; G = [200,1000]; Nsh = [1,3]; Psh = [20,80]
Total Standard = 91·800 = 72.800	Total SC = 91·800·72 = 5.241,600	Total SH = 91·800·3·60 = 13.104,000
Total = 72,800 + 5.241,600 + 13.104,000 = 18.418,400		

A 10000 entries database was constructed. It was decided to be this size due to computational time, and as can be seen, it is not a number that will cause overfitting for an ANN. To give an insight into the time spent on this process, 10000 entries took approximately 12 hours to complete, in a student's laptop. However, when measuring in a real system, there will be oscillations in values that cannot occur in a simulation. To create this effect in the database, a random matrix, with the same size of the dataset, was manufactured. It had values, in percentage, between -5 and +5 which were multiplied with the dataset, causing a variation in all variables, creating noise in the results. After finalizing both matrices, the simulated and the noise were put together creating a 20000 matrix database. In the appendix C a flowchart of the process of construction of the database is presented, displaying the most important steps.

Chapter 5

Artificial neural network development

In this chapter, the artificial neural network is presented, tested with several builds and explained in detail. An explanation for each algorithm chosen will be given, and all the choices made will be scrutinised.

5.1 Artificial neural network inputs

The beginning of an ANN starts with input variables essential to the output and to differentiate them from each result. Having a high number of inputs will increase the computational cost of an ANN and will augment its complexity. To get an I-V curve temperature and irradiance are essential, not having one of these will make it impossible to obtain accurately a behaviour. An easily understandable example would be a standard operation with a sudden decrease in irradiance, which would trigger a decrease in current. However, if there is no indication of that variable, the system may indicate an abnormal mode. In temperature, a decrease will cause a reduction of the voltage and a small increase in current, when compared to the variation of the voltage. Hence a reduction of power will be felt. Now thinking about only having access to the inverter constrains the utilization of the I-V curve in all its expansion. The inverter stores the points of maximum power, the voltage and the current, and these would be the other two variables to introduce in the ANN.

5.2 Pre-processing data

The input, as observed in the previous chapter, will contain four variables: temperature, irradiance, maximum power current and maximum power voltage. Each has its unit, and they all have different figures in their respective units. Making it more difficult for the artificial neural network algorithms to converge on a unique function capable of correlating input to output. To facilitate this process and also as a way of trying to reduce variables in the input, several approaches to normalizing the input were tried. First, the temperature was taken out and divided from the voltage and current. However, as it proved, removing this variable difficulted the process of the ANN, proving to be inefficient. The same happened with the irradiance, with even worst results for being the number with more figures in it causing a reduction in all variables to low numbers harder to distinguish. Finally, a better normalization was thought out to establish all maximum values close to the output code. The voltage was divided by the open-circuit voltage, the current by the short circuit current, the temperature by the maximum temperature and the irradiance by the maximum irradiance values.

5.3 Output code

Transforming three states of operation in a number is the main goal here. The ANN has a function that converges to the numbers chosen as output, and if balanced correctly, it can widely improve the accuracy of the method. The outputs will be each specific to the behaviours chosen: standard, short-

circuit and shading. To make them unique three output variables were created, and from that, an output configuration was designed, as can be seen in Table 5.1.

Table 5.1 - Modes of operation and their respective output code

Mode of operation	Standard	Short-circuit	Shading
Output code	[0 ; 0 ; 1]	[0 ; 1 ; 0]	[1 ; 0 ; 0]

5.4 Algorithms

A feed-forward network was chosen as the artificial neural network for this study. It applies what was thought out. It is designed to be trained by given inputs into respective outputs, depending on the combination of the inputs an output is obtained. The nominated behaviours all have specific combinations of inputs that produce the desired output. However, it is not a linear problem; thus, the utilization of the neural network, which provides a pattern for each mode. It consumes less time than obtaining all results. To allow the network to 'understand' the pattern training is required, here the database is divided into three parts: training, validation, testing. The training part is established to be 70% of the whole database, and the remaining parts are each divided into 15%. These values are previously chosen and can be changed; in Table 5.2, a database division can be seen.

Table 5.2 – Database division in ANN

Training (70%)	Validation (15%)	Testing (15%)

This process is random, meaning the second training of this ANN will have 70% of the database. However, it will be constituted by different entries. It is also a randomizing process to accurately provide the best pattern recognition tool. In this section, the inputs are given, and their respective output targets, a first initiation technique is instituted. The weights and biases are fitted to give an approximation to the target. Working to achieve the best fitting possible a scaled conjugate gradient backpropagation method is applied to the training [32]. From here a mathematical model is obtained, able to predict the output from similar inputs. After this close approximation, the validation part begins, where randomly picked entries will provide an unbiased evaluation of the mathematical model fit acquired in training. The weights and biases will be optimized and the hyperparameters established may also be changed. These are set before the learning step starts, and a simple example would be the number of hidden neurons. However, in this case, these are picked and fixed in the entirety of the process. Finally, the last step takes place, the testing starts, and it is used to provide an unbiased evaluation of the performance of the final model, after training fitting and validation optimization. Now to stop the routine one of the following actions need to happen:

- the maximum number of repetitions is hit
- the maximum amount of time is exceeded
- performance reaches the minimized goal
- performance gradient falls below a minimum
- validation performance increased more than a defined number of times since the last time it decreased

If it is well structured, it will stop when the generalization of the model stops improving, demonstrated by an increase in the cross-entropy error. It compares the desired output with the actual output attained, penalizing outputs that are incredibly inaccurate and for adequately correct classifications, provides a very low penalty. Its calculation is specified in (28), for each pair of the target (t) – output (y) pair an error is calculated. Given this is a classification problem cross-entropy provides a faster and more accurate convergence to an optimal solution than minimum squared error (MSE), as referred to in [33]. The decision boundary in a classification problem is significant, and MSE does not punish misclassifications enough.

$$ce = -t \cdot \log(y) \tag{28}$$

Assessing the performance of the finalized mathematical function is done by two methods, a performance graph of a cross-entropy error vs epochs (repetitions) is the first approach. It plots the error development along with the repetitions made to reduce it. In Figure 5.1 can be seen as an illustrative performance plot. It shows the best approximation it can reach and each of the three parts curves. To judge and understand how a performance curve is considered 'good' and thus decide that the ANN is well fitted, there are two factors: the first applies to the curve and if it can reach a convergence point, maintaining a constant slope without sudden peaks. The second is the relation of each curve. If they all converge together, it demonstrates the ANN is qualified for the problem at hand, considering there are no dissimilarities.
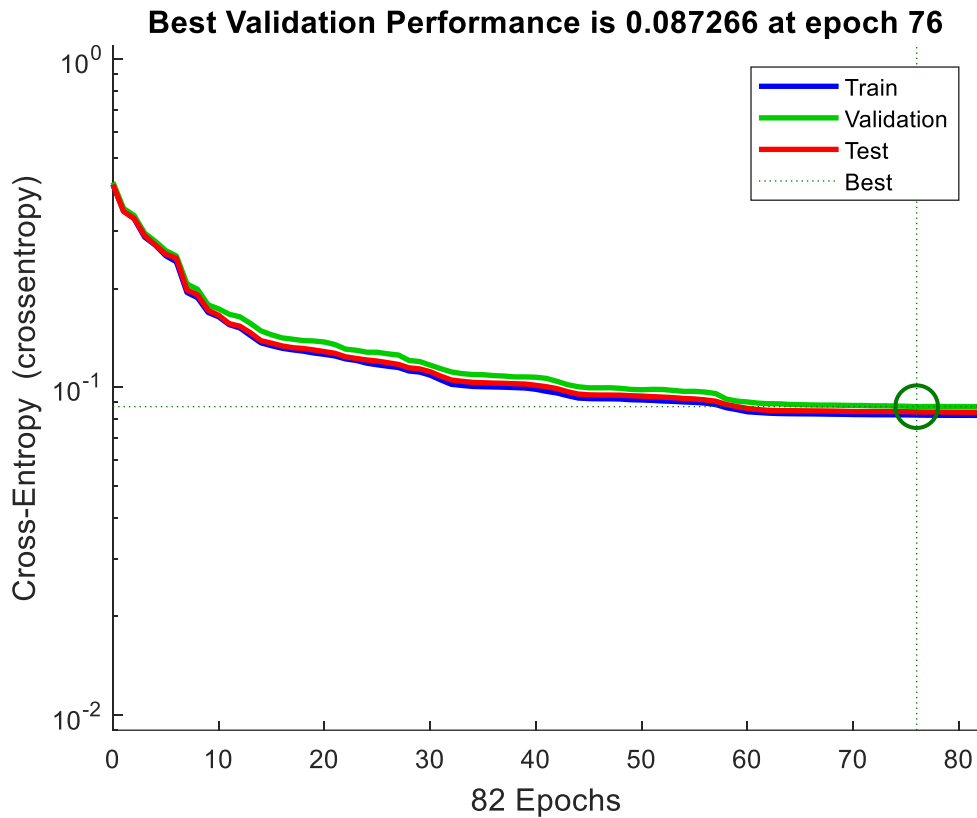
Figure 5.1 – Illustrative example of a performance plot

Consequently, the two next figures presented, Figure 5.2 and Figure 5.3, are the performance plots of the 10000 database with no noise added and the 20000 database with noise. As expected, the error will be lower in the dataset with no noise. However, there is a slight distance between the curves, causing a higher variation in the ANN provoking less correct predictions and more mistakes with other modes. In the performance graph of the 20000 dataset can be seen an apparent convergence of all curves.
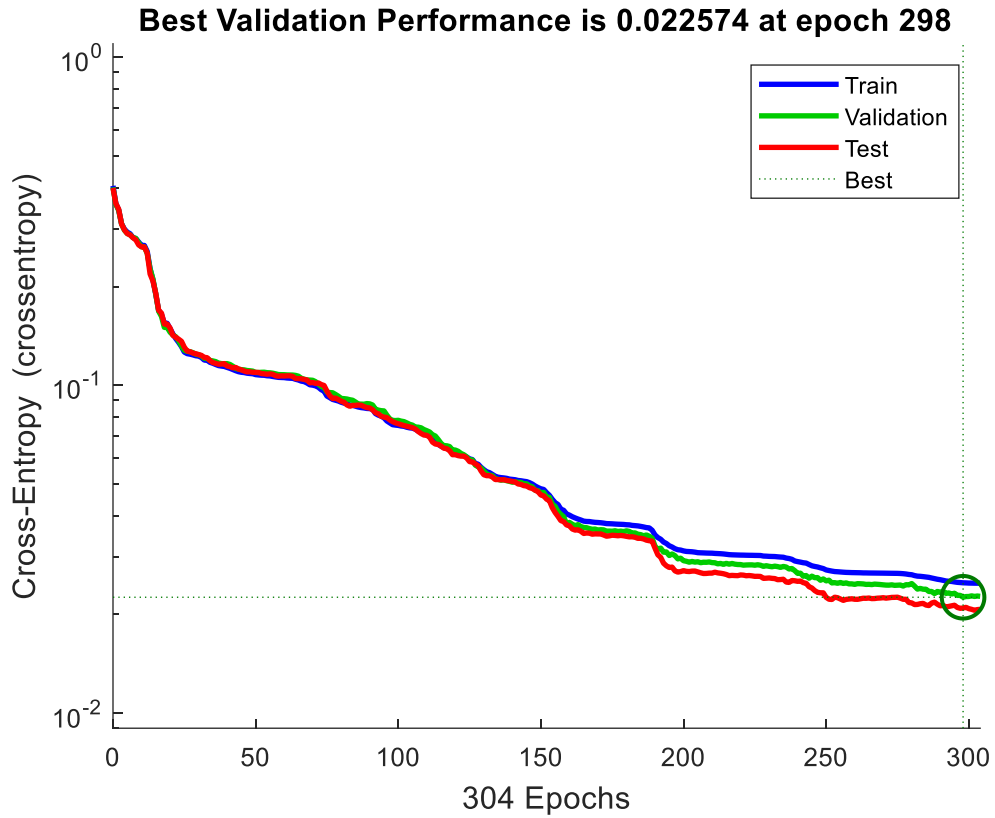
Figure 5.2 – Performance graph of the 10000 database with no noise added
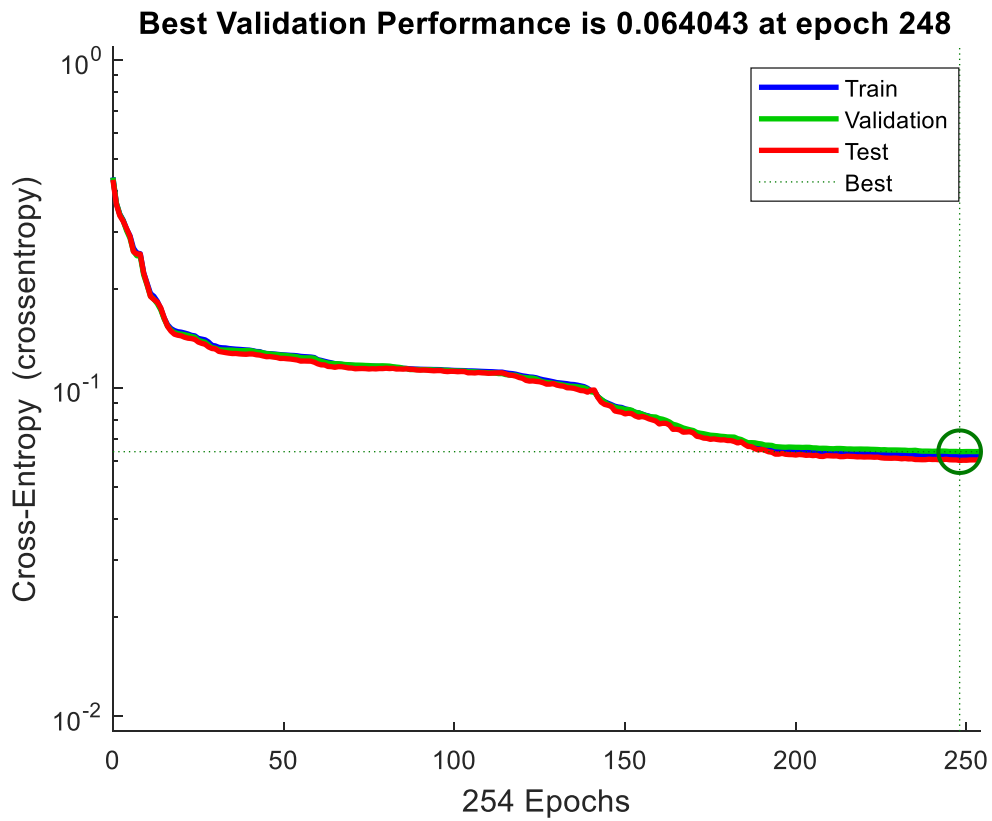
Figure 5.3 – Performance graph of the 20000 database with noise added

The second approach is with a confusion matrix, where a comparison between the different output codes obtained through the mathematical model and the targets is made and from that a matrix defining the relation within each output code and with whose it confused it. For a more direct and concise explanation, a confusion matrix is shown in Figure 5.4. There it can be seen in the rows of the matrix the output codes, 1 representing [1 0 0] (shading), 2 representing [0 1 0] (short-circuit), 3 representing [0 0 1] (standard). In the columns the target codes are represented, the interconnection between both express the failed try or a correct try. An example is easier to explain, the number in the position (1,1) in the matrix refers to the correct output code and how many it got right, specific to 1, obtained with the model. In position (2,1) is the number of times the model confused the target code 1 with output code 2, and in position (3,1) refers the number of times it missed target code 1 and swapped it with output code 3. The last row exhibits the percentage of correct and wrong fittings, where the target was designated 1, and the output was distributed by different codes, having a failure rate of 27.3% and a positive rate of 72.7%. When observing and comprehending if the final result is an acceptable neural network the most important trait to examine should be the test part, due to it being the final product and no more corrections are implemented, contrary to training and validation.
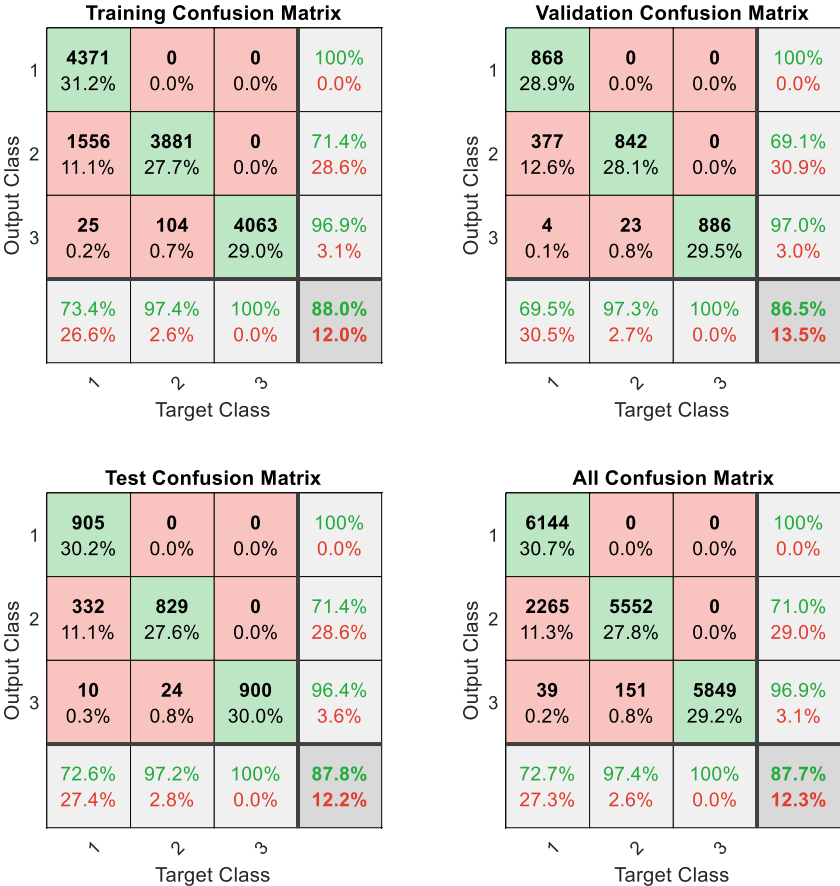
Figure 5.4 - Illustrative example of a confusion matrix

For each part will be done a confusion matrix, for the training, the validation and the testing. A final matrix is completed with all values of each singular part of the technique. In Figure 5.5, a visual representation of the four matrices is given. These are the matrices obtained from the dataset with 10000 normalized entries, and no noise added.
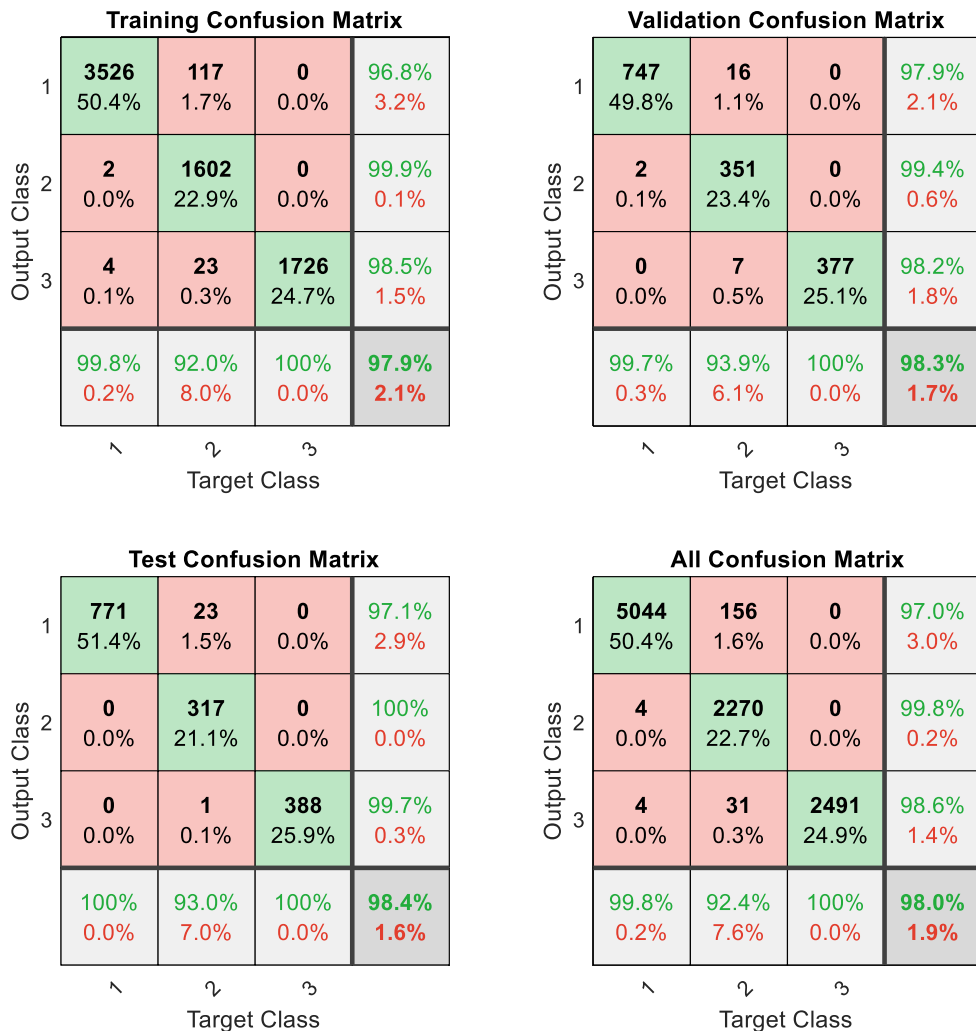
Figure 5.5 - Confusion matrices of all steps in a learning process of an ANN, in this specific case 10000 normalized entries with no noise added

As stated in chapter 4, random noise was added into the 10000 entries database to imitate possible discrepancies between simulated and experimental results. So, this dataset is a 20000 entry one with noise added, in the form of a 5% error in all input variables. In Figure 5.6 is represented the matrices equivalent to Figure 5.5 related to this new dataset.
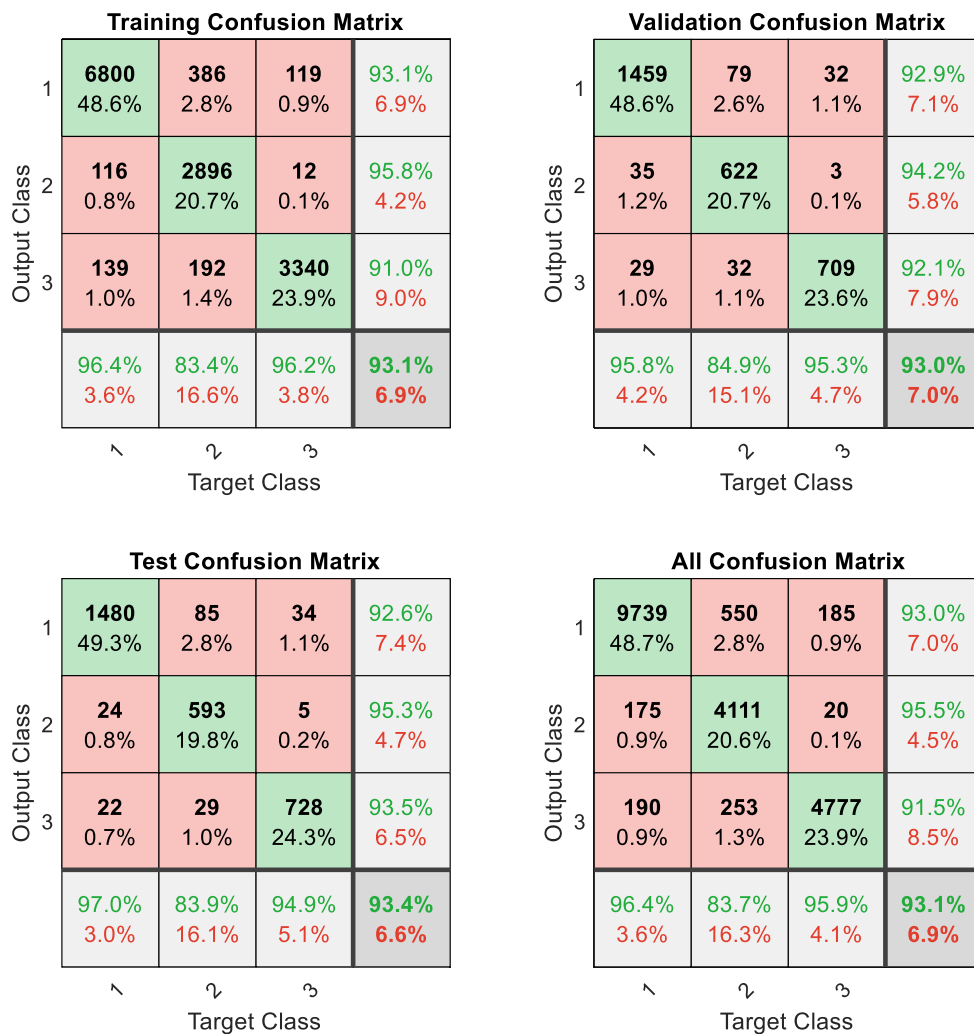
Figure 5.6 – Confusion matrices of all steps in a learning process of an ANN, in this specific case 20000 normalized entries with noise added

With the first dataset, there is overfitting. It is noticed when comparing the training error with the testing error, where an increase in this value is considerable, it proves there may be overfitting in the training step. There is also a definite drop in correct classification and a bump in error, due to the addition of noise. However, it guarantees reliable variations of experiments and measurements.

Chapter 6

Testing an artificial neural network

In this chapter, the final artificial neural network model will be used to provide conclusions about the behaviour of a photovoltaic panel. An outdoor experiment was accomplished, and different behaviours were tested and gathered all the inputs needed for the network. A comparison between what happened and what the ANN guessed will be presented, and the more common mistakes will be shown.

6.1 Outdoor testing conditions

An outdoor experiment was tested, as indicated in chapter 3, with known behaviours, described in Table 3.2. That experience was executed to have a term to compare with the results of the produced model. In the previous chapter, the artificial neural network was trained, for deducting the state of a single photovoltaic panel, and from that, a mathematical model was attained capable of that. Now an analogy is to be made between the established modes induced and the output of the prototype. The inputs can be gathered from the inverter in operating conditions, considering an mppt tracker included in the system, however in the experiment the whole I-V curve was drawn and from that the maximum power points were retrieved, both voltage and current. For cell temperature, two multimeters were used and for irradiance a solar meter. In a conventional system without resource to the temperature measuring devices, a formula (29) can be used. It only requires the datasheet of the panel, for the value of NOCT, the irradiance meter and the ambient temperature (θ_a).

$$T_{cell_k} = \theta_a + \frac{G(NOCT - 20)}{800} \tag{29}$$

6.2 Results from outdoor testing

A matrix compiling all the values of the outdoor experiment was inserted as input to the model. These are present in the appendix D. From that an output was reached, by the function, in Figure 6.1 is shown the correct estimates, in green, and the wrong, in red. It presents the correct output and its assumption.
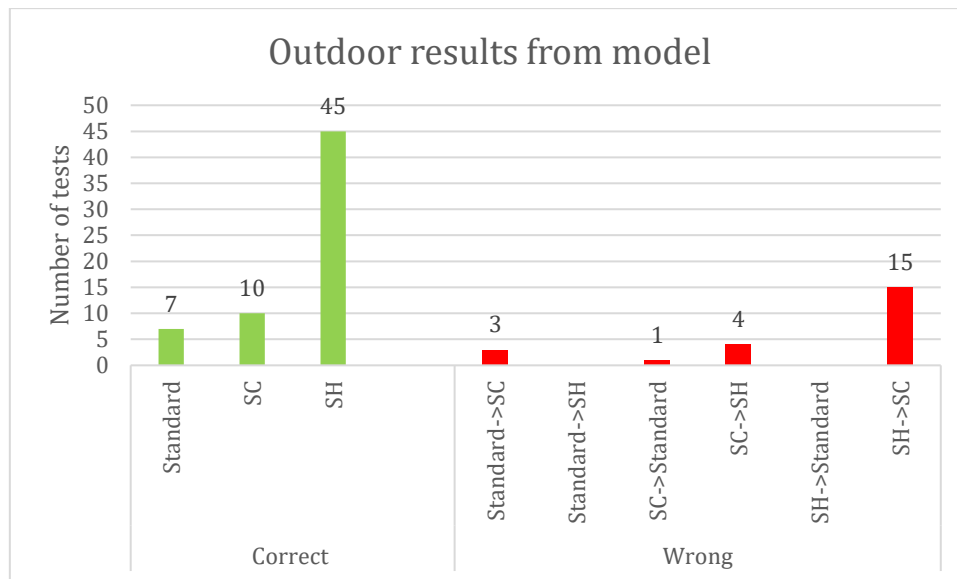
Figure 6.1 - Outdoor results after passing through the trained model

The first name tag indicates the targeted behaviour and the second indicates the one found by the model. For example, in Standard → SC, the panel is in a standard behaviour, however the model indicates a short circuit operation. In the cases where there is only one name tag, it indicates the model chose the mode of operation correctly. The graph is a more illustrative way of showing the results, hence being the first to be presented to the reader. However, it does not present all the information. The decision of choosing if it is determined mode is designated by the probabilities the model outputs. These are presented in five tables, Table 6.1, Table 6.2, Table 6.3, Table 6.4, Table 6.5. The tables were divided to better suit the page of the document and by fault condition. However, in shading another split was made based on subtypes of shading. In the graph previously presented, it is a defined behaviour when the output grants more than 50% of a mode. In a first look at the failure detection, caring only if it managed to know if the system was working without faults or not, it showed a performance of 95% efficiency. It indicated 81 correct outputs, 3 false alarms and 1 unnoticed failure.

Table 6.1 - Output results of standard condition, from outdoor testing, after passing through the mathematical model

Denomination	Shading (%)	Short-Circuit (%)	Standard (%)
Standard	0,0	23,4	**76,6**
	0,0	37,8	**62,2**
	0,0	29,5	**70,5**
	0,0	26,6	**73,5**
	0,0	32,6	**67,4**
Another Standard	0,0	**59,8**	40,3
	0,0	2,5	97,5
	0,0	**65,4**	34,7
	0,0	**52,7**	47,3
	0,0	1,5	**98,5**

Table 6.2 - Output results of short circuit condition, from outdoor testing, after passing through the mathematical model

Denomination	Shading (%)	Short-Circuit (%)	Standard (%)
Short circuit (1 cell)	0,0	**68,9**	31,1
	0,0	**67,7**	32,4
	0,0	**61,2**	38,8
	0,0	**63,6**	36,4
	0,0	**71,6**	28,4
Short circuit (24 cells)	0,0	11,5	**88,5**
	93,4	6,6	0,0
	15,5	**84,5**	0,0
	20,8	**79,2**	0,0
	96,4	3,6	0,0
Short circuit (48 cells)	**99,9**	0,0	0,0
	99,9	0,0	0,0
	3,5	**96,4**	0,0
	6,1	**93,9**	0,0
	43,4	**56,6**	0,0

Table 6.3 - Output results of one cell shading condition, from outdoor testing, after passing through the mathematical model

Denomination	Shading (%)	Short-Circuit (%)	Standard (%)
Shading (1 cell, 25%)	**99,9**	0,0	0,1
	99,9	0,0	0,1
	99,9	0,0	0,1
	99,9	0,0	0,1
	99,9	0,0	0,1
Shading (1 cell, 50%)	11,2	**88,8**	0,0
	16,3	**83,7**	0,0
	10,5	**89,5**	0,0
	2,7	**97,3**	0,0
	12,0	**88,0**	0,0
Shading (1 cell, 75%)	8,4	**91,7**	0,0
	6,1	**93,9**	0,0
	6,7	**93,3**	0,0
	8,8	**91,2**	0,0
	7,5	**92,5**	0,0
Shading (1 cell, 100%)	2,1	**97,9**	0,0
	8,5	**91,6**	0,0
	0,8	**99,2**	0,0
	8,0	**91,7**	0,0
	2,0	**98,0**	0,0

Table 6.4 - Output results of two cell shading, one in each substring, condition, from outdoor testing, after passing through the mathematical model

Denomination	Shading (%)	Short-Circuit (%)	Standard (%)
Shading (2 cells, one in each substring,25%)	99,9	0,0	0,1
	99,9	0,0	0,1
	99,7	0,0	0,3
	99,7	0,0	0,3
	99,7	0,0	0,3
Shading (2 cells, one in each substring,50%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
Shading (2 cells, one in each substring,75%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
Shading (2 cells, one in each substring,100%)	97,8	2,2	0,0
	99,0	0,9	0,0
	99,0	0,9	0,0
	99,1	0,9	0,0
	99,0	0,9	0,0

Table 6.5 - Output results of two cell shading, two in each substring, condition, from outdoor testing, after passing through the mathematical model

Denomination	Shading (%)	Short-Circuit (%)	Standard (%)
Shading (6 cells, two in each substring,25%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
Shading (6 cells, two in each substring,50%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
Shading (6 cells, two in each substring,75%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
Shading (6 cells, two in each substring,100%)	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0
	99,9	0,0	0,0

From the analysis of the graph, it can be seen that the standard operating condition is confused with short circuit only. It is caused, as previously mentioned in chapter 4, by a low difference in voltage drop, mainly in the test with only a short-circuited cell, a decrease in 0.5 volts. There is also a particular case in the short circuit of 24 cells in a substring that caused an error in the classification. It chose standard instead of short circuit due to the claw. The one used to short circuit the whole substring by connecting it with both terminals of the bypass diode, not providing contact to the metal. In Figure 6.2, the I-V curves of that sample of the test are drawn, and it shows the inefficiency of the claw in the blue curve where it fails to represent the loss of approximately 12 volts.
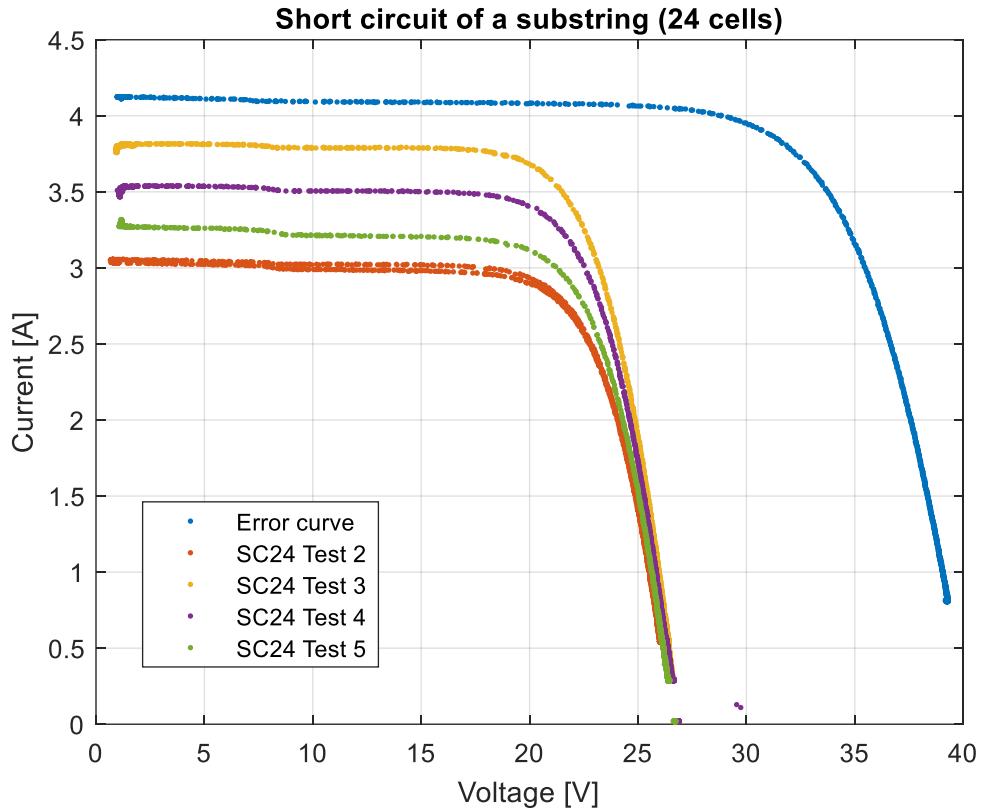
Figure 6.2 - I-V curves of a short circuit in a substring (24 cells) of the outdoor experiments

The rest of the tests, consisting of the standard mode were correct, just with different irradiances which caused changes in the current. By looking at Table 6.1 and Table 6.2, it can be noticed in some cases, proximity in probabilities between the short-circuit and standard operation. This is again caused by the proximity in voltage drops from one case to another. There was no complication in accessing the differences with standard and shading, meaning these two are not correlated and do not interfere with each other in the decision process. Moving on to short circuit evaluation, the same happens when a behaviour should be a short circuit. In one specific case, it classifies it as usual due to the low difference in voltage. Now the major upset is the confusion between short circuit and shading. It happens when the operating condition should be a short circuit, and it classifies it as shading. In the test occurred four times, two times in a substring short circuit and two times in a two substring short-circuit. This misinterpretation is created because when a high percentage of shading exists in a cell of a substring, said string would activate its bypass diode not to compromise the rest of the system. However, this originates a drop in voltage of approximately one substring causing confusion between a short circuit of a substring (24 cells) and shading. This behaviour is only similar when the short circuit is exclusively equal to the number of cells in one substring or more. Entirely comprises of the number of cells in short circuit, if the number, in a panel, is of 24, 48 or 72 it may confuse with shading. An example would be 30 short-circuited cells spread across the three substrings the behaviour compared with a high percentage of shading would be different. As the panel was enclosed and further destruction of it was not deemed necessary, these tests were not completed. In shading another failed misinterpretation of

the model arose, it misrepresented the three tests of on cell shaded with percentages of the shade of 50%, 75% and 100%, confusing it with a short circuit. Again, due to the misleading coincidences between them in these specific cases. Even with all these wrong assumptions, in all these cases, there is always a percentage stating it is not 100% correct.

The damaged panel was also tested, and abnormal behaviour was given. However, it is challenging to determine the level of damage the panel suffered, making it harder to envision its electrical parameters behaviour.

6.3 Results from model

An analysis of the outdoor results was appointed and discussed in the previous subchapter, yet another scrutiny was decided to be relevant to mention to the reader, that being the chapter 3 model results. The previous analysis gathered a 73% accuracy however here lies an associated error from the simulation and from the trained neural network results. Using the variables gathered in the outdoor results (temperature and irradiance), a simulation was made, and those values were inputted in the artificial neural network to demonstrate this propagated error. In Figure 6.3 the results can be seen.
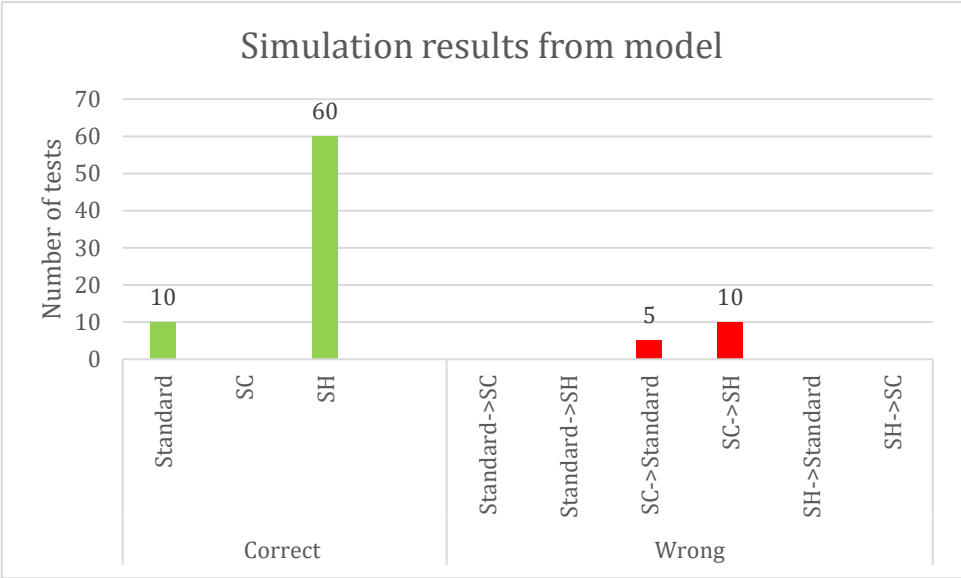
Figure 6.3 - Simulation results after passing through the trained model

In this situation, there are unambiguous correct patterns. The standard and shading modes demonstrate an unambiguous behaviour which the ANN model can unequivocally classify. It displays the wrong indicator in short circuit mode of operation. It misinterprets the one short-circuited cell with the standard condition and both the substring (24 cells) short circuit and the two substrings (48 cells) short circuit with shading operation.

6.4 Summary of results

Starting with the database results, there is a clear area where the trained model has difficulty accessing. It is the short circuit boundaries. If the short circuit is low, it struggles to decide between standard and short circuit. If the short circuit is equal to a substring or more than one, it coincides with the shading operation. From a total of 85 simulated tests, the trained model deduced correctly 70. It portraits an accuracy of 82,35%. The database and the trained model are correlated. One created the other, and this was mainly analysed to understand what could be improved, this will be referred to in the last chapter.

In the outdoor results, a more different situation is met, there are more imprecise outputs. Standard and short circuit get mistaken between each other with a low-level drop of voltage. The measurement error rate can even mistake a short circuit when standard behaviour is applied. Moreover, for shading and a short circuit occurs the misinterpretation stated in the above paragraph. However, there is an indication in the probabilities, showing the feasibility of another behaviour. A total of 85 outdoor tests were performed, and from that sample, 62 were correctly diagnosed, giving an accuracy of 72,94%. A decrease in performance was expected, considering the measurement errors of all the equipment utilized.

Chapter 7

Conclusions

A conclusion of the work is presented in this chapter. All major points are appointed, and improvements are documented. Recommendations for future work in this area and insight is given to continue the research on this theme.

In this thesis, known behaviours of photovoltaic systems were studied, namely, the standard operating condition, short circuit and shading. Open circuits were not developed in this work as there was no way of properly test them. However, this fault has a pattern equivalent to a high percentage of shading when the bypass diode activates. After studying and learning about those faults, a knowledge of the one diode five parameter model was gathered to understand the model used to simulate cell behaviour. Then the simulation model was optimized to be able to create the best reference values of each behaviour. First a static database was created, however it created overfitting in the artificial neural network training. A random database proved to be more efficient for ANN training as it did not overfit. Normalization of the input values established better results. However, only when normalized values were approximated to the interval [0;1]. Reducing the number of inputs was tried and it did not work, it proved to hinder the convergence process in the ANN. The four inputs cannot be taken out, they are all essential for an accurate detection and diagnosis. Even when one input was normalized using another it proved to complicate the analysis of the artificial intelligence method. Many networks were created with different hidden layer nodes however the ones with five showed the best results. The addition of noise occurred with the intention of creating a threshold in the values, to create the measurement error threshold in the network. It demonstrated overall positive effects, also in removing the overfitting. Then the outdoor test in a photovoltaic panel was done and compared the accuracy of the trained model of the ANN in outdoor circumstances and the accuracy of the simulation model. The primary observation made was of the detection method, where it achieved a performance of 95%. In the diagnosis the accuracy was 73%. In the detection department it displays a very solid achievement. In regard to diagnosis performance, it decreased due to measurement errors and error propagation of all applied methods and techniques, however it showed promising results.

From collected results, there are improvements recommended to be applied in the database creation. Topology consisting of more than one panel should be simulated. However, the computational cost will increase in the process of forming these new databases. Another enhancement would be focusing on the struggle of the artificial neural network on the differentiation of the short circuit and regular operation. It should include a restriction of the lower level of a short circuit, a threshold. Meaning the minimum in a panel detection would be 10% of the total voltage. This percentage was based on the measuring devices used in the experiments, and they were not optimal. In the example, this specific panel utilized in testing and simulation has close to 36 V, 10% of that would be 3.6 volts what amounts to approximately seven cells (3.5 V). It creates a more considerable barrier between these modes and prevents errors from measurements. In a system with more panels, this would still be equivalent to 10% of the total voltage of the system. If the measuring devices diminish their error rates, this percentage can also decrease.

In the artificial neural network enhancements, the more prone to be implemented would be a new input entry for topology. Allowing the information on the number of panels in series to be known by the network. If the model provides excellent accuracy, to simplify the design of the structure, the topology value can be inserted in the V_{mpp} input, by dividing it. It would cause a total voltage of a system to be normalized to a total panel voltage. If it leads to reduced accuracy, this idea should be discarded.

For open-circuit fault the model trained and constructed could still be used to detect them.

Considering these have the same behaviour as shading, which is a temporary fault when shading fault is identified, a program could run several times in different hours along the day. If it changes the diagnosis to standard, it should be a shading operation, that occurs only during a particular hour. However, if it stays along the day, it can indicate an open circuit.

Constructing databases across all photovoltaic systems can bring new awareness for diagnosis methods and possibilities for a more straightforward implementation of artificial intelligence in these. Crossing the simulated values with real systems can augment the perception of new patterns in PV diagnosis branch. Studies across new and older systems can even bring an ageing chart of various panels, making available a unique diagnosis method that can predict faults based on minimal factors influencing its future behaviour.

References

[1]   S. R. Madeti and S. N. Singh, "Monitoring system for photovoltaic plants: A review," *Renew. Sustain. Energy Rev.*, vol. 67, pp. 1180–1207, 2017.

[2]   "IRENA – International Renewable Energy Agency." [Online]. Available: https://www.irena.org/. [Acessed: 30-Oct-2019].

[3]   M. Green, *Improving Efficiency of PV Systems Using Statistical Performance Monitoring, IEA International Energy Agency, IEA PVPS Task 13, Subtask 2 Report IEA-PVPS T13-07:2017*. 2017.

[4]   "How To Visually Inspect A Solar - Renewables in Africa." [Online]. Available: https://www.renewablesinafrica.com/how-to-visually-inspect-a-solar/. [Accessed: 13-Sep-2019].

[5]   "Quality Inspection for Solar Modules - Raw Material, Manufacturing & …." [Online]. Available: https://www.slideshare.net/AnmolJaggi/quality-inspection-for-solar-modules-raw-material-manufacturing-lab-tests-150418. [Accessed: 13-Sep-2019].

[6]   "Solar Electricity - Photovoltaic Systems and Components, Grid-Connected Solar Electric Systems, Off-Grid (Stand Alone) Solar Electric Systems, PV Modules, PV Inverters, PV Chargers, PV Battery, PV Mounting, Small Solar Electric Devices." [Online]. Available: http://www.solardirect.com/pv/systems/systems.htm. [Accessed: 13-Sep-2019].

[7]   "Photovoltaic effect - Energy Education." [Online]. Available: https://energyeducation.ca/encyclopedia/Photovoltaic_effect. [Accessed: 13-Sep-2019].

[8]   "The Difference Between Solar Cell, Module & Array | Samlex Solar." [Online]. Available: https://www.samlexsolar.com/learning-center/solar-cell-module-array.aspx. [Accessed: 27-Oct-2019].

[9]   "Technology Fundamentals: Photovoltaic systems." [Online]. Available: https://www.volker-quaschning.de/articles/fundamentals3/index.php. [Accessed: 13-Sep-2019].

[10]   S. Lineykin, M. Averbukh, and A. Kuperman, "Five-parameter model of photovoltaic cell based on STC data and dimensionless," *2012 IEEE 27th Conv. Electr. Electron. Eng. Isr. IEEE 2012*, no. November 2012.

[11]   M. Köntges *et al.*, *Performance and reliability of photovoltaic systems subtask 3.2: Review of failures of photovoltaic modules : IEA PVPS task 13 : external final report IEA-PVPS*. 2014.

[12]   S. Daliento *et al.*, "Monitoring, diagnosis, and power forecasting for photovoltaic fields: A review," *Int. J. Photoenergy*, vol. 2017, 2017.

[13]   V. J. Chin, Z. Salam, and K. Ishaque, "Cell modelling and model parameters estimation techniques for photovoltaic simulator application: A review," *Appl. Energy*, vol. 154, pp.

500–519, 2015.

[14] E. Garoudja, F. Harrou, Y. Sun, K. Kara, A. Chouder, and S. Silvestre, "Statistical fault detection in photovoltaic systems," *Sol. Energy*, vol. 150, pp. 485–499, 2017.

[15] A. Woyte, M. Richter, D. Moser, S. Mau, N. Reich, and U. Jahn, "Monitoring of Photovoltaic Systems: Good Practices and Systematic Analysis," *28th Eur. PV Sol. Energy Conf. Exhib. 30 Sept. – 4 Oct.*, vol. 53, no. 9, pp. 1689–1699, 2013.

[16] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan, "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks," *Renew. Energy*, vol. 90, pp. 501–512, 2016.

[17] H. Braun *et al.*, *SYNTHESIS LECTURES ON POWER ELECTRONICS Signal Processing for Solar Array Monitoring, Fault Detection, and Optimization About SYNTHESIS LECTURES ON POWER ELECTRONICS.*

[18] Y. Zhao, L. Yang, and B. Lehman, "Decision Tree-Based Fault Detection and Classification in Solar Photovoltaic Arrays," pp. 93–99, 2012.

[19] L. Serrano-Luján, J. M. Cadenas, J. Faxas-Guzmán, and A. Urbina, "Case of study: Photovoltaic faults recognition method based on data mining techniques," *J. Renew. Sustain. Energy*, vol. 8, no. 4, 2016.

[20] D. Ji, C. Zhang, M. Lv, Y. Ma, and N. Guan, "Photovoltaic array fault detection by automatic reconfiguration," *Energies*, vol. 10, no. 5, 2017.

[21] A. Mellit and S. A. Kalogirou, "Artificial intelligence techniques for photovoltaic applications: A review," *Prog. Energy Combust. Sci.*, vol. 34, no. 5, pp. 574–632, 2008.

[22] "File:Artificial neural network.svg - Wikimedia Commons." [Online]. Available: https://commons.wikimedia.org/wiki/File:Artificial_neural_network.svg. [Accessed: 13-Sep-2019].

[23] "Artificial neural network." [Online]. Available: https://en.wikipedia.org/wiki/Artificial_neural_network#cite_note-Zell1994ch5.2-50. [Accessed: 16-Aug-2019].

[24] "What Is a Neural Network? - MATLAB & Simulink." [Online]. Available: https://www.mathworks.com/discovery/neural-network.html. [Accessed: 13-Sep-2019].

[25] "Preventing Deep Neural Network from Overfitting - Towards Data Science." [Online]. Available: https://towardsdatascience.com/preventing-deep-neural-network-from-overfitting-953458db800a. [Accessed: 13-Sep-2019].

[26] S. Laamami, M. Benhamed, and L. Sbita, "Artificial neural network-based fault detection and classification for photovoltaic system," *Int. Conf. Green Energy Convers. Syst. GECS 2017*, no. 1, 2017.

[27] E. A. S. Filho, F. F. Costa, A. P. N. Tahim, and A. C. C. De Lima, "Photovoltaic panel simulation based on individual cell condition," *ECCE 2016 - IEEE Energy Convers. Congr. Expo. Proc.*, 2016.

[28] P. Gilman, "SAM Photovoltaic Model Technical Reference SAM Photovoltaic Model Technical Reference," *Sol. Energy*, vol. 63, no. May, pp. 323–333, 2015.

[29] "Advancing the Science of Solar Data | National Solar Radiation Database (NSRDB)." [Online]. Available: https://nsrdb.nrel.gov/. [Accessed: 13-Sep-2019].

[30] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, "The National Solar Radiation Data Base (NSRDB)," *Renew. Sustain. Energy Rev.*, vol. 89, no. January

2018, pp. 51–60, 2018.

[31] "Error Propagation - Harvard university," no. 3, pp. 3–7, 2013.

[32] M. F. Møller, "A scaled conjugate gradient algorithm for fast supervised learning," *Neural Networks*, vol. 6, no. 4, pp. 525–533, 1993.

[33] P. Golik, P. Doetsch, and H. Ney, "Cross-entropy vs. Squared error training: A theoretical and experimental comparison," *Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH*, vol. 2, no. 2, pp. 1756–1760, 2013.

Appendix A

Wires' resistance

This section presents the wires' resistance calculated in the indoor experimental setup.

Table A.1 – Measurements, of one cell, for the calculation of the wires' resistance

Healthy Panel							
G (W/m2)	I_{sc} (A)	V (V)	V_{R1} (V)	V_{R2} (V)	V_R (V)	T (°C)	R (Ω)
1000	4,370	0,2100	0,0484	0,0685	0,1169	64,0	0,0268
800	3,250	0,1425	0,0384	0,0562	0,0946	55,0	0,0291
600	2,460	0,0946	0,0285	0,0382	0,0667	47,0	0,0271
400	1,595	0,0701	0,0192	0,0260	0,0452	39,0	0,0283
230	0,903	0,0373	0,0109	0,0136	0,0245	32,0	0,0271

Table A.2 – Error computation for correction related to the wires' resistance

V_{R1} error (%)	V_{R2} error (%)	I_{sc} error (%)	V_R error (%)	**R error (%)**
1,08	2,04	2,94	0,98	**3,10**
3,12	1,39	3,11	1,33	**3,38**
1,68	3,58	3,30	1,59	**3,66**
3,63	1,62	3,63	1,55	**3,94**
5,18	2,17	4,11	2,13	**4,63**

Appendix B

Simulation plots

This section presents the plot comparison between each simulated behaviour.

Figure B.1 – Normal and shading, in one cell (SH1), in two cells in different substrings (SH11) and three cells in different substrings (SH111), with a decrease of 50 W/m² in irradiance, both have G = 1000 W/m² and T = 70ºC

Figure B.2 - Normal and shading, in one cell (SH1), in two cells in different substrings (SH11) and three cells in different substrings (SH111), with a decrease of 500 W/m² in irradiance, both have G = 1000 W/m² and T = 70ºC

Figure B.3 – Short circuit and shading, in one cell (SH1), in two cells in different substrings (SH11) and three cells in different substrings (SH111), with a decrease of 50 W/m² in irradiance, both have G = 1000 W/m² and T = 70ºC
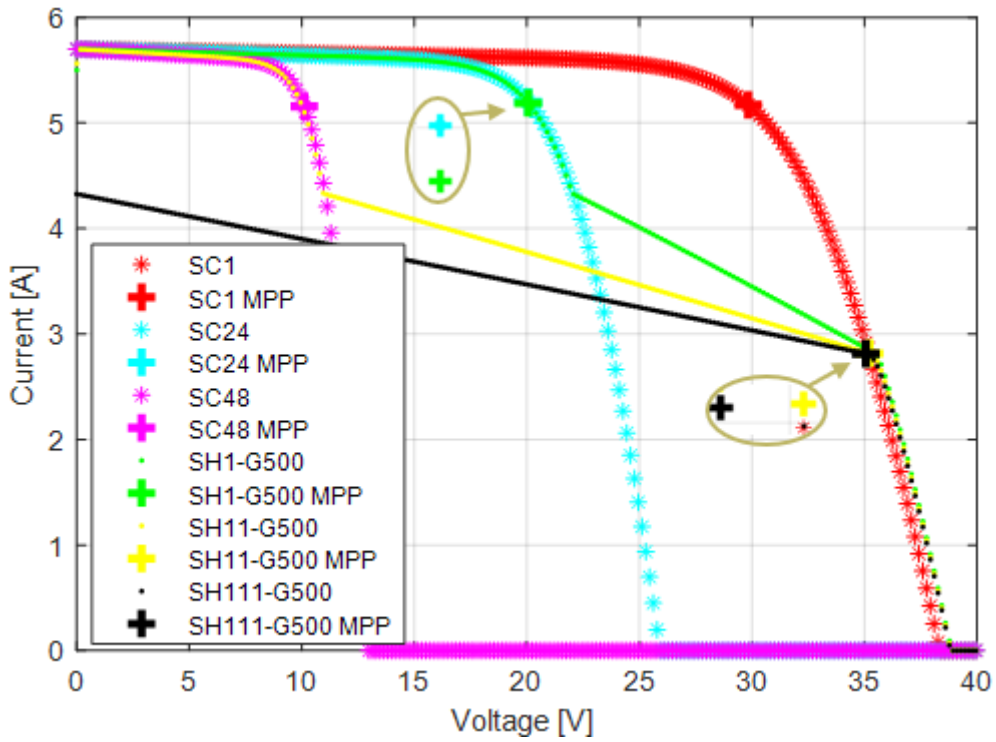
Figure B.4 - Short circuit and shading, in one cell (SH1), in two cells in different substrings (SH11) and three cells in different substrings (SH111), with a decrease of 500 W/m² in irradiance, both have G = 1000 W/m² and T = 70ºC

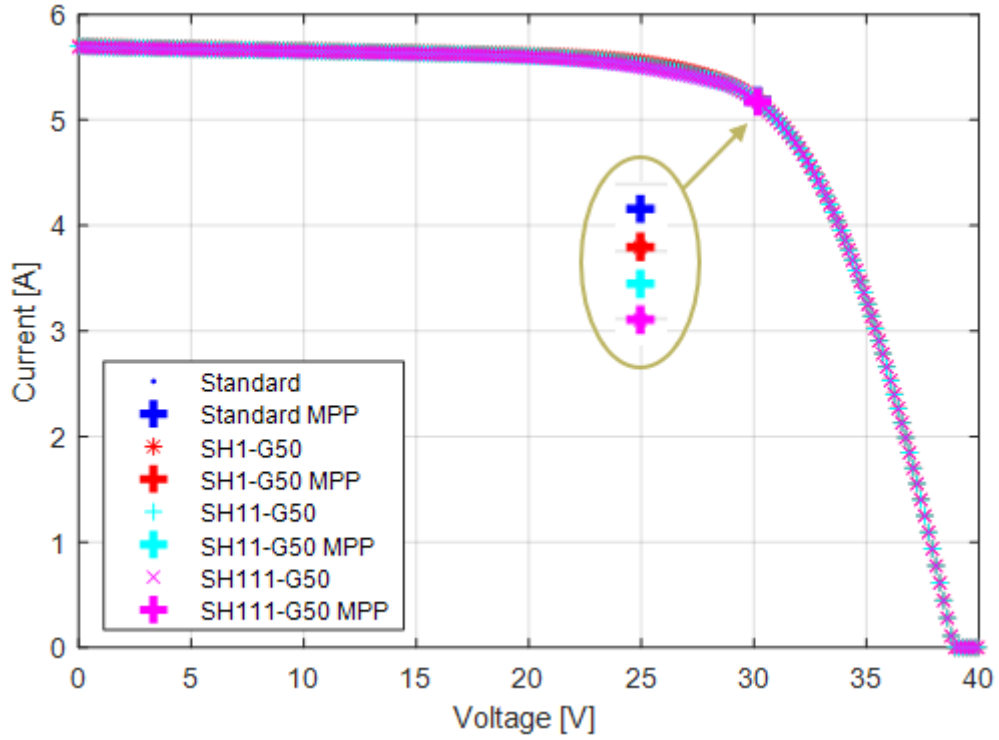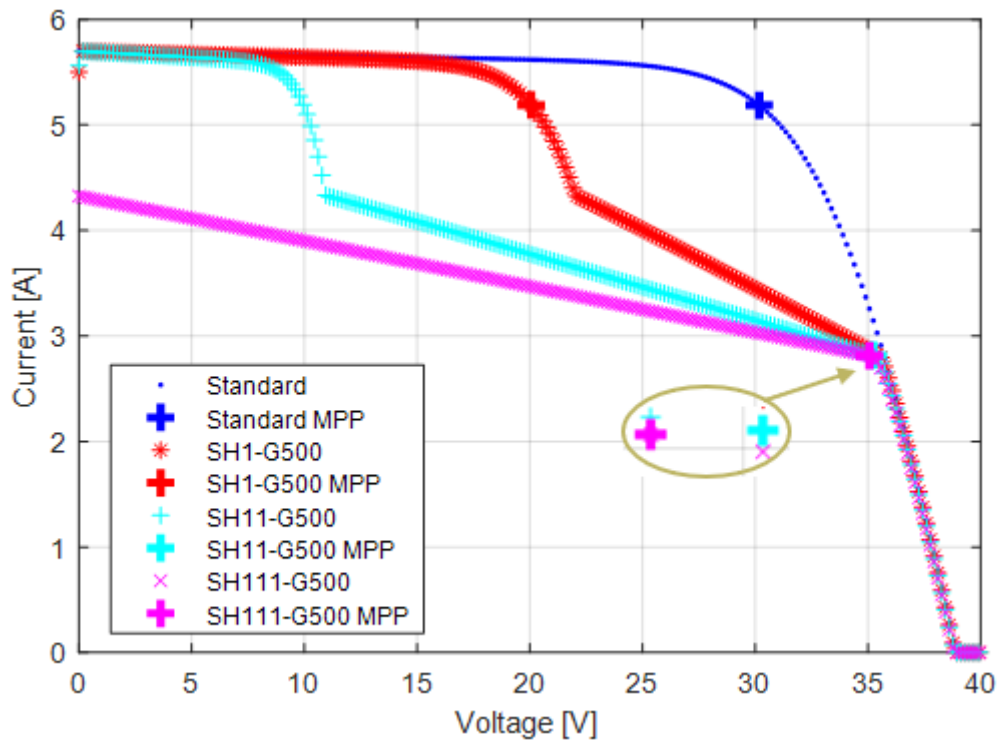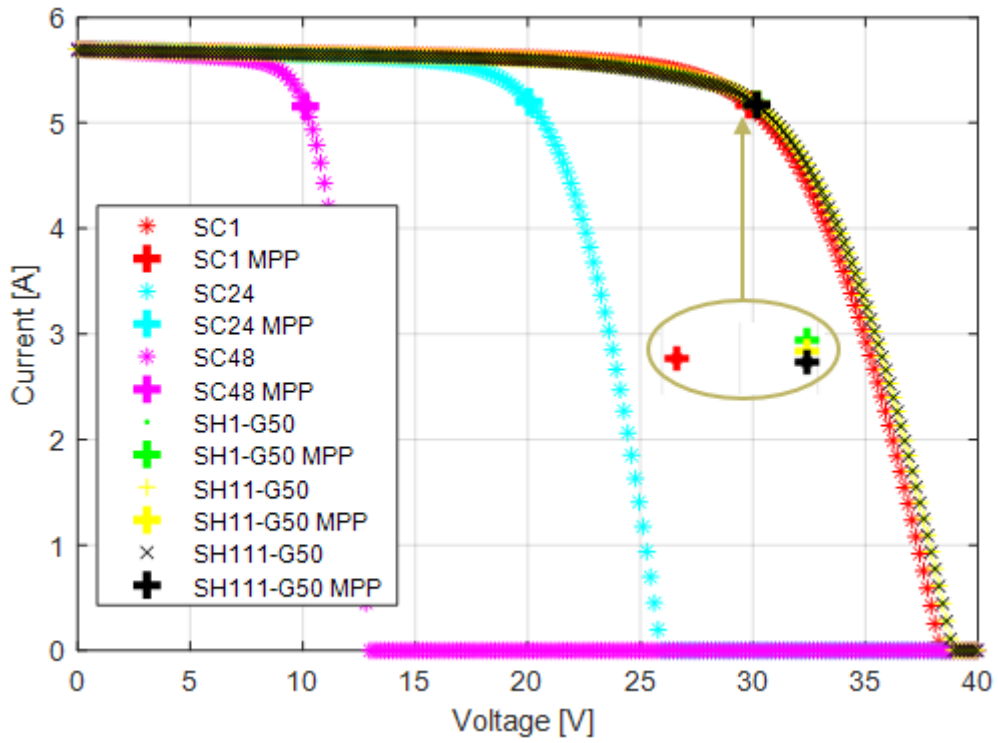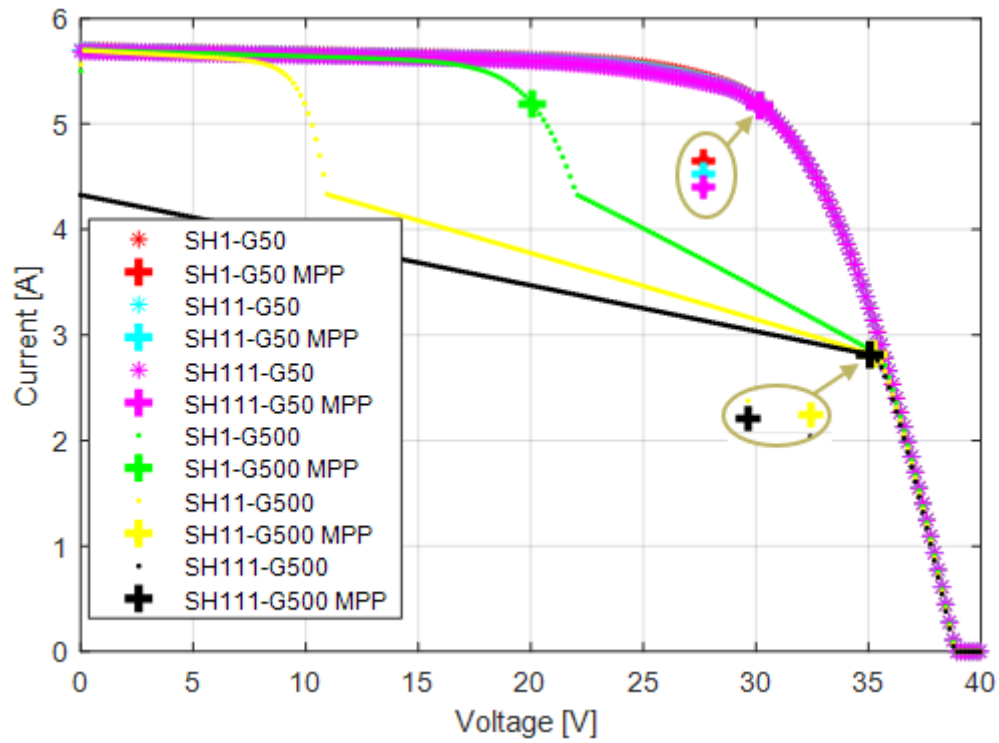*Figure B.5 - Shading, with a decrease of 50 W/m² in irradiance, and the other with a decrease of 500 W/m² both, in one cell (SH1), in two cells in different substrings (SH11) and in three cells in different substrings (SH111), both have G = 1000 W/m² and T = 70ºC*

Appendix C

Database flowchart

This section presents the flowchart of the database.
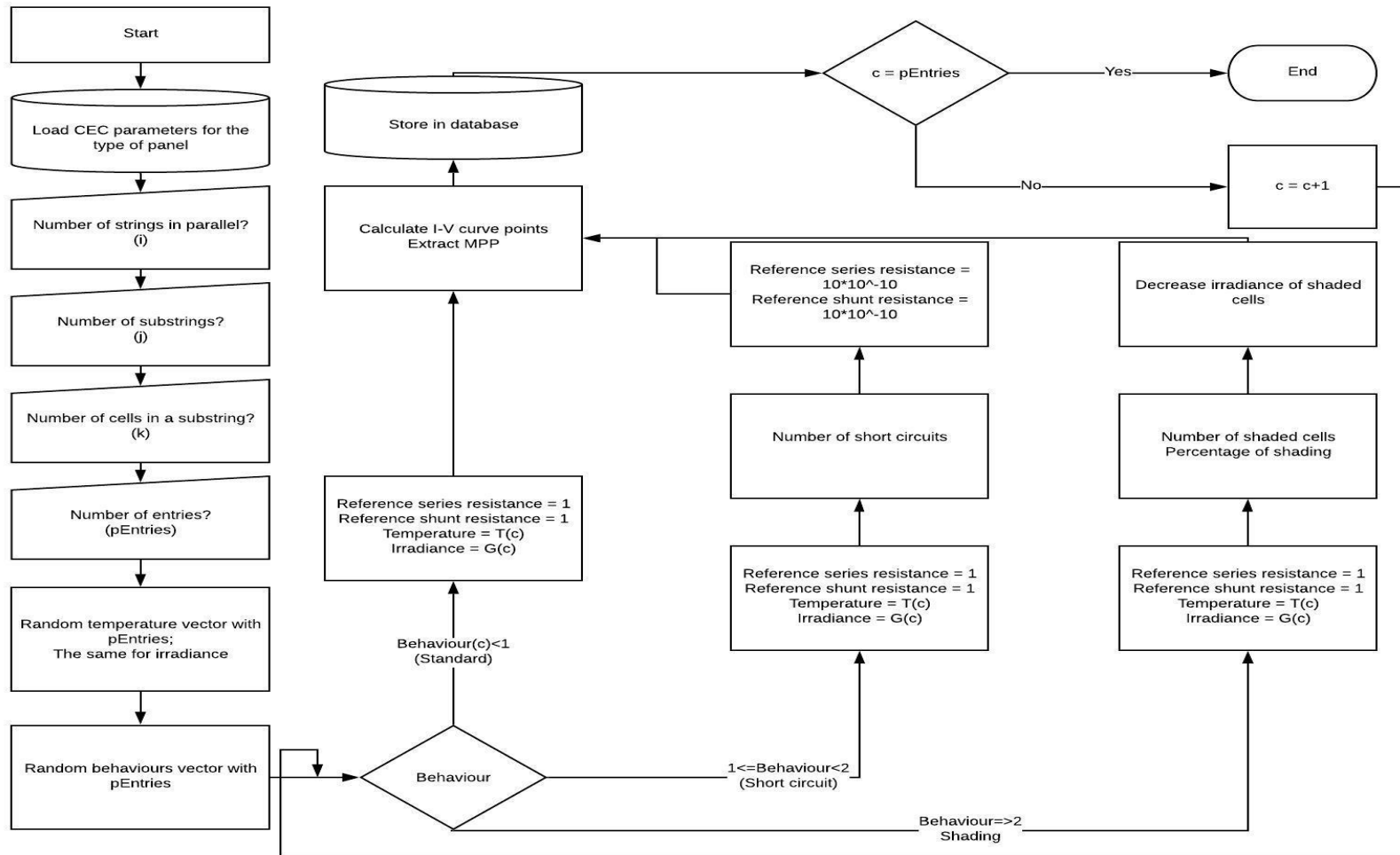
Figure C.1 – Flowchart of database

Appendix D

Outdoor test results

This section presents the outdoor test results, all normalized.

Table D.1 – Standard behaviour outdoor results, normalized

	Temperature	Irradiance	Voltage	Current
Standard	0,4667	0,6410	0,7305	0,6176
	0,4667	0,6360	0,7179	0,6244
	0,4778	0,6385	0,7215	0,6231
	0,4722	0,6415	0,7263	0,6210
	0,4667	0,6430	0,7233	0,6251
Another Standard	0,4778	0,7600	0,7089	0,7190
	0,4611	0,6490	0,6999	0,7068
	0,4611	0,6330	0,7101	0,6020
	0,4611	0,6270	0,7143	0,6081
	0,4667	0,5350	0,7149	0,6081

Table D.2 – Short circuit behaviour outdoor results, normalized

	Temperature	Irradiance	Voltage	Current
Short circuit of 1 cell	0,4778	0,7100	0,6975	0,6864
	0,4889	0,6400	0,6969	0,6210
	0,4833	0,6130	0,6999	0,6020
	0,4833	0,5900	0,7005	0,5775
	0,4833	0,5050	0,6987	0,4890
Short circuit of 24 cells	0,4556	0,6470	0,7047	0,6775
	0,4556	0,6600	0,4721	0,5006
	0,4667	0,5850	0,4721	0,6285
	0,4611	0,5630	0,4745	0,5789
	0,4611	0,7100	0,4691	0,5353
Short circuit of 48 cells	0,4444	0,8100	0,7143	0,4550
	0,4444	0,6730	0,7239	0,4931
	0,4556	0,6000	0,2262	0,8939
	0,4444	0,5200	0,2310	0,7748
	0,4444	0,4700	0,2322	0,6455

Table D.3 – Shading, of one cell in one substring, outdoor results, normalized

	Temperature	Irradiance	Voltage	Current
Shading (1 cell, 25%)	0,4611	0,6685	0,8061	0,4911
	0,4667	0,6690	0,8049	0,4918
	0,4611	0,6680	0,8055	0,4911
	0,4611	0,6680	0,8013	0,4945
	0,4611	0,6680	0,8019	0,4938
Shading (1 cell, 50%)	0,4611	0,6765	0,4721	0,6544
	0,4722	0,6765	0,4721	0,6510
	0,4722	0,6770	0,4697	0,6537
	0,4722	0,6770	0,4631	0,6619
	0,4667	0,6775	0,4715	0,6496
Shading (1 cell, 75%)	0,4556	0,7155	0,4721	0,6707
	0,4500	0,7060	0,4715	0,6626
	0,4444	0,6660	0,4727	0,6251
	0,4500	0,7000	0,4733	0,6381
	0,4556	0,7100	0,4715	0,6646
Shading (1 cell, 100%)	0,4667	0,6725	0,4631	0,6455
	0,4611	0,6930	0,4709	0,6551
	0,4667	0,6790	0,4589	0,6571
	0,4722	0,6870	0,4685	0,6544
	0,4722	0,6725	0,4619	0,6462

Table D.4 - Shading, of two cells, one in each substring, outdoor results, normalized

	Temperature	Irradiance	Voltage	Current
Shading (2 cells, one in each substring,25%)	0,4500	0,7395	0,7893	0,5516
	0,4444	0,7320	0,7833	0,5550
	0,4444	0,7050	0,7875	0,5578
	0,4556	0,7050	0,7851	0,5584
	0,4500	0,7050	0,7881	0,5578
Shading (2 cells, one in each substring,50%)	0,4556	0,7295	0,8336	0,3536
	0,4500	0,7320	0,8330	0,3536
	0,4556	0,7330	0,8342	0,3543
	0,4500	0,7330	0,8342	0,3550
	0,4500	0,7380	0,8354	0,3543
Shading (2 cells, one in each substring,75%)	0,4611	0,7395	0,8582	0,1828
	0,4722	0,7535	0,8582	0,1849
	0,4611	0,7480	0,8576	0,1842
	0,4611	0,7525	0,8570	0,1849
	0,4611	0,7435	0,8558	0,1828
Shading (2 cells, one in each substring,100%)	0,4556	0,7500	0,2004	0,8361
	0,4556	0,7390	0,2118	0,7007
	0,4556	0,7570	0,2136	0,7088
	0,4556	0,7570	0,2082	0,7306
	0,4556	0,7550	0,2136	0,7075

Table D.5 - Shading, of two cells one in each substring, outdoor results, normalized

	Temperature	Irradiance	Voltage	Current
Shading (6 cells, two in each substring,25%)	0,4500	0,7100	0,7857	0,5081
	0,4500	0,7160	0,7893	0,5047
	0,4444	0,7210	0,7923	0,4958
	0,4444	0,7445	0,7917	0,5094
	0,4500	0,7530	0,7929	0,5142
Shading (6 cells, two in each substring,50%)	0,4389	0,7910	0,8402	0,3094
	0,4333	0,7945	0,8390	0,3039
	0,4389	0,7955	0,8414	0,2971
	0,4389	0,7960	0,8426	0,2903
	0,4500	0,7945	0,8420	0,2829
Shading (6 cells, two in each substring,75%)	0,4833	0,8210	0,8318	0,2128
	0,4833	0,8140	0,8300	0,2107
	0,4833	0,8190	0,8318	0,2114
	0,4889	0,7970	0,8300	0,2073
	0,4944	0,7960	0,8306	0,2053
Shading (6 cells, two in each substring,100%)	0,4833	0,8120	0,8528	0,0039
	0,4944	0,8100	0,8522	0,0039
	0,4889	0,7910	0,8504	0,0039
	0,4889	0,7770	0,8492	0,0039
	0,4889	0,7720	0,8498	0,0039

Annex A

Equipment datasheets

In this section, the equipment's datasheets are presented.

CENTER 120 122 MULTIMETER

GENERAL SPECIFICATIONS

Display: 3-3/4 digital(4000 count) display
Sample Rate: 2 times/sec for digital reading;
20 times/sec for analog bar indication
Safety: IEC 1010 1000V CAT III
Auto Power Off: 30 minutes
Battery Type: 9V, NEDA 1604, IEC 6F22, JIS 006P
Battery Life: 100 hours typical (alkaline)
Size: 198×86×38mm (7.8"×3, 4"×1, 5")
With holster: 209×94×48mm (8.2"×307"×1.9")
Weight: Approx.430g   **With holster:** Approx.600g
Accessories: Operation manual, Teat lead, Battery ,Holster,
Temperature converter, K Type bead sensor (CENTER 122)
RS-232 cable and CENTER SE-120 software (CENTER 122)

Download data for analysis or monitoring the real time data through easy access "Slim Com" "RS-232 connection.

ELECTRICAL SPECIFICATIONS

Function	Range	Resolution	Accuracy
DCV	400mV	0.1mV	±(0.3%rdg+2dgt)
	4V	0.001V	
	40V	0.01V	
	400V	0.1V	
	1000V	1V	±(0.5%rdg+2dgt)
ACV	400mV	0.1mV	±(0.5%rdg+5dgt)
	4V	0.001V	
	40V	0.01V	
	400V	0.1V	
	750V	1V	±(0.8%rdg+5dgt)
DCA	400uA	0.1uA	±(0.8%rdg+2dgt)
	4000uA	1uA	
	40mA	0.01mA	
	400mA	0.1mA	
	10A	0.01A	±(1.0%rdg+2dgt)
ACA	400uA	0.1uA	±(1%rdg+5dgt)
	4000uA	1uA	
	40mA	0.01mA	
	400mA	0.1mA	
	10A	0.01A	±(1.2%rdg+5dgt)
Ohms	400Ω	0.1Ω	±(0.6%rdg+2dgt)
	4KΩ	1Ω	
	40KΩ	10Ω	
	400KΩ	100Ω	
	4MΩ	1KΩ	
	40MΩ	10KΩ	±(1.0%rdg+3dgt)
⊣⊢	4nF	1pF	±(1.9%rdg+20dgt)
	40nF	10pF	
	400nF	100pF	
	4uF	1nF	
	40uF	10nF	
	400uF	100nF	
	4mF	1uF	
	40mF	10uF	
Hz VA-Hz	4KHz	1Hz	±(0.05%rdg+1dgt) VA-Hz Max. to 1MHz
	40KHz	10Hz	
	400KHz	100Hz	
	4MHz	1KHz	
	40MHz	10KHz	
RPM	40KRPM	0.01KRPM	±(0.05%rdg+1dgt)
	400KRPM	0.1KRPM	
	4MRPM	1KRPM	
	40MRPM	10KRPM	
	400MRPM	100KRPM	
TEMP	-50°C~1000°C	1°C	1%rdg+2°C
	-58°F~1832°F	1°F	1%rdg+4°F
⊣⊢	⊣⊢		Test Current: 0.6mA
•)))	•)))		Active Sound Level: ≤40 Ω

LED back light system for operation under poor illumination.

Shock absorbing holster add an extra protection and provide a buit-in tilt stand.

Other colorchoices for Holster

' The specifications and other information in this catalogue are subject to change without prior notice.

CENTER®

CENTER TECHNOLOGY CORP.
4/F., No.415, Jung-Jeng Rd., 238 Shu-Lin Chien, Taipei, Taiwan
TEL: 886-2-26763926   E-Mail: center@centertek.com
FAX: 886-2-26763925   http : //www.centertek.com

Distributed by:

Figure A.1 - Multimeter Datasheet

SPECIFICATIONS

Display	LCD display, 4 digit LCD reading
Range	2000 W/m^2 634 Btu / (ft^2 x h)
Resolution	0.1 W/m^2 0.1 Btu / (ft^2 x h)
Spectral response	400-1100nm
Accuracy	Typically within ±10 W/m^2 [±3 Btu / (ft^2 x h)] or ±5%, whichever is greater in sunlight ; Additional temperature induced error ±0.38 W/m^2/℃ [±0.12 Btu / (ft^2 x h)/℃] from 25℃
Angular accuracy	Cosine corrected <5% for angles < 60°
Drift	< ±2% per year
Calibration	User recalibration available
Over-Input	Display shows "OL"
Sampling Rate	4 times/sec
Battery	4 pcs size AAA
Battery Life	Approx. 100 hours
Operating Temp and Humidity	0℃ to 50℃ (32℉ to 122℉) below 80%RH
Storage Temp and Humidity	-10℃ to 60℃ below 70%RH
Weight	Approx. 158g
Dimension	110(L) × 64 (W) × 34(H) mm
Accessories Included	Operation Manual , 4 pcs size AAA

Figure A.2 – Solar meter datasheet

AC Current

Range	Resolution	Accuracy 50Hz~500Hz
10A	1mA	2%+10
80A	10mA	
80A~100A		3.5%+10

DC Current

Range	Resolution	Accuracy
10A	1mA	2.5%+10
80A	10mA	
80A~100A		4.5%+10

AC Voltage

Range	Resolution	Accuracy 50Hz~500Hz
600V	0.1V	1.5%+5

DC Voltage

Range	Resolution	Accuracy
600V	0.1V	1%+2

Resistance

Range	Resolution	Accuracy
10KΩ	1Ω	1%+3

Figure A.3 – Ammeter datasheet

Absolute accuracy at full scale, differential ‹		
Range	Typical at 25 °C (mV)	Maximum over Temperature (mV)
±20	14.7	138
±10	7.73	84.8
±5	4.28	58.4
±4	3.59	53.1
±2.5	2.56	45.1
±2	2.21	42.5
±1.25	1.70	38.9
±1	1.53	37.5

Figure A.4 - DAQ datasheet