BLIDA 1 UNIVERSITY

Faculty of technology
Electronics Department

DOCTORAL THESIS
In Electrical Engineering

FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING ARTIFICIAL INTELLIGENCE TOOLS

Defended by:

GAROUDJA Elyes

Before a jury composed of:

A. NAMANE	Professor, Blida 1 University	Chairman
A. FERDJOUNI	MC-A, Blida 1 University	Examiner
A. KHALDOUN	MC-A, MB University of Boumerdes	Examiner
K. KARA	Professor, Blida 1 University	Supervisor
A. CHOUDER	MC-A, M'sila University	Co-supervisor

Blida, 2018

ABSTRACT

In this thesis, two efficient strategies of faults detection and diagnosis in Photovoltaic (PV) systems are developed. The first strategy uses the probabilistic neural networks (PNN) classifiers to detect and diagnose faults in the Direct Current (DC) side of Grid Connected Photovoltaic (GCPV) systems. The second strategy suggests the development of two statistical methods (such as the improved-ratio and control charts based methods) to detect and diagnose the faults. The improved ratio based method consists on the evaluation of three coefficients to detect and diagnose short-circuits and open-circuits faults. While, the control charts based method applies the exponentially weighted moving average (EWMA) and Shewhart charts to detect and diagnose the faults in GCPV systems. However, the developed strategies require the availability of a high-quality database that describes the system behavior for both healthy and faulty operations. To deal with this concern, a PSIM^{TM}/Matlab^{TM} co-simulation strategy is developed to elaborate a trusted simulation model. This model requires the use of the One Diode Model (ODM) electrical parameters. For this, an efficient strategy, based on the artificial bee colony (ABC) and the best-so-far ABC algorithms, are developed to identify the ODM parameters. Finally, the ODM identified parameters are used to elaborate an efficient strategy of maximum power point (MPP) estimation. The efficiency of the developed strategies is experimentally evaluated by using real measured data, collected from two actual GCPV systems. The first one is a 9.54 kWp PV system located at Algiers (Algeria), while the second one is a 0.9 kWp PV system, located at Jaen University (Spain).

Keywords: Grid connected, faults, probabilistic neural networks, statistical methods, improved-ratio, control charts, EWMA, Shewhart, ABC, best-so-far ABC, one diode model.

ملخص

في هذه الأطروحة، تم اقتراح استراتيجيتين فعالتين لكشف و تشخيص الأعطاب في الأنظمة الكهروفولطية. تعتمد الاستراتيجية الأولى على استخدام مصنفات الشبكات العصبية الاحتمالية (PNN) لكشف وتشخيص الأعطاب جهة التيار المستمر (ت.م) للأنظمة الكهروفولطية المتصلة بالشبكة. كما تقترح الاستراتيجية الثانية طريقتين إحصائيتين (طريقة النسبة المطورة وطريقة مخططات التحكم) لكشف الأعطاب وتشخيصها. تتمحور طريقة النسبة المطورة في فحص ثلاث معاملات لكشف وتشخيص أعطاب الدارة المقصورة و الدارة المفتوحة. من جهة أخرى، تتمحور طريقة مخططات التحكم في استعمال مخططي EWMA و Shewhart لكشف و تشخيص أعطاب الأنظمة الكهروفولطية المتصلة بالشبكة. تستلزم الاستراتيجيات المقترحة توفر قاعدة بيانات ذات نوعية رفيعة لمحاكاة سلوك النظام الكهروضوئي في حالته السليمة و المرفوقة بالأعطاب. في هذا الصدد، تم اقتراح استراتيجية تعتمد على برنامجي PSIM^{TM}/Matlab^{TM} لتطوير نموذج محاكاة فعال. يستوجب هذا النموذج معرفة قيم معلمات النموذج الأحادي الصمام. من أجل استخلاص قيم هذه المعلمات، تم اقتراح استراتيجية فعالة تستند على خوارزميتي ABC و best-so-far ABC. ختاما، تستعمل القيم المستخلصة للمعلمات في تطوير استراتيجية فعالة لتقدير نقطة القوة القصوى. تم تقييم أداء الاستراتيجيات المقترحة بصفة تجريبية باستخدام قياسات تم جمعها من نظامين كهروفولطيين النظام الأول هو نظام طاقة كهروفولطية سعة 9.54 كيلوواط موجود في الجزائر العاصمة (الجزائر)، في حين أن النظام الثاني هو نظام طاقة كهروفولطية سعة 0.9 كيلوواط، موجود في جامعة Jaen (إسبانيا).

RÈSUMÈ

Dans cette thèse, deux stratégies efficaces de détection et de diagnostic des défauts dans les systèmes photovoltaïques (PV) sont développées. La première stratégie utilise les classificateurs à base des réseaux de neurones probabilistes (PNN) pour détecter et diagnostiquer les défauts du côté courant continue (CC) des systèmes photovoltaïques connectés au réseau (SPVCR). La deuxième stratégie suggère le développement de deux méthodes statistiques (telles que les méthodes basées sur les rapports améliorés et les cartes de contrôle) pour détecter et diagnostiquer les défauts. La méthode basée sur les rapports améliorés consiste à l'évaluation de trois coefficients pour détecter et diagnostiquer les défauts des court-circuits et des circuits-ouverts. Tandis que la méthode basée sur les cartes de contrôle applique les cartes EWMA et Shewhart pour détecter et diagnostiquer les défauts dans les SPVCR. Cependant, les stratégies développées nécessitent la disponibilité d'une base de données de très bonne qualité qui décrit le comportement du système pour les opérations saines et défectueuses. A cet effet, une stratégie de co-simulation PSIM^{TM}/Matlab^{TM} est développée afin d'élaborer un modèle fiable de simulation. Ce modèle nécessite l'utilisation des paramètres électriques du Modèle à Une Diode (MUD). Pour cela, une stratégie efficace, basée sur les algorithmes des colonies d'abeilles artificielles (ABC) et best-so-far ABC, est développée afin d'identifier les paramètres du MUD. Finalement, les paramètres identifiés de ce modèle sont utilisés pour élaborer une stratégie efficace d'estimation du point de puissance maximale (PPM). L'efficacité des stratégies développées est évaluée expérimentalement en utilisant des mesures réelles, collectées à partir de deux SPVCR. Le premier est un système photovoltaïque de 9,54 kWp situé à Alger (Algérie), tandis que le second est un système photovoltaïque de 0,9 kWp, situé à l'Université de Jaen (Espagne).

Mots clés: Connectés au réseau, défauts, réseaux de neurones probabilistes, méthodes statistiques, rapports améliorés, cartes de contrôle, EWMA, Shewhart, ABC, best-so-far ABC, modèle à une diode.

DEDICATION

To the best family in the world: my father *Toufik*, my mother *Faiza*, my brother *Salim*, my little sister *Asmaa* and my fiancée *Hanane*.

Without you, this work would never been elaborated.
Proud of you!

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

CHAPTER 3: FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING PROBABILISTIC NEURAL NETWORKS

CHAPTER 4: FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING STATISTICAL APPROACHES.

LIST OF ILLUSTRATIONS, GRAPHICS AND TABLES

ILLUSTRATIONS and GRAPHICS

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

INTRODUCTION

Context:

The need of sustainable energy solutions in the worldwide is a demand nowadays, due to the high electricity consumption in combination with the desired environmental friendly solutions for power production development. Photovoltaic (PV) energy, which has gained a central place in governments' energy policies, as it is more suitable for grid connection and stand-alone schemes, is a promising solution to address the aforementioned concerns [1, 2]. In addition, its eco-friendly nature and abundance have given additional advantages that motivated its worldwide deployment.

Problem statement:

Although the advanced tools for the PV power generation, in practice, several factors can affect significantly the PV system performance by decreasing its efficiency. Indeed, PV systems are frequently exposed to different sources of faults and anomalies that affect the power generated by the PV generators. These faults could considerably reduce the production efficiency and the lifespan of PV arrays [3-7]. This is mainly due to external interferences or faults resulting from dust accumulation on the PV modules, aging of PV modules, shading, MPPT error, and inverters faults.

PV system, especially its DC side, may be subjected to defects and anomalies causing a drop of the overall system performance or even to their total unavailability [8]. Therefore, a real time fault detection and diagnosis procedure is crucial, not only for lowering maintenance cost, but also to avoid any energy loss, damage to equipment and safety hazards.

The increased attention given to fault detection and safety in PV systems has led to the development of several methods of fault detection and diagnosis. These methods are classified into three main types: Process-History Based methods, Quantitative-Model Based methods, and Signal-Processing Based methods.

Process History based methods rely mainly on machine learning and computational intelligence methods. These methods depend on the availability of quality input data, and their implementation is not an easy task, especially for real-time applications. Moreover, multi-layer perceptron neural networks suffer from their slow training step, they do not support noisy data, and could fall in local minima instead of global one [9, 10]. These weaknesses could obviously affect the reliability, the efficiency and even the real time utilization of these methods.

Moreover, statistical monitoring charts [11-13] are important tools for monitoring sequential systems to make sure that they work stably and satisfactory. However, until recently statistical control charts have not been widely used to improve the performance of PV systems. In addition, the main shortcoming of the proposed monitoring chart-based methods is their limitation to detect incipient faults because they make decisions based only on the recent observations [11].

Quantitative-Model based methods compare analytically the real measured outputs with the model-based outputs (simulated outputs) to detect and diagnose faults in PV systems [14]. This category of methods requires the use of an accurate simulation model that mimics the system operation under healthy and faulty conditions. Also, it suffers from the manual strategy of threshold computation and it requires that the PV array works at the maximum power point (MPP).

Signal-Processing Based methods rely on applying advanced techniques of signal processing to detect and diagnose faults in PV systems. This category of methods is an expensive strategy because it requires the use of additional expensive devices (such as LCR meter, thermal camera… etc.) of signal processing to detect and diagnose the faults [15, 16].

Objectives:

In the light of the above discussion, two novel strategies of faults detection and diagnosis in GCPV systems, based on neuronal classifiers and statistical methods, will be suggested in this thesis.

The first strategy consists on using probabilistic neural network (PNN) [9] classifiers to detect and diagnose faults in the DC side of GCPV systems. This choice is essentially due to the PNN convergence speed, its simplicity and

its no-need to weights adaptations. Moreover, PNN classifier has a strong robustness against noisy data generated by the measurement equipment and involved sensors [9, 10].

The second strategy suggests the development of two statistical methods (such as the improved-ratio [17] and control charts [14] based methods) to detect and diagnose faults in GCPV systems.

The improved ratio based method consists on the evaluation of three coefficients: currents coefficient, voltage coefficient and power coefficient, to detect and diagnose short-circuits and open-circuits faults. While, the control charts based method will apply two control charts, EWMA [11, 12] and Shewhart [11] charts, to detect and diagnose the DC side faults.

The efficiency of the developed strategies will be experimentally assessed by using real measurements collected from Algerian and Spanish PV systems.

Thesis organization:

The thesis is divided into four chapters in which the last three chapters denote original works that have already been published.

 In the **first chapter**, we will discuss the subject through a general study of PV systems, as well as their different types of faults. Then, a literature review on the already proposed methods of faults detection and diagnosis will be conducted.

 The second chapter will be entirely dedicated to PV array modeling. Indeed, the ODM of the PV module will be firstly introduced. Then, the five electrical parameters of this model will be identified by two heuristics optimization algorithms (ABC [18, 19] and best-so-far ABC [20, 21]). Lastly, a new strategy for estimating the maximum power will be suggested.

 The third chapter will be devoted to the application of PNN classifiers for faults detection and diagnosis. In this chapter, the developed strategy is based essentially on (i) the parameters identification results, (ii) the elaboration of a PSIM^{TM}/Matlab^{TM} co-simulation model for the real PV system, and (iii) the development of two PNN classifiers for faults detection and diagnosis.

The fourth chapter will be dedicated to the development of two statistical methods (such as the improved ratio and control chart based methods) for the supervision, faults detection and diagnosis of PV systems.

Finally, some conclusions and future perspectives will be depicted.

CHAPTER 1

STATE OF THE ART

1.1. Introduction:

Until recently, energy consumption in the world still grows up very obviously. For this reason, one of the best efficient strategies, recently applied in the developed countries, is to use renewable energy sources to satisfy their energetic need [22, 23].

Among the different available renewable energy sources, PV energy is one of the best alternative solutions [24]. Indeed, the sun is the most important source of energy in our planet. Furthermore, PV sources are renewable, sustainable and of eco-friendly nature. PV energy sources reduce significantly the pollution effect on the environment, contrary to the traditional ones, (such as: oil, coal and nuclear), which contaminate our atmosphere by releasing several venomous gazes (carbon dioxide, sulphur dioxide, nitrous oxide and mercury).

In the other side, during its operation, PV systems are usually exposed to different sources of failures and anomalies, which may noticeably reduce the system energetic efficiency, degrade its performance and even decrease the PV arrays lifetime [6]. Thus, efficient strategies to detect and diagnose PV system failures at an earlier stage must be developed.

In this chapter, PV systems and their main components are briefly described. Then, the different types of PV system failures and their main causes are listed. Finally, a state of the art for the recently proposed strategies of faults detection and diagnosis is presented.

1.2. PV systems:

PV systems are commonly divided into two main classes: Grid-Connected PV (GCPV) systems and Stand-Alone PV systems. GCPV systems (also known as grid-tied), are always interfaced to the local electricity grid, and they are mainly used to solve the energy demand issue. In the other side, stand-alone PV systems are self-contained, and they are mainly constructed to satisfy the energy need of isolated customers. During the present research work, only GCPV systems

are considered. Figure (1.1) shows the main components of a GCPV system, while its electrical synoptic is depicted in Figure (1.2) [25].
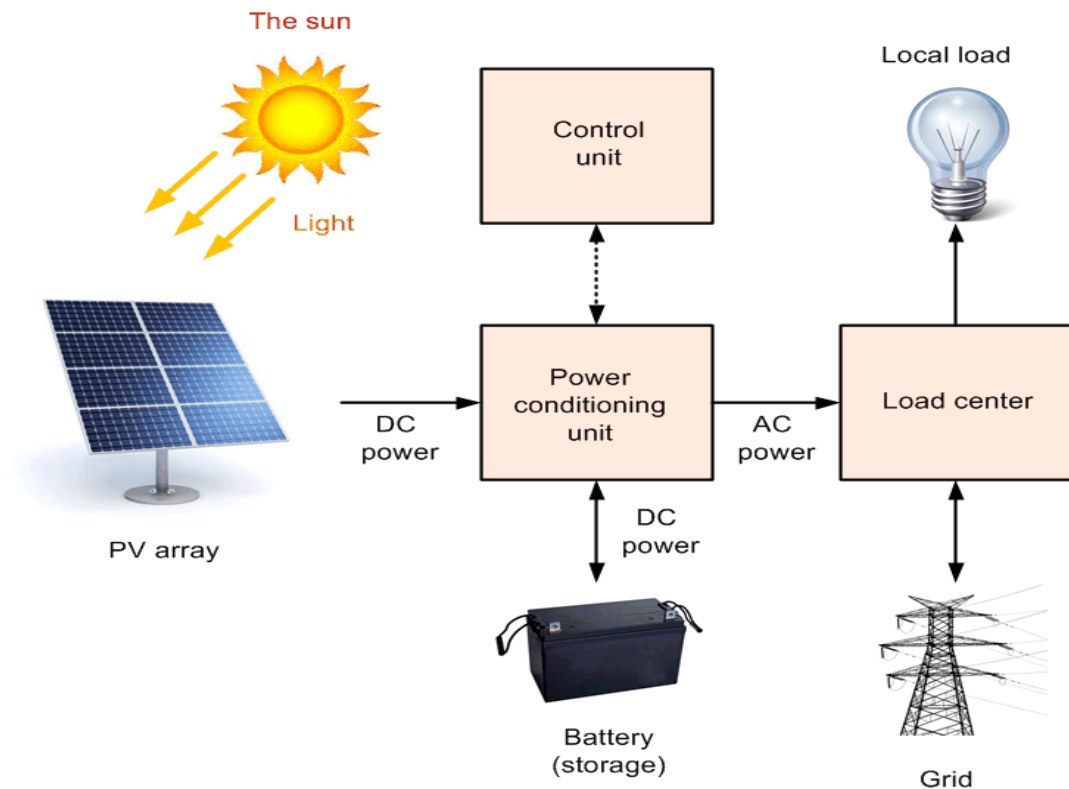
Figure (1.1): The main components of a grid connected PV system.

Whatever their types, PV systems are mainly composed of PV generators, converters, cables, junction boxes and protection devices.

1.2.1. PV generator:

The PV generator is the essential unit of energy production. This unit converts the solar energy into electrical energy through the photovoltaic effect. In other term, the sunlight beams will be converted to electricity when they will be exposed to some materials that exhibit the photovoltaic effect property.

1.2.1.1. PV cell:

The elementary component of PV generator is the PV cell. Generally speaking, the generated voltage from commercial solar cell is about 0.6V, while its produced current is mainly depending on the sunlight intensity and the solar cell area [25].
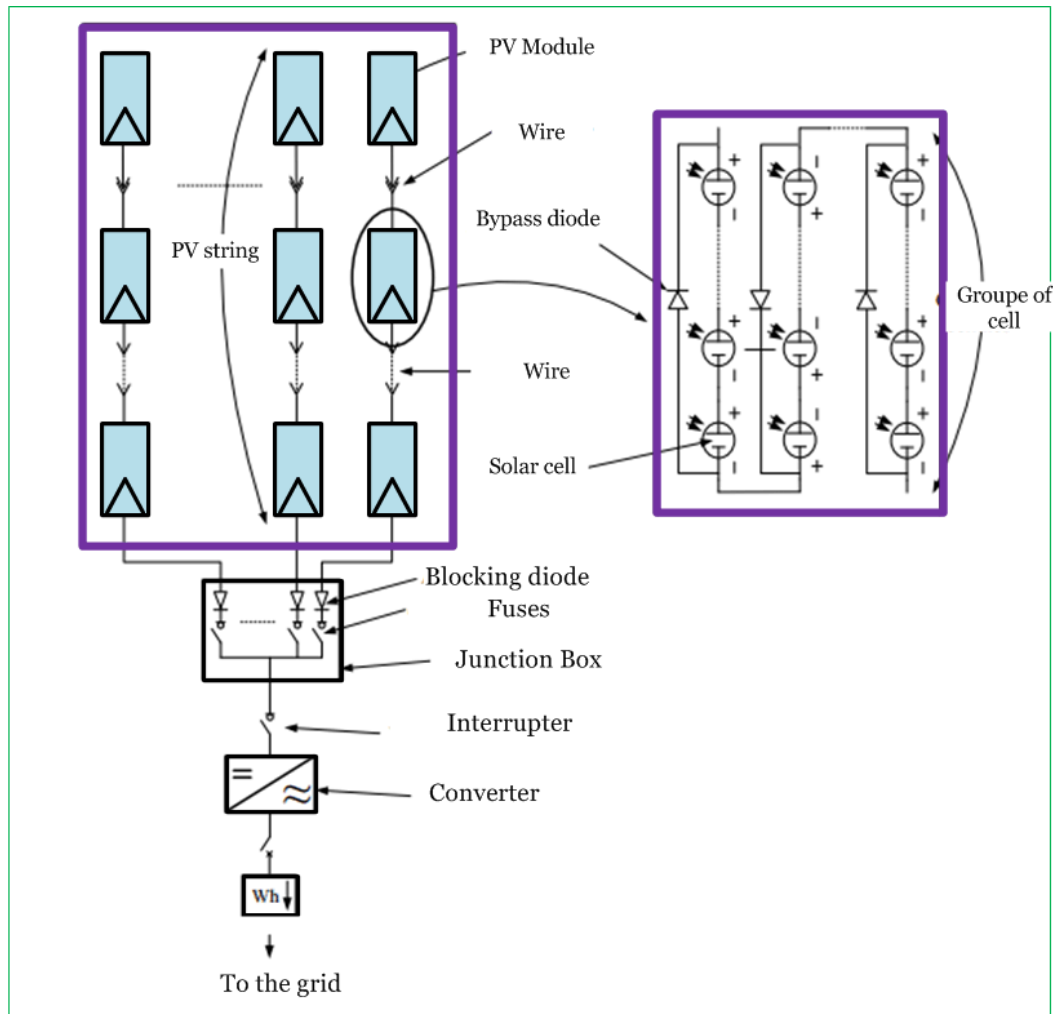
Figure 1.2: The electrical synoptic of a grid connected PV system [25].

1.2.1.2. PV module:

The amount of generated current and voltage, i.e. produced power, of one solar cell is not enough for real applications. Hence, the cells are connected in series to increase the generated voltage and in parallel to increase the generated current.

In fact, series and parallel interconnection of several solar cells forms a PV module. The commercialized PV modules available nowadays are constructed of 36, 40, 54, 60, 72, or even 108 solar cells in series. These cells are sub-grouped into several groups. Each one is connected in parallel with one bypass diode. The main goal of this diode is to protect the group of cell from the reversed polarization.

1.2.1.3. PV string:

A PV string is the connection of several PV modules in series in order to reach the required output voltage.

1.2.1.4. PV array:

In order to produce a required power under a tolerable voltage, PV modules should be connected in series-parallel to form a PV array.

1.2.2. Converters:

PV systems are mainly composed of two types of converters, DC/DC and DC/AC converters. The main goal of these converters is to extract the PV array maximum power (DC power), and convert it into alternative power (AC power), before the grid utility injection.

1.2.2.1. DC/DC converters:

When connecting a PV generator (PVG) to a load R, the operating point (current and voltage) is defined as the intersection of the (I-V) characteristic curve of the PVG and the (I-V) characteristic curve of the load R, as depicted in Figure (1.3). This operating point depends on R, and the slope of the load characteristic at this operating point is 1/R. However, there is only one optimal operating point at which the PV generator produces the maximum power (P_{mpp}), and its corresponding coordinates are noted as I_{mpp} and V_{mpp}. Thus, adaptation blocks are required for the maximum power extraction.

These blocks of adaptation consist of DC/DC converters, which are equipped with a maximum power point tracking (MPPT) algorithms. The main goal of these algorithms is to operate the PV generator under its maximum power point.

1.2.2.2. DC/AC converters:

The DC/AC converters (inverters) are crucial blocks for the GCPV systems. In fact, these blocks guaranty the conversion of the PV generator energy from DC to AC. The DC/AC conversion is necessary to reach the grid utility injection in one side, and to supply the AC electrical devices in the other side.
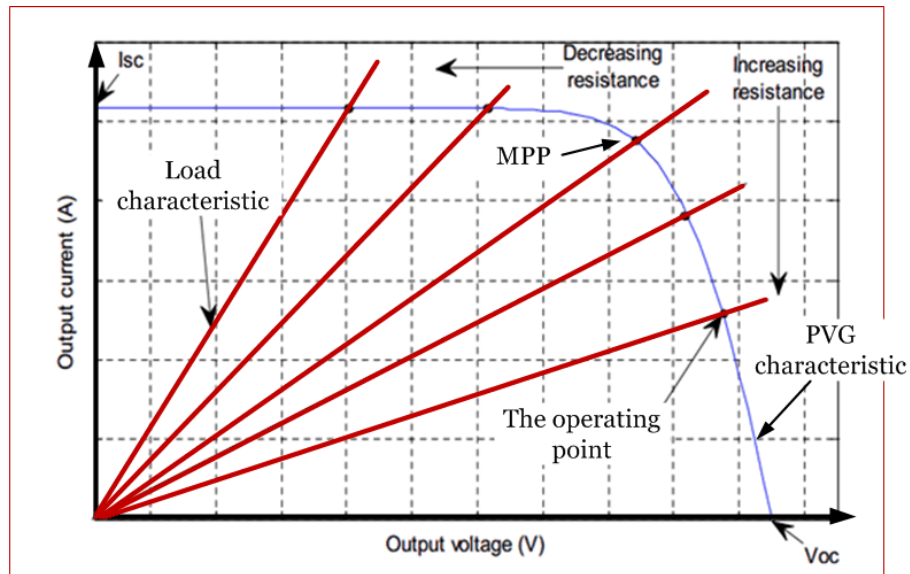
Figure 1.3: The operating point variations for different values of resistive load.

1.2.2.3: <u>Converters topologies:</u>

GCPV systems consist of two topologies [26], central topologies (Figure (1.4)) and modular topologies (Figure (1.5)).
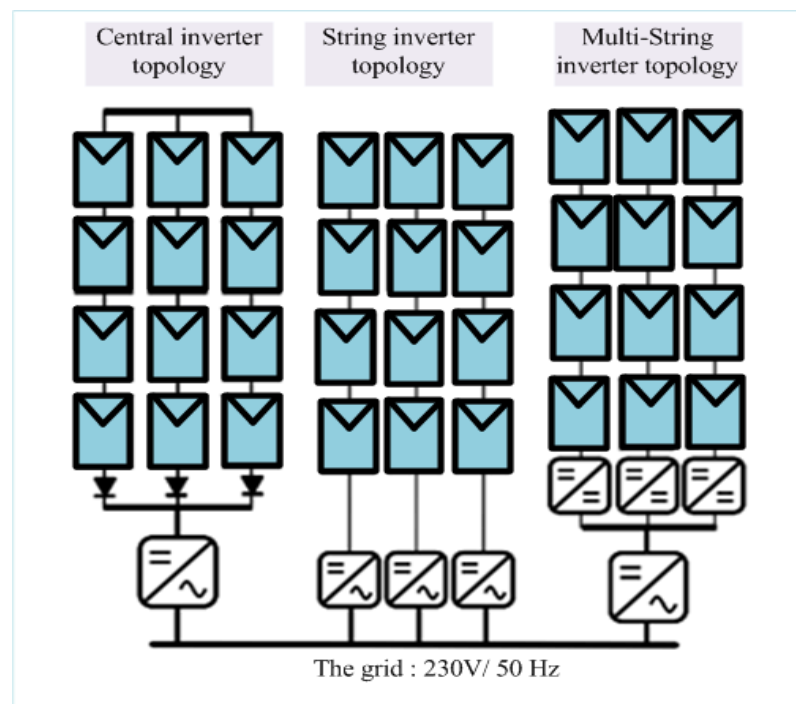
Figure 1.4: The central topologies of a GCPV system [25].
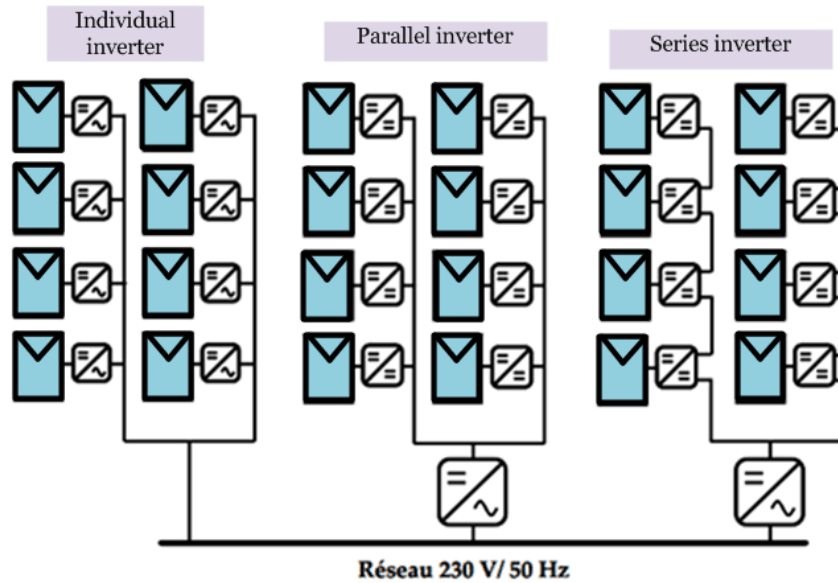
Figure 1.5: The modular topologies of a GCPV system [25].

GCPV system has three main central topologies: central inverter, string inverter and multi-string inverter.

- Central inverter: This type of topology is the frequently used topology nowadays. It consists of using one DC/DC converter and one DC/AC converter to interface the PV generator with the grid utility [25]. However, using only one DC/DC converter, for the entire PV array, does not allow an accurately extraction of the optimal maximum power point under partial shading conditions.

- String inverter: In this topology, one DC/DC converter and one DC/AC converter are used for each PV string. This type of topology enhances the MPP extraction procedure by increasing the number of MPP extraction blocks (number of DC/DC converters). Moreover, this type of topology guaranties the energy availability in case of failure of one inverter for instance [25].

- Multi-string inverter: In this last central topology, each PV string is equipped with a single DC/DC converter, while only one DC/AC converter is associated to the entire PV array [25]. The main advantage of this topology is the low cost comparing with the string inverter one. However, it cannot guaranty the electricity availability in case of inverter failure.

In the other side, the modular topologies are also vastly used nowadays. The main goal of these topologies is to reduce the effect of one faulty module

on the entire PV array operation. In other term, these topologies enhance the maximum power point extraction procedure even in the presence of faulty PV modules.

GCPV systems have three types of modular topologies: individual inverter, parallel inverter and series inverter [26].

- Individual inverter: This first topology allows PV modules to supply directly the grid. In other term, for each PV module, we associate individual DC/DC and DC/AC converters.

- Parallel inverter: In this topology, each PV module is connected with a separate DC/DC converter, while the whole DC/DC converters are connected with only one inverter.

- Series inverter: This topology is similar to the parallel inverter topology, except that in this particular case, the DC/DC converters are connected in series.

1.2.3. Cables:

Grid connected PV system requires the use of double isolated cables, which are water resistant, to link the PV generator with the inverter. These cables should be carefully sized to provide a reduced voltage drop [27].

1.2.4. Junction box:

The parallel connection of several PV strings is achieved by using a junction box. Moreover, this block can comprise several safety devices such as: fuses and interrupters [25].

1.2.5. Protection devices:

GCPV systems usually comprise several protection components, such as bypass diodes and blocking diodes:

1.2.5.1. Bypass diode:

The bypass diode is usually used to protect solar cells from the reverse polarization. In fact, when one of the serially connected solar cells/modules of the PV generator receives less amount of photo-current (Iph) than the others, due to the partial shading occurrence, it will become reversed biased. In other word, this solar cell/module will dissipate energy instead of generating it, which results in

the hotspot phenomenon [28]. Under these conditions, the cell temperature will rise substantially till reaching a high level at which the PV cell/module will be damaged. This phenomenon could considerably affect the entire PV module/array safety. To avoid this problem, the most frequently used strategy is to add a bypass diode in parallel with a preset group of serially connected PV cells. This diode is reverse biased in the case of uniform insolation, i.e. it has no effect in this case. However, it will be forward biased when the solar cells are partially shaded. In fact, under the partial shading condition, the current will pass via this diode instead of the shaded solar cells/modules.

1.2.5.2. Blocking diode:

For a multi-strings PV system, the generated voltage of some strings could be different from the others. Under this circumstance, the PV string with lower voltage can dissipate a reversed current, originally generated from the other strings. A similar situation could lead to energy reduction and a PV string damage [25]. The most commonly used strategy to protect the system from the reversed current, is to connect a blocking diode in series of each PV string.

1.3. Faults in PV systems:

PV systems are usually exposed to several sorts of failures, which could affect significantly its performance by decreasing its efficiency [6].

The flowchart of Figure (1.6) shows the different types of faults usually occurred in PV systems. As shown in this figure, faults in PV systems consist mainly of two classes [29]: faults in the direct current (DC) side, and faults in the alternative current (AC) side.

Hereafter a detailed description of the most commonly occurred faults in PV systems is given:
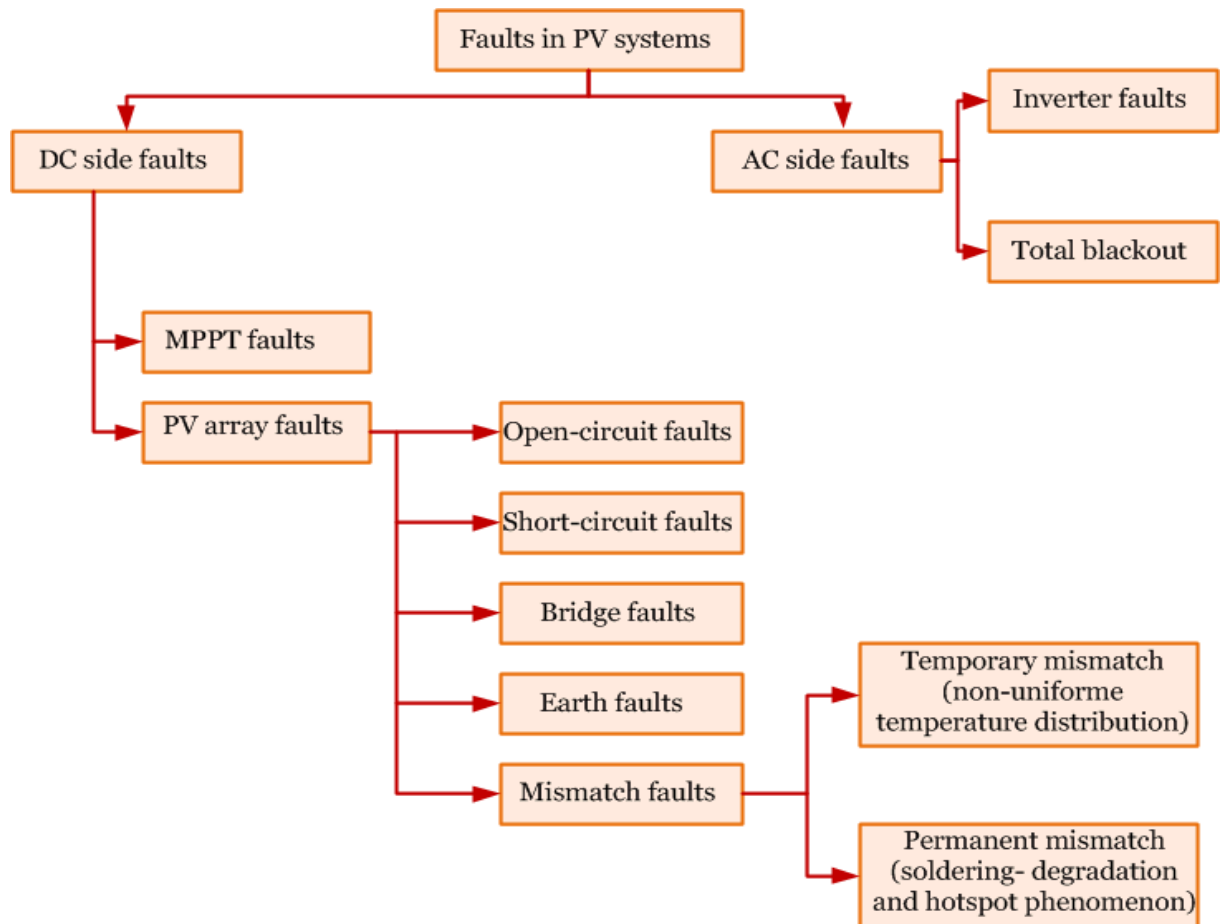
Figure 1.6: Types of faults usually occurred in PV systems.

1.3.1. DC side faults:

Generally, faults of PV arrays and MPPT controllers are the main DC side faults, usually occurred in PV systems [29]. The detection of this first category of faults is often a mandatory stage to avoid energy loss, system shutdown and even disastrous fires. Details about DC side faults, their main causes and impacts are clarified in the following subsections.

1.3.1.1. PV array faults:

During its operation, PV array may be exposed to five crucial types of faults [29]: mismatch faults, open-circuits faults, earth faults, bridge faults (Line-Line fault) and short-circuits faults.

a. Mismatch faults:

Mismatch fault can be defined as the connection of several solar cells or PV modules, which have different electrical parameters, or which are exposed to different operating conditions (temperature and irradiance) [29].

This type of fault has been considered as a serious problem because of its hard detection and enormous effect on power reduction. Mismatch faults are categorized into two main groups: temporary mismatch faults and permanents mismatch faults [29].

- Temporary mismatch faults:

These faults rely on the non-uniform temperature distribution (covering of snow) [29]. In fact, during winter season, the covered snow on the top of PV modules may cause a non-uniform temperature distribution. This effect can reduce considerably the PV systems energetic production.

- Permanents mismatch faults:

This subcategory of mismatch faults comprises mainly three faults [29]: degradation fault, soldering fault and hotspot phenomenon.

Degradation fault:

PV array degradation is due essentially to the ageing of PV cells. In fact, this type of fault occurs when the series resistance between PV cells increased over the time, due to the reduced adherence of contacts, or corrosion caused by water vapor [29, 30].

Soldering fault:

This type of fault occurs when the solder bond between solar cells and their contacted ribbons are accidently disconnected [29].

Hotspot phenomenon:

As previously explained (subsection 1.2.5.1), the hotspot phenomenon occurs when one of the series connected solar cells receives less amount of photo-current (Iph) than the others, due to the partial shading occurrence. During this stage, this cell becomes reversed biased and it dissipates power instead of generating it. A similar situation leads to a significant temperature rise of this cell, and even to its damage.

b. Open-circuit fault:

This type of fault occurs when the series connection between the current-carrying path and the load is accidently broken, opened or damaged [29].

A similar situation may occur due to the poor connection between solar cells, and to the break of cables between PV modules [29].

c. Earth fault:

This type of fault has been considered as the most commonly happened PV arrays faults. Earth fault can be reached when an unintentional connection to the ground is accidentally established [29].

d. Bridge fault:

Bridge fault (called also Line-Line fault) occurs when a connection of low-resistance is established between two different points of dissimilar potential within different PV strings [29]. The main sources of this type of faults are: corrosion, mechanical damages and failure of cables' insulation [29].

e. Short-circuit fault:

This fault affects mainly solar cells, bypass diodes or PV modules. It is mainly due to the water infiltration into the modules, and to the bad wiring between modules and inverter [29]. This type of fault can be generated according to an accidental connection between two points of different voltages within one string or PV module.

1.3.1.2. MPPT faults:

The aim of MPPT controller (DC/DC converters) is to extract the MPP (I_{mpp}, V_{mpp}) of the PV array. However, during the real operation of PV systems, inefficient tracking of MPP could occur, especially under partial shading conditions [29].

1.3.2. AC side faults:

Two main types of faults could happen in the AC side of PV systems: inverter faults and total blackout [29].

1.3.2.1. Inverter faults:

The failure of each part of the inverter (transformer, capacitors, IGBTs, and drive circuitry) leads to the inverter faults [29].

1.3.2.2. <u>Total blackout:</u>

This type of fault can be defined as a permanent power loss, mainly due to the occurrence of an accident or natural disaster [29]. Indeed, under these circumstances, the inverter will shut down immediately the power supply from the PV generator to the distributed grid.

1.4. <u>Faults detection and diagnosis methods in the literature:</u>

Up to now, methods of faults detection and diagnosis in PV systems are ranked into three main types: Process-History Based methods, Quantitative-Model Based methods, and Signal-Processing Based methods. These methods employ several measurements to detect and diagnose failures, such as: meteorological measurements (Temperature (T) and Irradiance (G)), the maximum power point coordinates of currents I_{mpp} and voltage V_{mpp}, inverters outputs of current I_{ac} and voltage V_{ac}, and finally the current-voltage (I-V) characteristic.

Several criteria have been used to assess the efficiency of a such method, some of them are listed below:

- The capacity to detect the occurrence of faults at an earlier stage.
- The aptitude to diagnose correctly the type of detected faults.
- The robustness of the developed method against modeling errors and noisy conditions.
- The quality and the amount of the required data to be used during fault detection and diagnosis.

1.4.1. <u>Process History-based methods:</u>

In these methods, implicit empirical models, mainly derived from the available data analysis, are used to detect and diagnose PV system faults. Process History-based methods rely mainly on machine learning and computational intelligence methods. Hereafter, a number of the most recently proposed works in this category are given.

An Artificial neural network based method to detect partial shading occurrence has been proposed in [31]. It uses an ANN type Multi-Layer Perceptron (MLP) to estimate the output PV current and voltage based on solar irradiance and cell temperature. Detection and diagnosis of the partial shading has been achieved

by comparing the measured and the estimated PV current and voltage. The network inputs are solar irradiance and cell temperature, while its outputs are the PV current and voltage.

The detailed flowchart of this method is depicted in figure (1.7) [31], where Ev and Ei are residual errors, while Si and Sv are their threshold values, respectively.
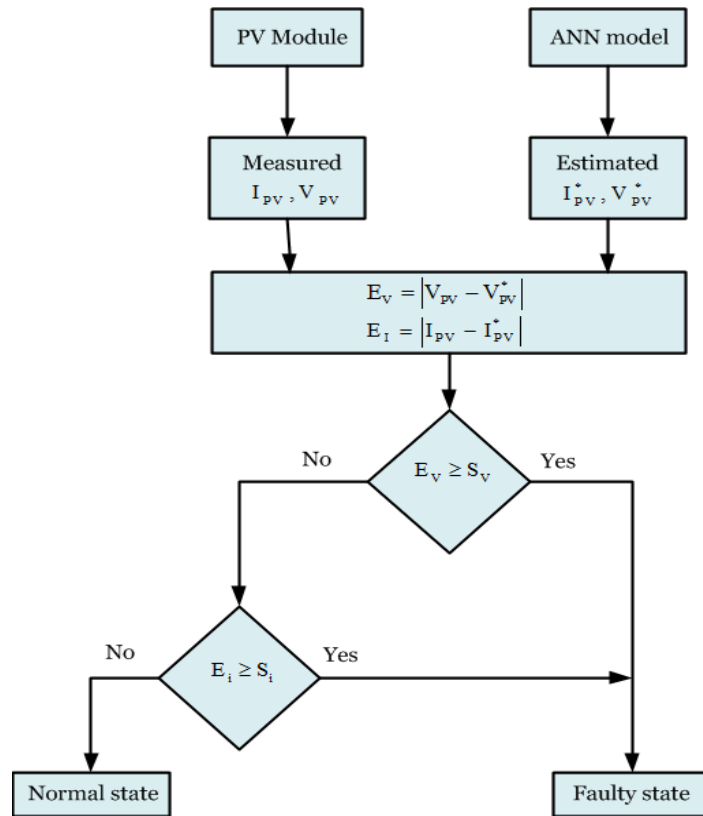
Figure 1.7: Flowchart of the ANN-based strategy for partial shading detection [31].

In addition, a hybrid Neuro-Fuzzy approach of fault detection and diagnosis has been proposed in [32]. This work consists mainly on (i) the development of a Neuro-Fuzzy model of PV modules ,(ii) the analysis and extraction of six attributes (Isc, Voc, Impp, Vmpp, S1 and S2) using the I-V characteristic, and finally (iii) the application of Norm-test to detect and diagnose faults.

S1 and S2 are the incremental derivative ratios, and they are computed as follows:

$$S_1 = \frac{I_{Mpp} - I_{sc}}{V_{Mpp} - V_{oc}} \qquad (1.1)$$

$$S_2 = \frac{0 - I_{Mpp}}{V_{oc} - V_{Mpp}} \qquad (1.2)$$

In this work, three Neuro-Fuzzy models (blocks) have been developed to model the system under ideal, normal and faulty conditions. The ideal condition block defines the case when all the inputs and outputs are considered perfectly constants and noiseless. However, normal and faulty blocks reflect the real behavior that can be expected in healthy and faulty practical operating cases.

The general flowchart of this strategy is depicted in Figure (1.8) [32].
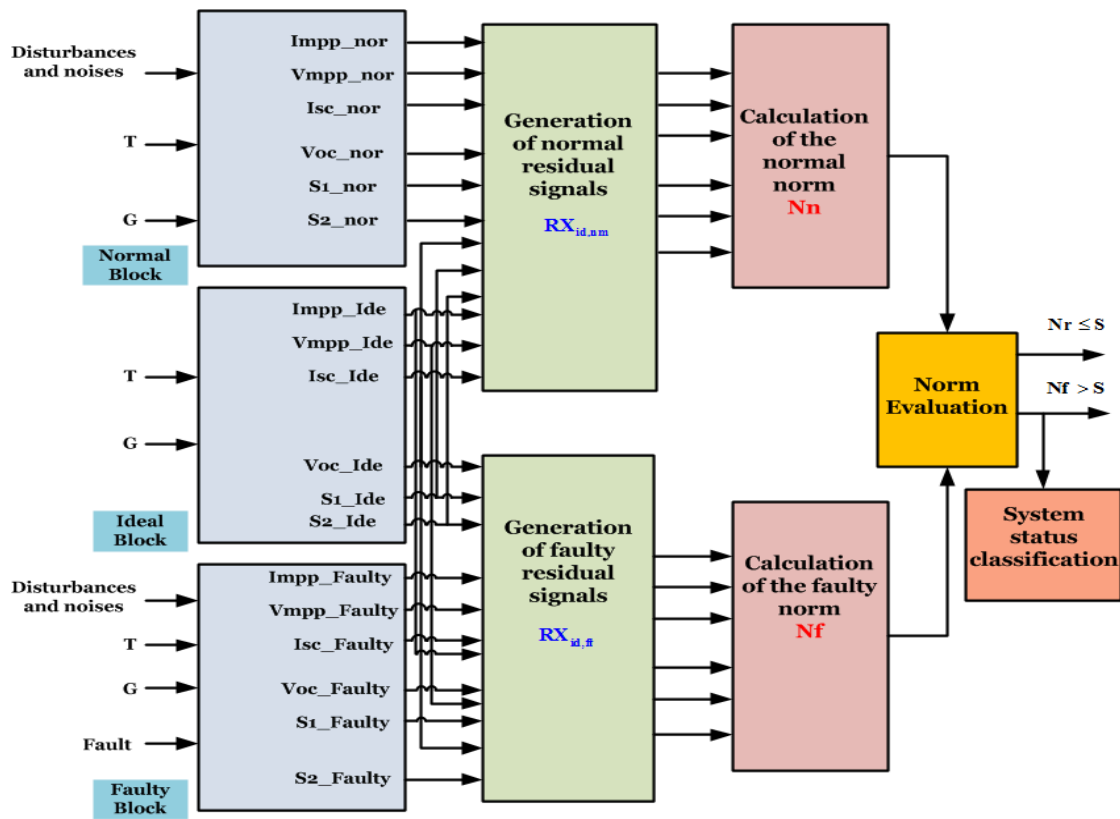
Figure 1.8: Flowchart of the Neuro-Fuzzy based strategy of fault detection and diagnosis [32].

Normal and faulty residuals signals, $RX_{id,nm}$ and $RX_{id,ft}$, are computed as follows [32]:

$$RX_{id,nm} = X_{id} - X_{nm} \qquad (1.3)$$

$$RX_{id,ft} = X_{id} - X_{ft} \qquad (1.4)$$

Where: X is the adequate attribute; the indexes (id, nm, ft) denote respectively ideal, normal and faulty status.

Normal and faulty norms, Nn and Nf, are computed as follows [32]:

$$\left|Nn\right| = \sqrt{\sum_{i=1}^{6}\left(RX_{id,nm}\right)^2} \qquad (1.5)$$

$$\left|Nf\right| = \sqrt{\sum_{i=1}^{6}\left(RX_{id,ft}\right)^2} \qquad (1.6)$$

To finish with, norms evaluation consisted on comparing the obtained normal norms (Nn and Nf) with experimentally set threshold values (S). This comparison could identify five operating cases: normal operating case, diode short-circuit operating case, lower earth fault operating case, upper earth fault operating case and partial shading operating case. System status classification strategy based on the norm test is summarized in Figure (1.9) [32].
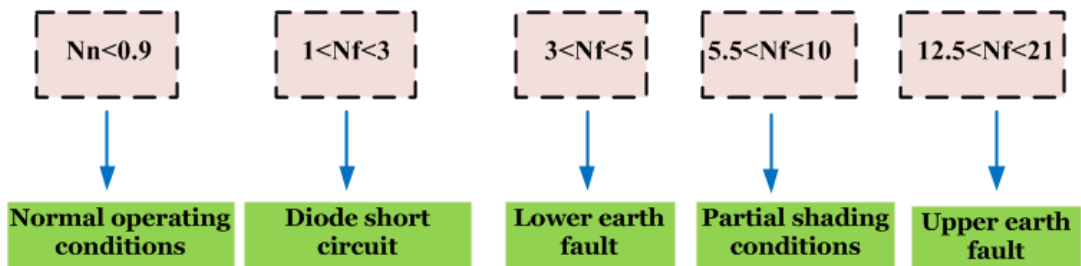
Figure 1.9: System status classification by the norm test [32].

Moreover, a machine learning based strategy has also been proposed in [33] as an efficient strategy of fault detection and classification. In fact, Zhao et al have suggested a Graph Based Semi-Supervised Learning (GBSSL) method to detect and diagnose PV system faults, using only few labeled training data [33]. Contrary to the traditional machine learning strategies, this method does not require a high amount of expensive labeled data, and its training strategy can automatically update the model under changing weather conditions. The proposed strategy used only readily available measurements of PV array voltage and current, PV module operating temperature and solar irradiance, to detect and diagnose faults.

In addition, this method focuses on two categories of faults: bridge faults and open-circuit faults. It is worth mentioning that the detection of the aforementioned faults is difficult task by using the conventional Over-Current-Protection-Devices (OCPD).

As summary, this strategy presents the following contributions [33]:

- A novel strategy of data normalization has been developed to enhance the detection and classification efficiency of machine learning methods, and to better improve data visualization.
- GBSSL model has been used for the first time in this field of application. The advantages raised from using this model are: the low training cost and the high ability of self-learning over the time.
- The developed strategy can be integrated into all PV inverter topologies and takes advantages of using only readily available measurements.

The normalized PV current I_{Norm} and voltage V_{Norm} are computed as follows [33]:

$$V_{Norm} = \frac{V_{mpp}}{N_{Mod} * V_{OC\_Ref}} \qquad (1.7)$$

$$I_{Norm} = \frac{I_{mpp}}{N_{Str} * I_{SC\_Ref}} \qquad (1.8)$$

Where: N_{mod} and N_{Str} define the number of modules and strings respectively; V_{oc_ref} and I_{sc_ref} are the reference open circuit voltage and short circuit current, respectively.

The overview of the GBSSL based strategy of fault detection and diagnosis is summarized in five main steps, as can be depicted in the flowchart of Figure (1.10) [33].

X is a labeled data matrix; Y is a matrix that includes the corresponding classes of these labeled data; Z is a binary matrix; and W is the weight matrix. More details about Z and W matrixes can be founded in [33].

The efficiency of this method has been tested on a small-scale grid connected PV system. Results have clearly demonstrated the efficiency of GBSSL based strategy in fault detection and diagnosis of bridge fault and open-circuit fault.
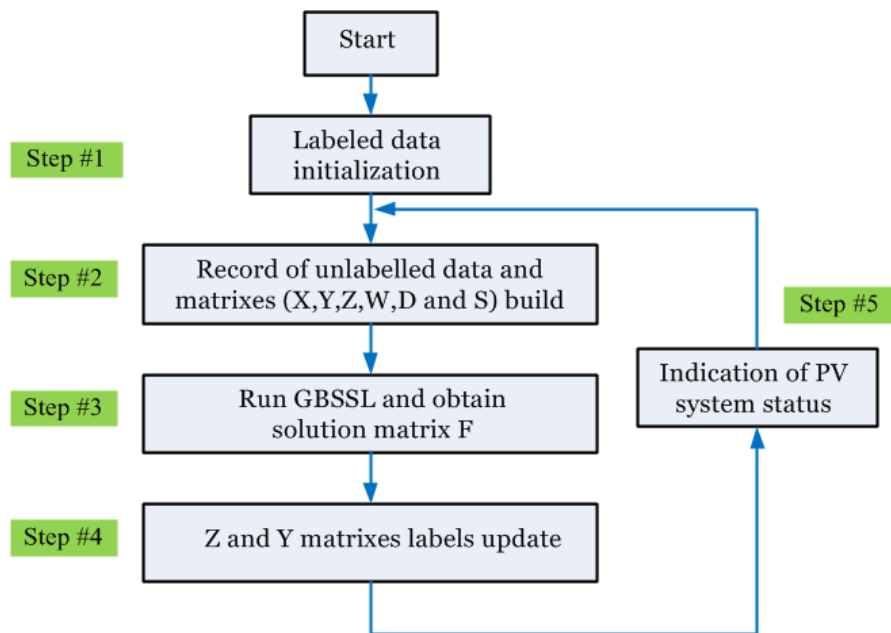
Figure 1.10: Flowchart of the GBSSL based strategy
of faults detection and diagnosis [33].

Chine et al have proposed an efficient strategy to detect, diagnose and localize faults in PV systems, using the ANN [34]. In this method, PV current, voltage and the number of peaks in the current-voltage (I-V) characteristic are computed based on a simulation model. The difference between the measured and simulated PV array output power is firstly computed and compared with the threshold value (S), to detect faults occurrence. Then, the analysis of the main attributes, derived from (I-V) characteristic of each separate PV string, is elaborated to diagnose and localize the faults.

The different steps of this method for the detection, diagnosis and localization are summarized in Figure (1.11) [34].
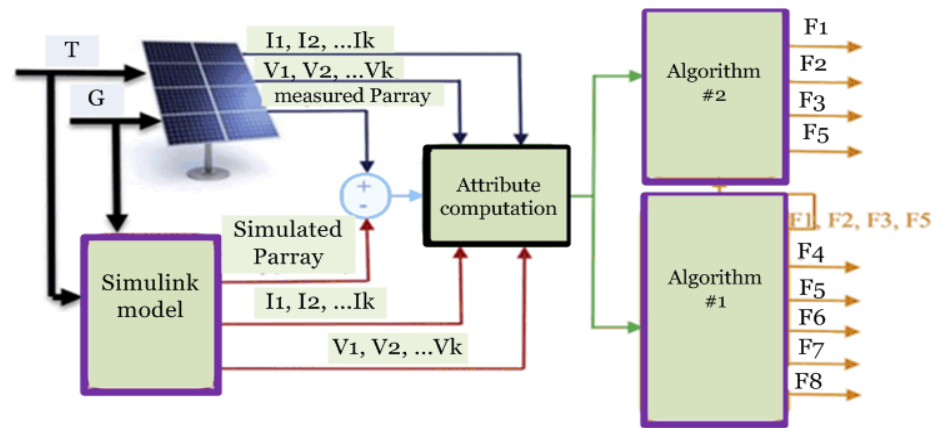
Figure 1.11: The different steps of the ANN-based method [34].

This method could accurately detect, diagnose and localize eight types of faults, which are described in table (1.1) [34].

Table 1.1: The possibly occurred faults that could be correctly detected by Chine's strategy [34]

Type of fault	Name	Symbol
	Short-circuit fault in any bypass diode or (cell or module)	F1
	Inversed bypass diode fault or (cell or module)	F2
Module	Shunted bypass diode fault of (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 string.	F6
Shadow effect with faulty bypass diode	Shadow effect in a group of cells equipped by a faulted bypass diode.	F7
Shadow effect with connection fault	Shadow effect in a group of modules connected by a connected resistance.	F8

As can be seen from Figure (1.11), this method employs two different algorithms, named algorithm #1 and algorithm #2.

The first algorithm, entitled signal threshold based approach, is described in detail in Figure (1.12) [34], and it allows the discrimination of six faulty cases. Parameters that are used in this first algorithm and their descriptions are summarized in table (1.2) [34].

In the other side, the second algorithm is devoted to distinguish between the faulty cases F1, F2, F3 and F5, by using two ANN classifiers (multi-layer perceptron (MLP) and radial basis function (RBF) classifiers).
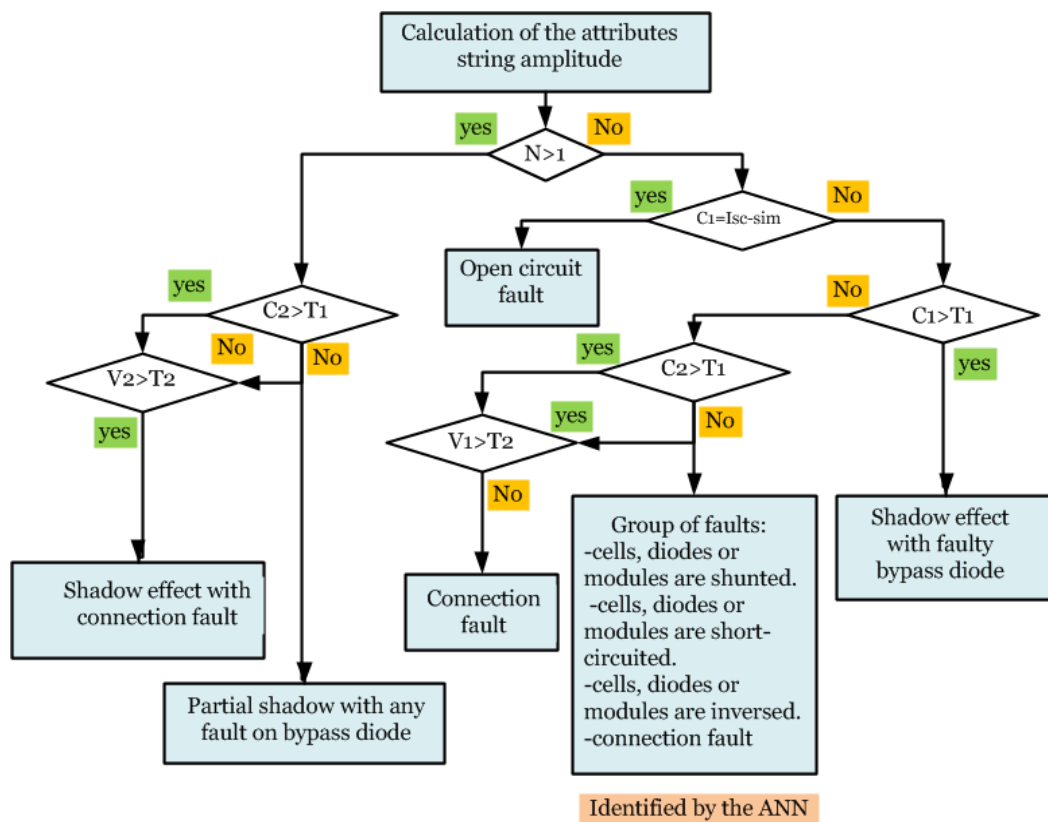
Figure 1.12: The detailed flowchart of the first algorithm [34].

Table 1.2: The description of parameters used in Algorithm #1

Variable	description
C1	a reduction in the short circuit current
V1	a reduction in the open circuit voltage
C2	a reduction or an increase in the output current
V2	a reduction or an increase in the output voltage
N	An increased number of MPPs in the I-V characteristic

Moreover, Dhimish et al proposed an efficient strategy of faults detection based on theoretical curve modeling and fuzzy classification system [35]. This method uses LabVIEW software to simulate the meteorological conditions (T and G), the voltage ratio VR, and the power ratio PR.

In this approach, a third order polynomial function has been employed to compute two detection limits (high and low detection limits) of VR and PR ratios. Then, these limits are compared with real measured data of an actual PV generator. Samples that lie out of these limits are then introduced to a fuzzy-logic-based classification system to diagnose the fault type. The general structure of this method is depicted in Figure (1.13) [35], while its detailed flowchart is depicted in Figure (1.14) [35].

The theoretical power ratio (PR) and voltage ratio (VR) are computed using Equations (1.9-1.10). These ratios are used to categorize the faults regions.

$$PR = \frac{P_{G,T}}{P_{G,T} - QP_0} \tag{1.9}$$

$$VR = \frac{V_{G,T}}{V_{G,T} - QV_0} \tag{1.10}$$

Where: $P_{G,T}$ and $V_{G,T}$ are respectively the theoretical power and voltage, generated from the PV generator, under specific meteorological conditions of temperature (T) and irradiance (G). Q is the number of PV modules. P_0 and V_0 denote respectively the maximum operating power and voltage, measured under standard test conditions (STC: T=25°C, G=1000W/m²).

The low limits of PR and VR are given by:

$$PR \text{ Low limit} = \frac{P_{G,T}}{\left(P_{G,T} - nP_0\right)\eta_{sensor}} \tag{1.11}$$

$$VR \text{ Low limit} = \frac{V_{G,T}}{\left(V_{G,T} - nV_0\right)\eta_{sensor1}} \tag{1.12}$$

Where $\eta_{sensor1}$ is the efficiency of voltage sensor, while η_{sensor} is the efficiency of both current and voltage sensors and it is expressed as follows:

$$\eta_{sensor} = \eta_{sensor1}(\text{voltage sensor efficiency}) + \eta_{sensor2}(\text{current sensor efficiency}) \quad (1.13)$$

The PR and VR high limits are those given by equations (1.9-1.10), and their real measured values are computed as follows:

$$\text{Measured PR} = \frac{P_{G,T}}{P_{measured}} \quad (1.14)$$

$$\text{Measured VR} = \frac{V_{G,T}}{V_{measured}} \quad (1.15)$$

This method could efficiently detect the following faults [35]:

- The presence of partial shading (PS) within the PV generator.
- One short-circuited PV module and PS.
- Two short-circuited PV modules and PS.
- (Q-1) short-circuited PV modules and PS, where Q denotes the total number of PV modules.

The main weakness of this method is that it depends on power and voltage ratios. Therefore, its efficiency depends highly of the efficiency of the instrumentation components.
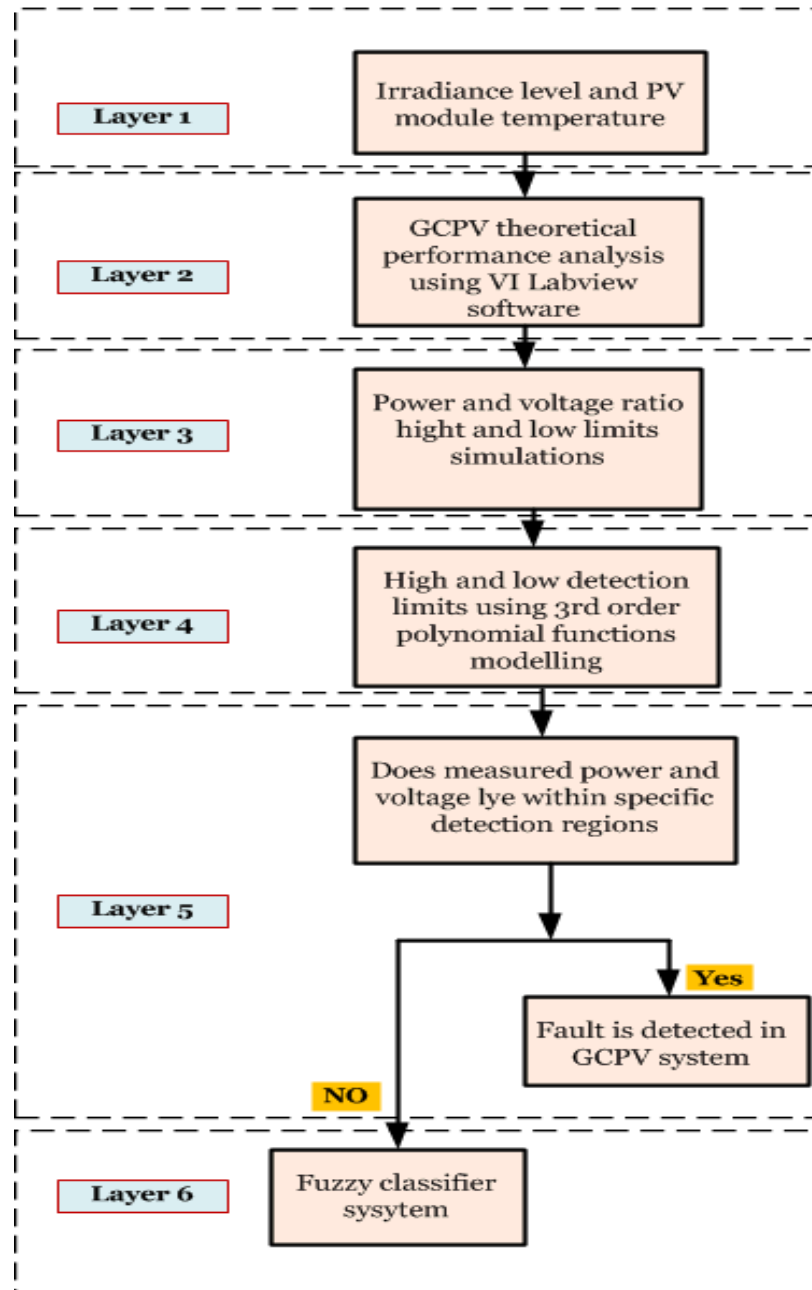
Figure 1.13: The general structure of the theoretical curve based method [35].
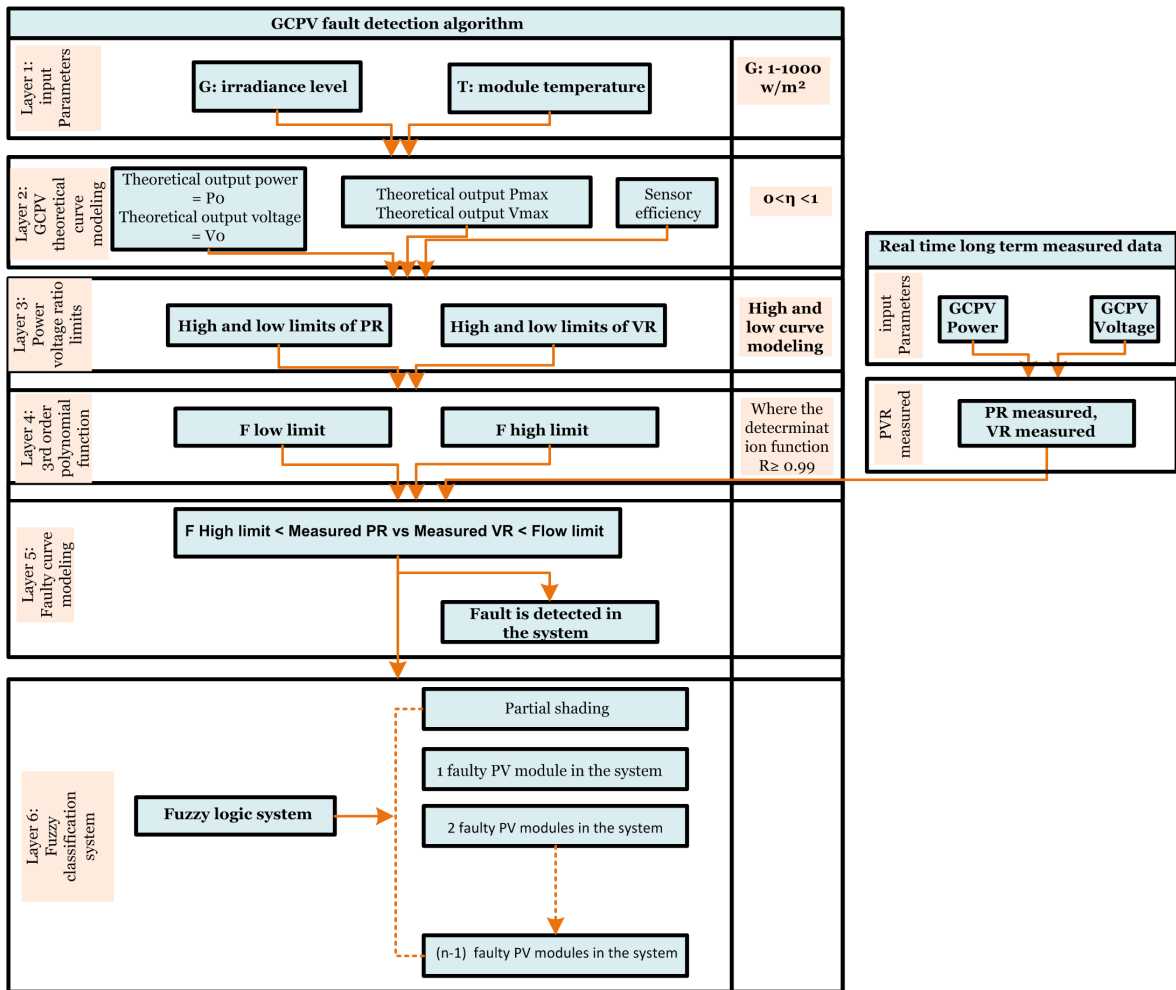
Figure 1.14: The detailed structure of the theoretical curve based method [35].

In addition, Lian et al have proposed an automatic fault detection and diagnosis method, for PV systems, by combining ANN and analytical based method [36]. In this method, a two-layered ANN type MLP was employed to predict the expected power of the PV system, which will be compared with the real measured one according to equation (1.16). Based on this comparison, a fault diagnosis strategy was developed by using a conventional analytical method.

$$\Delta_p = \frac{\left| P_{meas} - P_{pred} \right|}{P_{pred}} \tag{1.16}$$

The flowchart of this method is illustrated in Figure (1.15) [36], where ε_1 and ε_2 are predefined errors. N_{ss} is the number of PV modules in each string, $V_{oc\text{-}Si}$ denotes the open-circuit voltage of one string, and μ is the allowance variable.
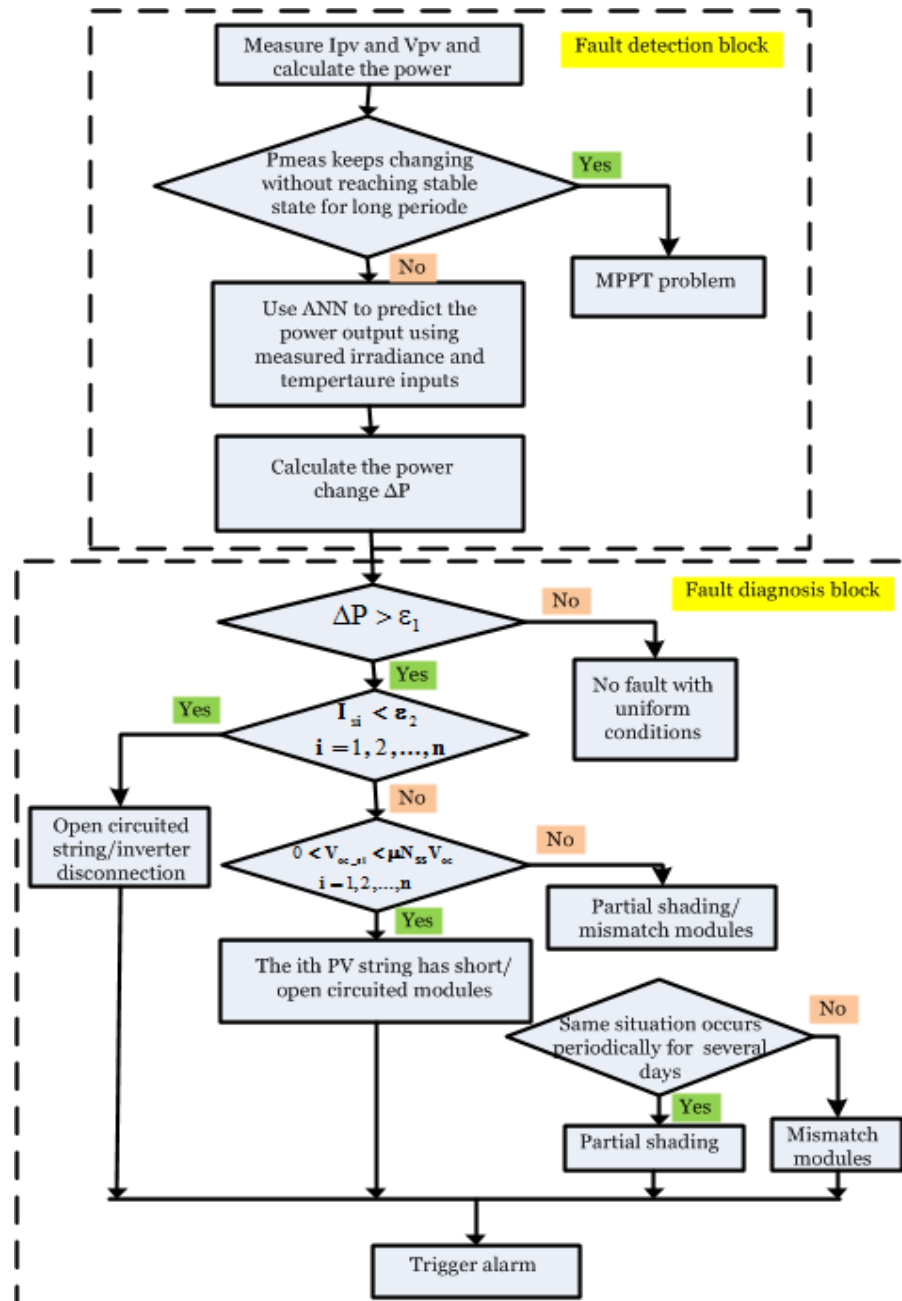
Figure 1.15: The detailed flowchart of lian's method
for faults detection and diagnosis [36].

Moreover, Todizara et al have proposed a fault detection strategy that uses the Least Square Method (LSM) and Fuzzy Logic system classification [37]. In this work, the Bishop model of the PV module has been used to simulate the PV system behavior under healthy and partial shading conditions. Simulations showed that the direct analysis of the generated voltage and power does not give sufficient information to detect and diagnose faults. For this reason, the LSM is used to compute the residuals, which are the least square error between measurements

and predictions. These residuals will be exploited by a fuzzy logic decision block to diagnose the type of fault.

The Bishop model of the PV module is defined using equation (1.17) [37].

$$I_{PV} = I_{Ph} - I_0 \left[\exp\left(\frac{V_{PV} + R_s I_{PV}}{V_t} \right) - 1 \right] - \frac{V_{PV} + R_s I_{PV}}{R_{sh}} \left[1 + L \left(1 - \frac{V_{PV} + R_s I_{PV}}{V_b} \right)^{-m} \right] \quad (1.17)$$

where: I_{PV} and V_{PV} are the PV cell current and voltage respectively. I_{Ph} is the light-generated current, I_0 is the dark saturation current, V_t is the thermal voltage, R_s and R_{sh} denote series and shunt resistances, respectively, L and m are the Bishop adjustment coefficients ($3.4 \le L \le 4$ and $m = 0.1$) and V_b is the cell breakdown voltage (-10V to -30V).

The diagram of the fuzzy logic classification system is depicted in Figure (1.16) [37].
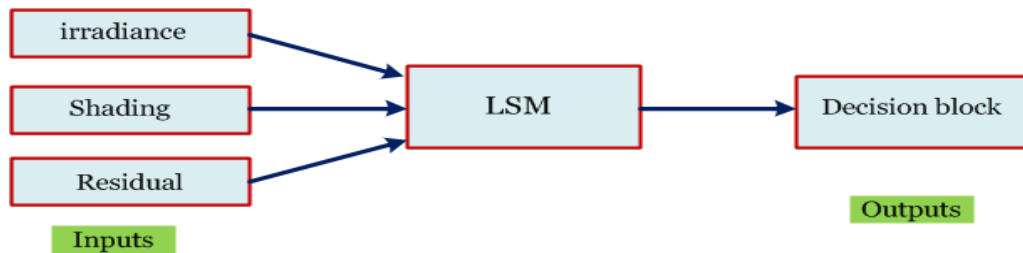
Figure 1.16: The diagram block of the fuzzy decision classification [37].

In addition, the (I-V) characteristic and the fuzzy logic classification system have also been used to elaborate a diagnosis method for PV systems [38]. This method relies on analyzing the effects of faults on the PV array/string (I-V) characteristic, to detect three faults: shading, increased series resistance losses and potential inducing degradation (PID) affecting the PV string. Moreover, the proposed strategy has been formalized as fuzzy logic sets and rules to gain more robustness and flexibility, such that it will be useful for wide range of crystalline-based PV systems. The different steps of this method are summarized in Figure (1.17) [38].
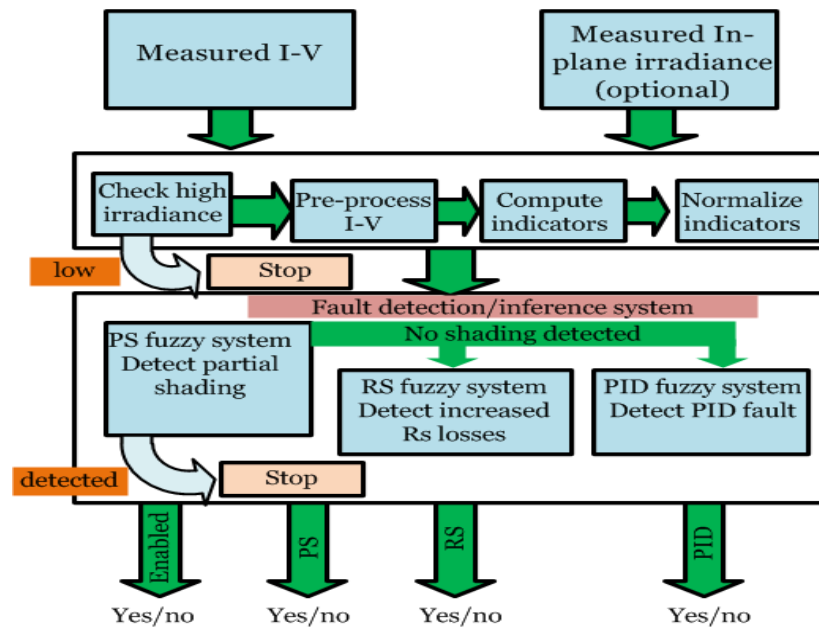
Figure 1.17: General structure of the diagnosis method based on I-V curve and fuzzy logic classification [38].

This method consists mainly on three stages. The first one relies on measuring the (I-V) curve and the in-plane irradiance (G) of the tested PV generator (module, string or array). The in-plane irradiance level (G) could be obtained by using irradiance sensors or mathematical estimations [38].

In the second stage, the in-plane irradiance level (G) is analyzed, such that if its value is lower than 500 W/m², the diagnosis procedure will be stopped. The main reason of this limitation is that the power losses, due to the shading and the increase in series resistance, cannot be visible, using I-V curves under low irradiance level [38]. After that, if the in-plane irradiance is higher than 500 W/m², the measured I-V characteristic will be filtered from any noise and used to compute a set of diagnostic indicators. The indicators are then normalized to maintain the independency with the system configuration. Finally, the normalized indicators of diagnosis are analyzed automatically, based on three fuzzy logic classifiers, to detect and diagnose the aforementioned faults [38].

The indicators that have been used in this method are: the I-V curve inflexion factor IV_f, the maximum power point factor MPP_f, the equivalent series resistance R_{se} and the fill factor FF.

This method can only be applied for the string inverter topology, which is not the most frequently used topology nowadays, and it gives efficient results only under high irradiance level (G>500W/m²).

Also, a methodology that uses three ANN classifiers was developed to detect the occurrence of partial shading (PS) [39]. The first ANN aims to detect the PS occurrence and distinguish it from the uniform change of climatic conditions, while the two remaining networks compute the shading factor and the number of shaded modules. The shading factor is defined as the ratio of the lower irradiance level on the shaded PV modules to the higher irradiance level on the rest of the PV array.

The inputs of the ANN that is used to detect the PS are the generated power and the short circuit current. The value 'one' of its output indicates that there is a PS, otherwise the ANN output is equal to 'zero'.

The inputs of the second network, which computes the shading factor, are the maximum power and the relative absolute change (ROC) of the output power due to shading. The output neuron gives the shadowing factor value. Finally, the last network, which is responsible of computing the number of shaded PV modules, considers the shading factor, generated from the second network, as input vector.

In addition, Vergura et al have proposed a descriptive and inferential statistical based method to supervise and monitor the PV system operation [40]. In this work, the performance of the PV system is firstly assessed, using an offline supervising step. This has been accomplished by using both descriptive and inferential statistics. After that, an efficient strategy of real-time monitoring and fault detection has been elaborated by using an online inferential algorithm. This method is used to detect bad connections between PV modules and inverter, using the inverter generated energy as input data. It has been validated using both Gaussian and non-Gaussian data distribution.

However, even though its efficiency to detect the aforementioned fault, this method has been considered as a complex statistical-based strategy. In fact, it requires the use of high complex tests (variance homogeneity test, normality test, ANOVA test and Kruskal-Wallis test) to detect only bad connections between PV modules and inverters.

Finally, an efficient strategy of faults detection and monitoring in PV systems using outlier detection rules was proposed in [13]. In this work, three outlier monitoring rules, named three-sigma (Shewhart), Hampel identifier and Boxplot rules have been used to detect line-line faults, open-circuits faults, degradation faults and partial shading occurrence.

This method uses the instantaneous current, generated from PV strings, to monitor the PV system operation and detect the faults occurrences.

The main advantages of this method are [13]:

- No model training procedure is required.
- The total independency to the weather conditions, which make it a costly inexpensive strategy.
- Its high simplicity for real-time implementation.

This method is mainly based on comparing the PV string normalized current I_{str_norm}, given by equation (1.18), with the healthy system outlier detection limits (high and low limits). If the normalized PV string current lies outside these limits, then a faulty operation is identified [13].

$$I_{str\_norm} = \frac{I_{str}}{I_{sc}} \tag{1.18}$$

Where I_{str} is the PV string operating current and I_{sc} is the PV module short-circuit current measured under STC.

Because it is mainly based on the value of the PV string current, this method only detects the faults occurrence and cannot diagnose its type.

Process History-based methods: Advantages and Drawbacks:

Methods of this class have showed their efficiencies in fault detection and diagnosis of PV systems. In fact, they require a minimal a prior knowledge of the system configuration, and use only healthy and faulty system data to monitor, detect and diagnose the faults.

However, these methods depend on the availability of quality input data, and their implementation is not an easy task, especially for real-time applications.

Moreover, multi-layer perceptron neural networks suffer from their slow training step, they do not support noisy data, and could fall in local minima instead of global one [9, 10]. These weaknesses could obviously affect the reliability, the efficiency and even the real time utilization of these methods.

Moreover, statistical monitoring charts [11-13] are important tools for monitoring sequential systems to make sure that they work stably and satisfactory. However, until recently statistical control charts have not been widely used to improve the performance of PV systems. In addition, the main shortcoming of the proposed monitoring chart-based methods is their limitation to detect incipient faults because they make decisions based only on the recent observations [11].

1.4.2. Quantitative Model-based methods:

The methods of this second category analytically compare real measured outputs with model-based outputs (simulated outputs) to detect and diagnose PV system failures. In this type of method, a fault will be declared when a large difference between the measured and estimated outputs is identified. Hereafter, a list of the most recently proposed works in this second category of methods is given.

Mahmoud Dhimish et al have proposed an efficient method of fault detection and diagnosis in GCPV systems based on voltage and power ratios [3]. In this work, the theoretical performance of the PV system is firstly simulated to compute the theoretical outputs voltage and power. Then, the ratios between the theoretical and measured voltage and power are computed and then analyzed, to detect the fault occurrence and diagnose its type. Current, Voltage and Power ratios are expressed by the following equations:

$$VR = \frac{V_{theoretical}}{V_{measured}} \tag{1.19}$$

$$PR = \frac{P_{theoretical}}{P_{measured}} \tag{1.20}$$

$$CR = \frac{I_{theoretical}}{I_{measured}} \tag{1.21}$$

Where: $P_{theoretical}$, $V_{theoretical}$ and $I_{theoretical}$ are the simulated outputs of power, voltage and current, respectively. $P_{measured}$, $V_{measured}$ and $I_{measured}$ are the real measured outputs of power, voltage and current, respectively.

The general structure of this strategy is depicted in Figure (1.18), where VR and PR computation details are shown in Figure (1.19). Moreover, detailed flowchart of the first region is illustrated in Figure (1.20). While those of the second and the third regions are shown in Figure (1.21).

Despite of its high capacity in fault detection and diagnosis, this method depends completely on the power and voltage ratios. Therefore, it depends on the accuracy of both simulation model and measurements devices.
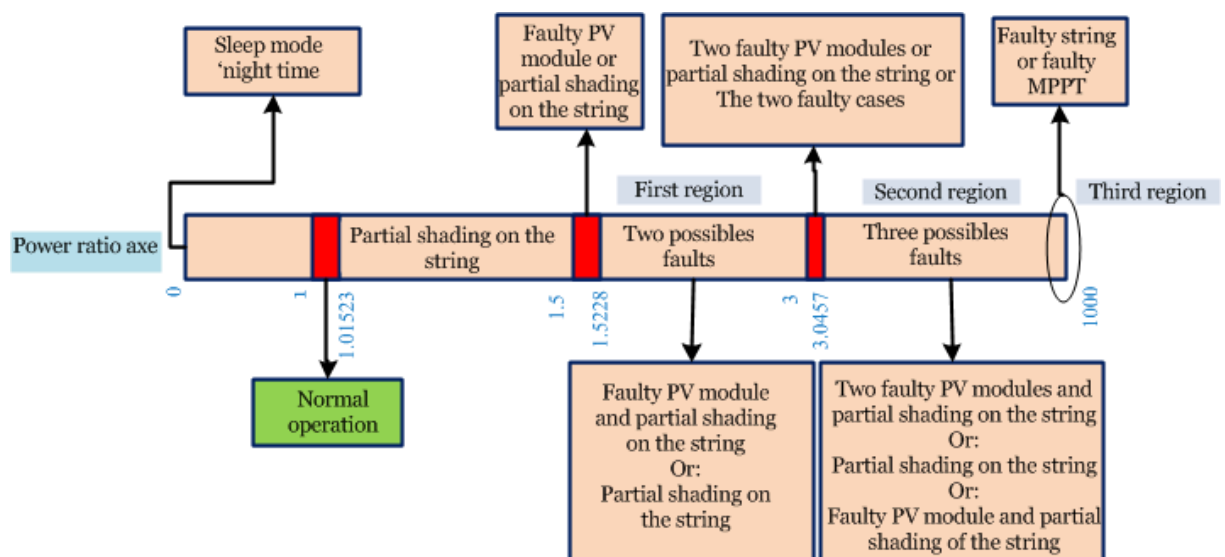
Figure 1.18: General structure of Dimish method for fault detection and diagnosis [3].
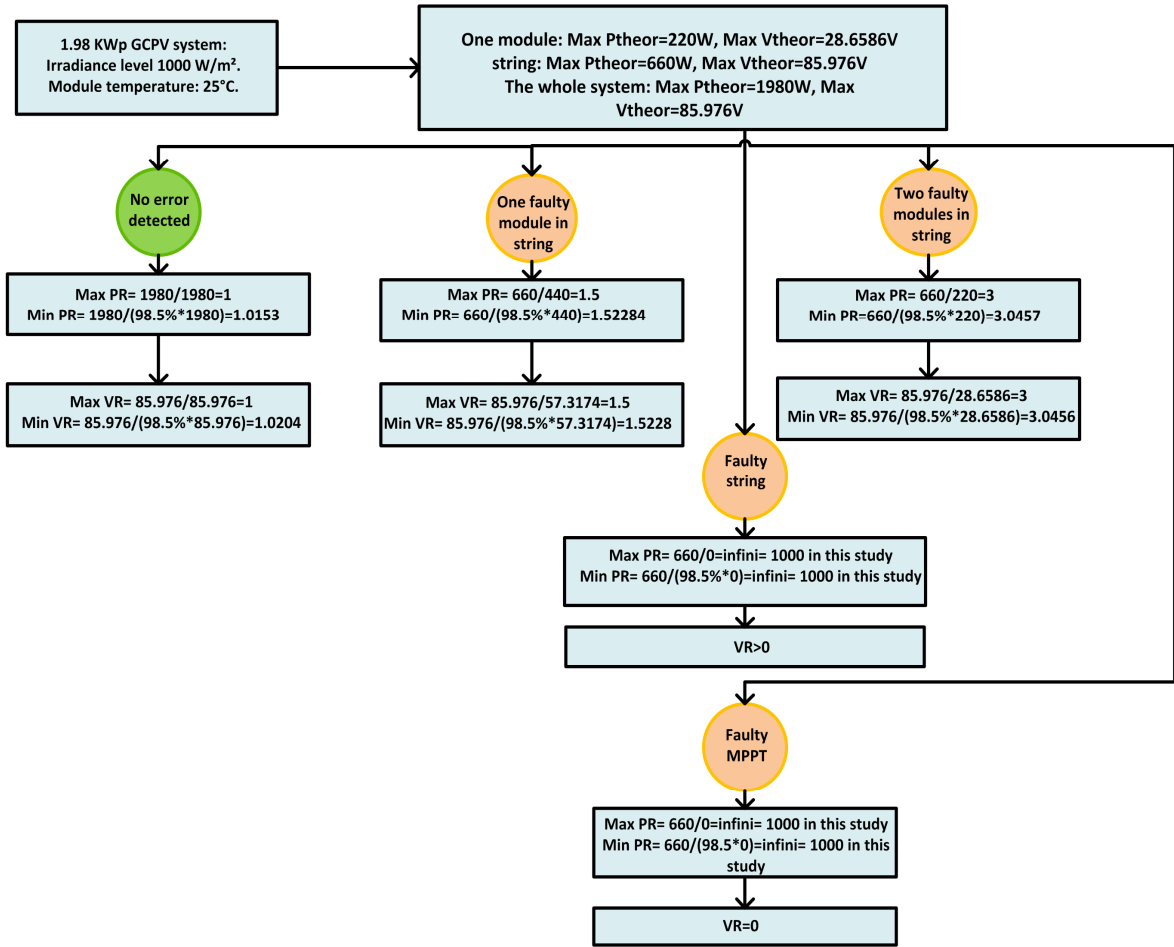
Figure 1.19: VR and PR computation details [3].
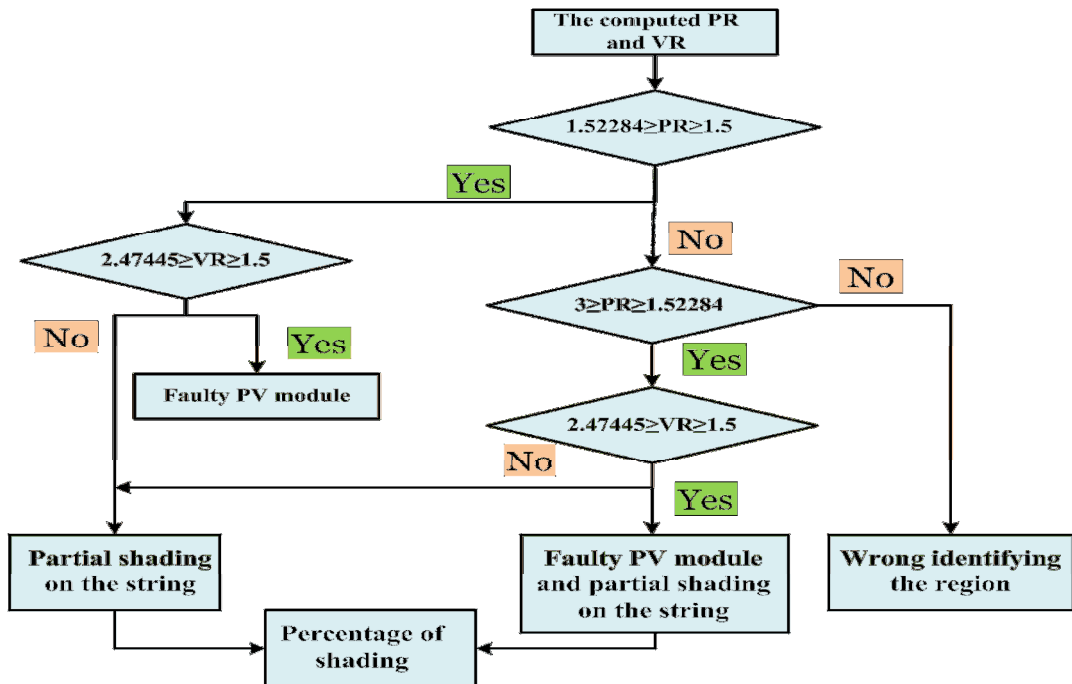
Figure 1.20: The detailed flowchart of the first region of Dimish method [3].
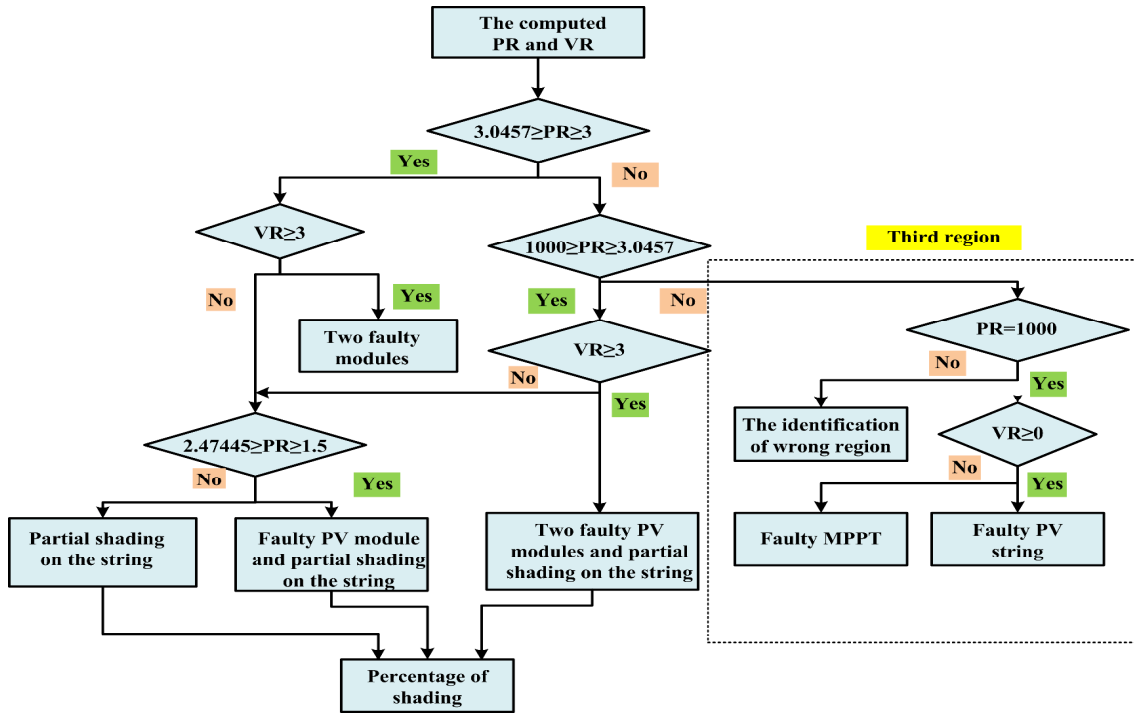
Figure 1.21: The detailed flowchart of the second

and third regions of Dimish method [3].

In addition, Chouder et al have developed a diagnosis method for GCPV systems based on the power losses analysis [41]. The main idea of this work consists to continuously measure the DC side power losses (capture losses) and then examine if these losses lie within some theoretical boundaries. The system is considered under faulty operation if the measured power losses are beyond these boundaries. Finally, in order to diagnose the fault type, current and voltage ratios are evaluated and monitored.

In this work, two novel indicators of power losses have been defined: Thermal capture losses (L_{ct}) and Miscellaneous capture losses (L_{cm}). The evaluation of these indicators allows the fault detection. While the evaluation of current and voltage ratios, CR and VR, is used to diagnose the fault type.

The DC side's total power losses, L_c, are defined as follows:

$$L_c = Y_r(G, T_c) - Y_a(G, T_c) = \frac{H_i}{G_{ref}}(G, T_c) - \frac{E_{dc}}{P_{ref}}$$

(1.22)

Where: $Y_r(G, T_c)$ and $Y_a(G, T_c)$ denote respectively the reference and the array yields, measured under real operating conditions of irradiance G and cell

temperature T_c. H_i is the total in-plan irradiance of the PV array. G_{ref} is the reference irradiance measured at STC (G_{ref}=1000W/m²). E_{dc} is the array produced energy and P_{ref} is the array output of maximum power.

Thermal capture losses occur when the PV module works at a temperature level higher than 25°C. Under this situation, the produced DC power will significantly decrease. In the other side, miscellaneous capture losses could be cause by: MPP tracker failures, module failure, string failure, angle of incidence losses and shading losses. The computation of thermal and miscellaneous capture losses is established by using the following equations [41]:

$$L_{ct} = Y_a(G, 25°C) - Y_a(G, T_c) \qquad (1.23)$$

$$L_{cm} = L_c - L_{ct} \qquad (1.24)$$

Detailed flowcharts of this method are given in Figures (1.22-1.23).
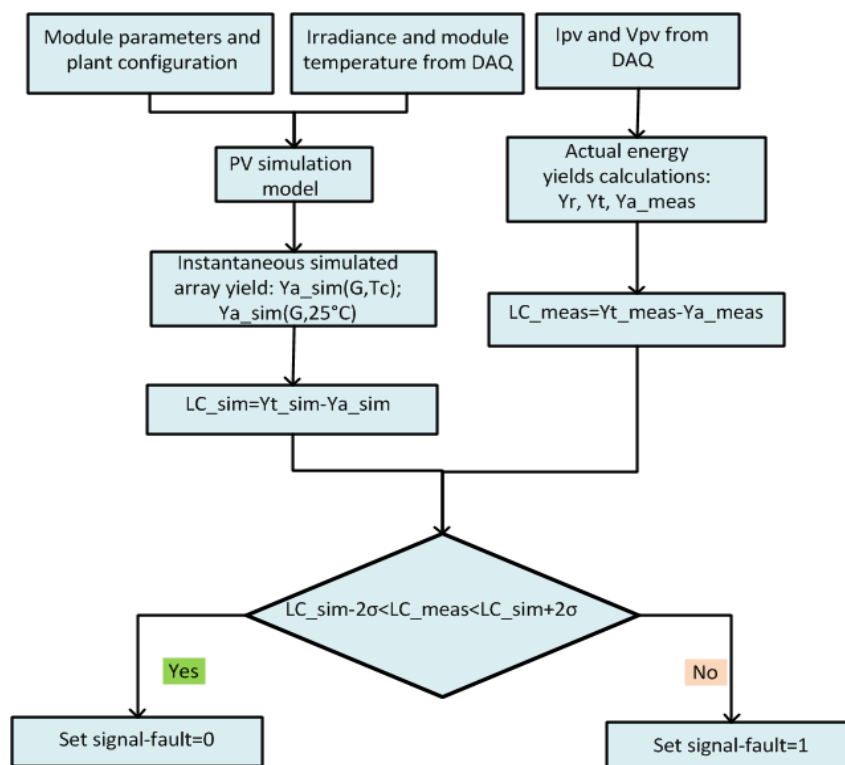
Figure 1.22: The flowchart of Chouder's method for faults detection and diagnosis[41].
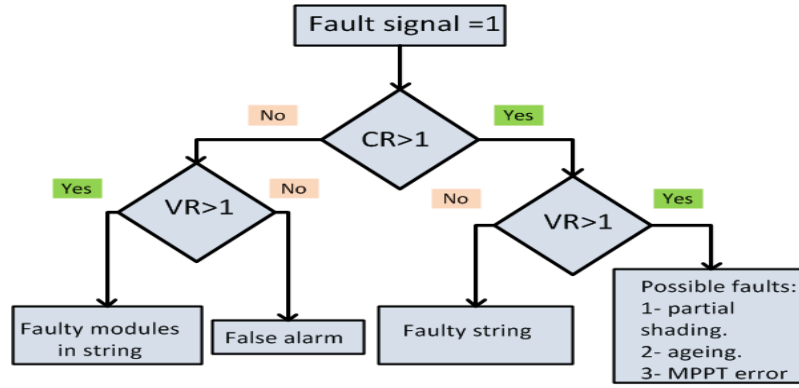
Figure 1.23: The flowchart of current and voltage ratios evaluation strategy [41].

The efficiency of this method has been experimentally tested for three operating cases: healthy system operation, faulty string operation (string open-circuited) and operation in the presence of partial shadowing.

Moreover, a detailed procedure of supervision and fault diagnosis in GCPV systems have been proposed in [42]. In this work, the fault detection is achieved by comparing the measured and simulated yields, while the fault diagnosis is carried out by analyzing and comparing DC current and DC voltage with a set of healthy system thresholds.

In this work, the measured meteorological conditions and the electrical parameters have been used to develop a LabVIEW-based software of: (i) capture losses computations, (ii) PV system simulation, (iii) online parameters supervision, and finally (iiii) fault detection and diagnosis [42].

The error between measured and simulated capture losses is given by:

$$\mathrm{EL}_c = \left| L_{c\_meas} - L_{c\_sim} \right| \tag{1.25}$$

The regular check of this error is required to detect any fault occurrence; i.e. if its value does not exceed predefined thresholds, given by equation (1.26), the system is considered working under healthy operation.

$$\mathrm{EL}_{c\_ref} - 2\sigma\left(\mathrm{EL}_{c\_ref}\right) \leq \mathrm{EL}_c \leq \mathrm{EL}_{c\_ref} + 2\sigma\left(\mathrm{EL}_{c\_ref}\right) \tag{1.26}$$

Where: EL_{c_ref} is the error between measured and simulated capture losses of the healthy system. σ is standard deviation of this error.

The values of the thresholds in equation (1.26) are set after statistical analysis of this error when the system is under healthy operation.

The detailed flowchart of the fault detection strategy is depicted in Figure (1.24).
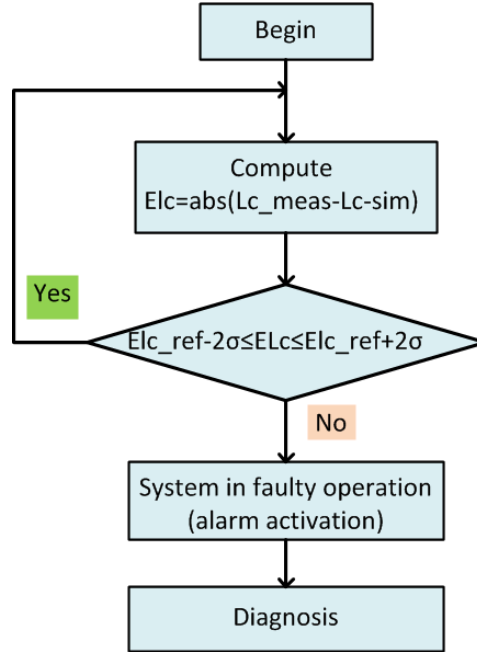
Figure 1.24:The flowchart of the fault detection procedure [42].

When the presence of faults is detected, two novels indicators, called current error E_i and voltage error E_v are computed using the following equations:

$$E_i = \left| I_{dc\_meas} - I_{dc\_sim} \right| \tag{1.27}$$

$$E_V = \left| V_{dc\_meas} - V_{dc\_sim} \right| \tag{1.28}$$

Where: I_{dc_sim} and I_{dc_meas} are simulated and measured DC current. V_{dc_sim} and V_{dc_meas} are simulated and measured DC voltage.

Similarly to the detection procedure described above, a set of predefined thresholds for E_i and E_v should be computed to diagnose the detected faults. In other term, the system is considered under healthy operation if:

$$E_{i\_ref} - 2\sigma(E_i) \le E_i \le E_{i\_ref} + 2\sigma(E_i) \tag{1.29}$$

$$E_{v\_ref} - 2\sigma(E_v) \le E_v \le E_{v\_ref} + 2\sigma(E_v) \tag{1.30}$$

Where: E_{v_ref} and E_{i_ref} are respectively the errors between the measured and simulated DC voltage and current for the healthy system. σ is the standard deviation of these errors.

The flowchart given in Figure (1.25) illustrates the detailed strategy of faults diagnosis.
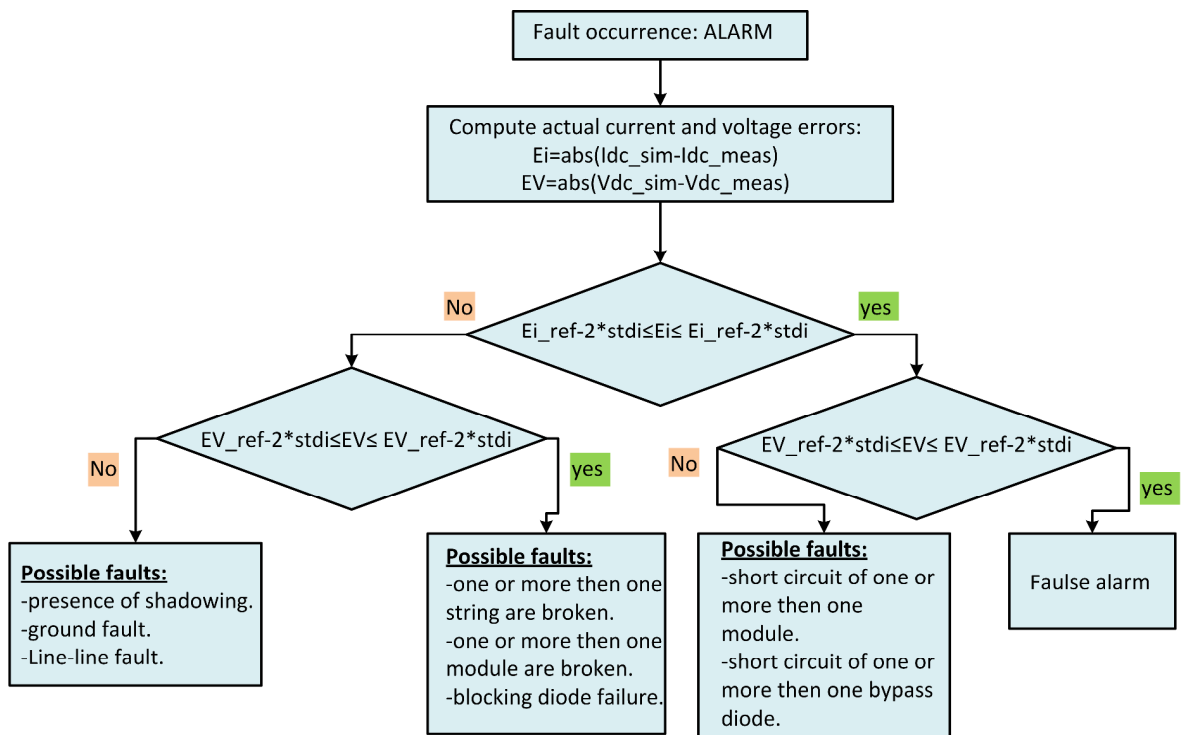
Figure 1.25: The detailed flowchart of the faults diagnosis strategy [42].

This method has been experimentally tested under an actual GCPV system to detect and diagnose the occurrence of: inverter disconnection, partial shading and string disconnection in the array. Nevertheless, its main drawback is that it categorizes the entire possibly occurred faults into three groups instead of individually separates each faulty case.

In addition, Mohammed Tadj et al have proposed an efficient approach of fault detection and diagnosis using GISTEL (Gisement solaire par télédétection: Solar Radiation by TeleDetection) approach [24]. In this work, the hourly global horizontal solar irradiance is firstly estimated using an improved GISTEL-Fuzzy model. This last is then used to estimate the DC power. To detect the faults occurrences, the estimated DC power is compared with the real measured one.

The flowchart of this method is shown in Figure (1.26), where d is the standard deviation of the estimated (simulated) output DC power, for a clear sky. CR and VR denote respectively the current and voltage ratios.
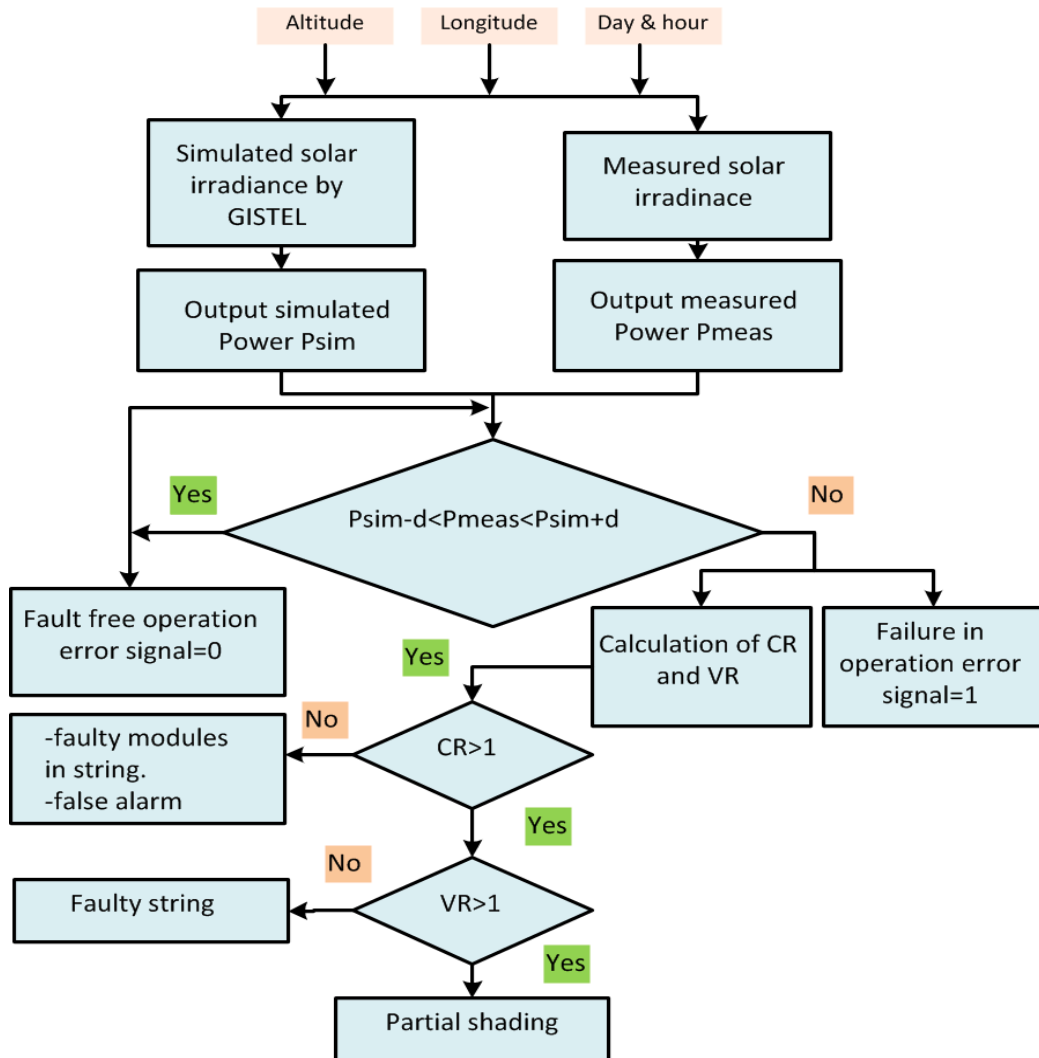
Figure 1.26: The flowchart of the GISTEL_Fuzzy based method
for fault detection and diagnosis [24].

As can be seen from the given flowchart, the presence of a fault is identified by comparing measured and simulated DC power, while, the detection of its type can be achieved by analyzing both current and voltage ratios.

Also, Drews et al have developed a fault detection method that uses satellite observation data to estimate the PV energetic yield [43]. Based on this strategy, the fault occurrence is declared when a significant deviation between the real measured and the simulated energetic yields is identified.

Moreover, another efficient approach to detect PV arrays faults and partial shading, based on PV array voltage, PV array current and measured irradiances was proposed by Hariharan et al [44]. This method has the ability to distinguish between three possible cases: normal operation, partial shading, and permanent faults (bridge fault and open-circuits fault).

The main idea of this method is to compute predefined thresholds, obtained from simple formulas, and which may be deemed constant for a given PV array under all conditions, to detect and diagnose faults [44]. This method does not require the use of a large amount of experimental training data.

In this method, two novel variables, named gamma (γ) and array losses (L_{array}), have been introduced as follows:

$$\gamma = \frac{V_{pv} * I_{pv}}{G} \tag{1.31}$$

$$L_{array} = \underbrace{\left(P_m \left[\frac{G}{G_0}\right]\right)}_{\text{Expected Power}} - \underbrace{\left(V_{pv} * I_{pv}\right)}_{\text{Actual Power}} \tag{1.32}$$

Where: G is the instantaneous irradiance of an un-shaded portion of PV array. V_{pv} and I_{pv} are the instantaneous PV array voltage and current respectively. L_{array} is the difference between the instantaneous expected power and the actual power. P_m is the PV array maximum power, measured at the reference irradiance G_0=1000 W/m².

The detailed flowchart of this method is shown in Figure (1.27). The constants used in this flowchart are summarized in table (1.3).
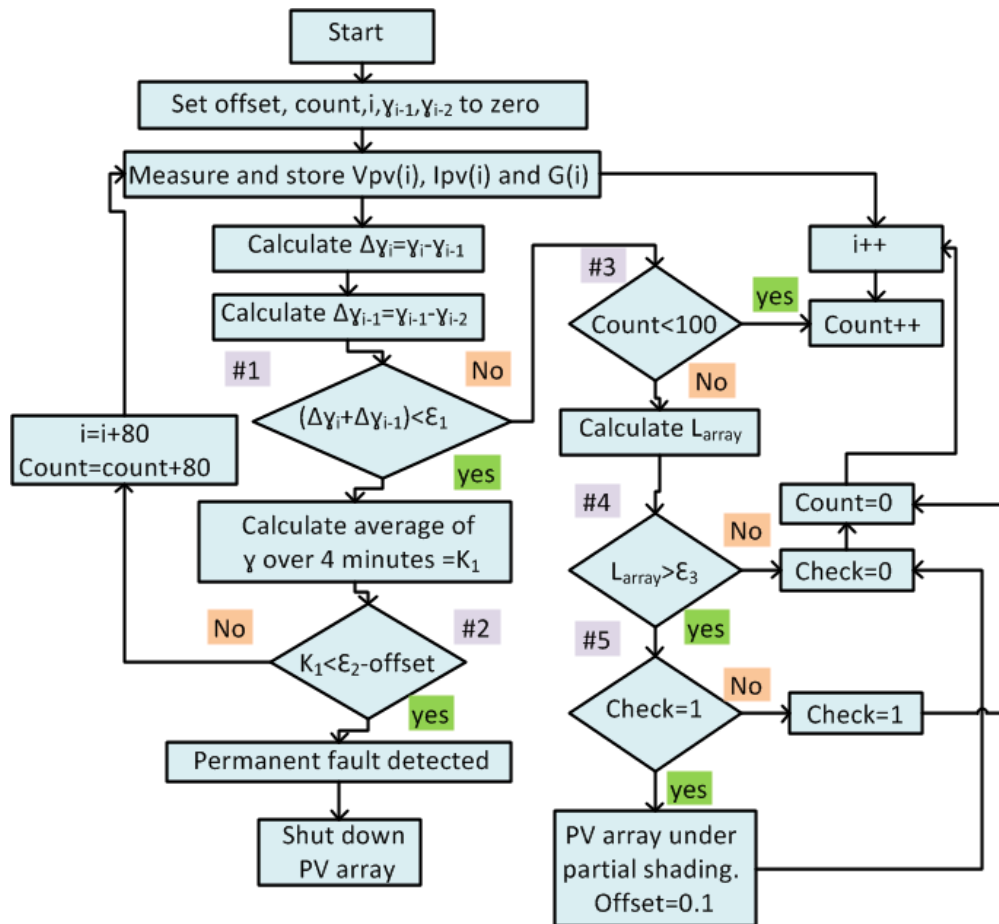
Figure 1.27: The flowchart of Hariharan's method

for fault detection and diagnosis [44].

Table 1.3: Lists of constants used in Hariharan's method [44].

Constant	Formula	Value
ε_1	$-\dfrac{\left(L_{array} \text{ for one module mismatch fault under STC}\right)}{G_0}$	-0.2 m²
ε_2	$\left(\dfrac{\text{MPP of PV array under STC}}{G_0}\right)+\varepsilon_1$	1.4 m²
ε_3	L_{array} at G=100W/m²	100 W
offset	$\dfrac{\varepsilon_3}{G_0}$	0.1m²
K_1	$\dfrac{\sum\limits_{j=i}^{i+80}\gamma_j}{80}$	/

Results from the aforementioned works show the high efficiency of this category of method to detect and diagnose failures. However, the use of an accurate simulation model that mimics the system operation under healthy and faulty operations is required. Moreover, this type of methods suffers from the manual strategy of threshold computation and it requires that the PV array works at MPP.

1.4.3. Signal Processing-based methods:

This category of methods relies on applying advanced techniques of signal processing to detect and diagnose PV systems failures. Hereafter, a list of the most recently proposed works is introduced.

Il-Song Kim has suggested an online fault detection algorithm in PV systems based on the Wavelet Transformation [45]. In this work, the Multi-Level Decomposition (MLD) wavelet transformation has been used to detect, localize and diagnose faults in power conditioning systems (DC/DC and DC/AC converters) [45].

The developed strategy of fault detection employs 3-level MLD tree to detect switch open and over harmonics failures. In fact, power conditioning systems (PCS) transforms the generated DC voltage into AC grid voltage via switches (**FET**: Field Effect Transistor; **IGBT**: Insulated Gate Bipolar Transistor). In the case of breakdown of these switches, a fault of switch open will be occurred. During this fault, the distorted waveform of PCS current will keep providing high order harmonics to the grid [45].

The proposed algorithm mainly relies on decomposing the measured signal of each wavelet tree level, and extracts its coefficients. The variation of these extracted coefficients will be then analyzed to identify the types of faults.

Even though this is a simple and efficient strategy, it is useful only to detect AC side failures.

In addition, Yihua Hu et al have proposed an efficient strategy to detect the mismatch of PV modules using a thermography-based temperature distribution analysis [15]. In this work, thermal images of an actual operating PV system

are recorded, processed and introduced to a mathematical model, to extract useful information of mismatch fault. This information is mainly based on the temperature distribution, and it is used to detect mismatch faults and to develop an enhanced MPPT strategy that overcomes the mismatch fault effect on MPPT efficiency.

In this work, PV system modeling has been elaborated by combining both electrical and thermal models of solar cell. These models are interlinked via the energy balance principle [15]. Based on the severity of mismatch fault, three classes of mismatch have been defined and analyzed: minor, medium and heavy mismatch faults.

This method can be summarized as follows:

- Image that corresponds to the surface temperature of the tested PV panel is recorded via a thermal camera.
- This image is send to a central computer to analyze its thermal feature.
- The thermography (temperature distribution) of each PV module is extracted via a freehand cropping program of Matlab^{TM} software.
- For each PV module, mismatch fault occurrence can be detected by computing its relative temperature with a reference.
- In order to distinguish between the three types of mismatch fault (minor, medium and heavy faults) the generated (I-V) curve, from each PV module, is analyzed.

This strategy requires the use of a sophisticated thermal camera, for each PV array, to detect and diagnose mismatch faults. Therefore, it is a costly expensive fault detection strategy.

Based on Time-Domain-Reflectometry (TDR), Schirone et al have developed an efficient strategy for detection and localization of short-circuit and open-circuit faults [46]. TDR is a well-known electrical method, usually used to measure the electrical characteristic of a transmission line, and to detect its breakdown point. Figure (1.28) shows the fundamental principle of TDR method [46]. As can be seen from this figure, an input signal is introduced to a transmission line. This signal will be then compared with the reflected signal (response signal from the devices under test (DUTs)). Comparison between these two signals forms a

shifted signal. This last signal is used to localize the fault position, while its waveform is used to diagnose the type of faults (short-circuit, open-circuit, degradations… etc.).
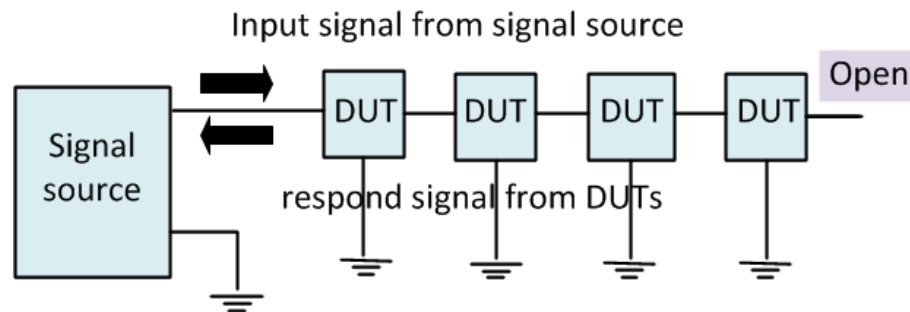
Figure 1.28: The fundamental principle of TDR method [46].

Also, an experimental study of fault detection and localization in PV systems was carried out by Takumi et al [47]. In this study, the efficiency of Earth-Capacitance-Measurement (ECM) and TDR methods to detect and localize string disconnection and degradation faults was tested. ECM is also a frequently applied strategy that is used to localize the disconnection position of a transmission line (Figure (1.29)). In this method, the distance (x) from the start point of the transmission line to the fault point is computed using equation (1.33).

$$x = \left(\frac{Cx}{Cd}\right)D \qquad (1.33)$$

Where: Cx is the earth capacitance value, measured from start point to the fault point (disconnection point); Cd is the earth capacitance value, measured for the whole transmission line; D is the length of the whole transmission line.
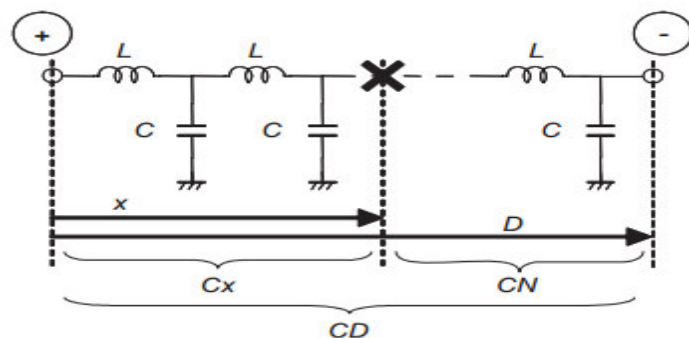
Figure 1.29: Fundamental principle of ECM method [47].

Detection and localization of string disconnection fault, based on ECM method, has been reached by considering the PV string as a transmission line. Therefor, the module number (num_mod) to the disconnection position is computed using equation (1.34).

$$num\_mod = \left(\frac{Cx}{Cd}\right)M \qquad (1.34)$$

Where M is the total number of PV modules within the string.

Moreover, TDR has been used to detect and localize degradation fault.

Signal Processing-based Methods: advantages and drawbacks:

The analysis of the aforementioned strategies shows the efficiency of this category of methods to detect and diagnose PV system faults. However, it suffers from several serious drawbacks:

The TDR based method is limited by its technical requirements to turn off the entire PV system, which leads in reducing the system's energetic performance. In addition, this method requires expensive and sophisticated tools since it needs to analyze input and output reflected signals.

In the other side, ECM method requires additional and expensive monitoring tools, like LCL meter. Moreover, this method cannot identify the partially-shaded PV module disconnection due to the presence of bypass diodes.

Finally, the thermal-camera based method is an expensive strategy, and it cannot guaranty the fault detection until reaching a dangerous faulty situation.

1.5. Conclusion:

In this chapter, the research topic has been firstly discussed through a general study of PV systems, as well as their different types of faults. Then, a literature review of the different methods for faults detection and diagnosis in the PV systems has been conducted. This review has allowed us to classify the methods of faults detection and diagnosis into three main types: process-history based methods, quantitative-model based methods and signal-processing based methods. The analysis of these methods has allowed us to clearly identify their limitations, which will be addressed in the next chapters.

CHAPTER 2

PV ARRAY MODELING AND VALIDATION

2.1. Introduction:

This chapter clarifies the developed strategy of PV array modeling and validation. In fact, the one diode model (ODM) of the PV module is firstly described, then, the procedure of PV module parameters identification is explained in detail. After that, the developed strategies of cost criterion minimization, using ABC [18, 19] and best-so-far ABC [20, 21] algorithms, are presented. Finally, an efficient strategy of MPP estimation based on the identified parameters is elaborated.

The accuracy of the identified parameters is tested using real static (I-V) curves of different PV modules from several technologies. While, the developed strategy of MPP estimation is experimentally validated using real measurements collected from Algerian and Spanish PV systems.

2.2. PV module modeling:

Usually, the solar cell/module is described by the well-known one diode model (ODM). This model describes the PV module behavior by the following electrical circuit [48-50]:
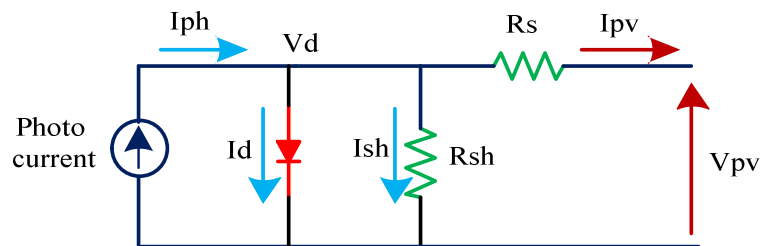
Figure 2.1: The one diode model of a solar cell.

In this model, the output current versus the output voltage is given by the following equation [41]:

$$I_{PV} = I_{ph} - I_0\left(\exp\left(\frac{q(V_{PV} + R_s I_{PV})}{nk_B T}\right) - 1\right) - \frac{V_{PV} + R_s I_{PV}}{R_{sh}} \qquad (2.1)$$

where:

I_{PV} and V_{PV} denote respectively current and voltage outputs of the PV module; I_{ph} stands for the light-generated current. I_0 is the diode saturation current; R_s and R_{sh} are series and shunt resistances respectively; n is the diode ideality factors; k_B is the Boltzmann constant $\left(k_B = 1.3806503 \times 10^{-23}\, j/k\right)$, T is the cell temperature and q is the electronic charge $\left(q = 1.60217646 \times 10^{-19}\, c\right)$.

2.3. PV module parameters identification:

2.3.1: Fundamental principle:

The ODM parameters identification stage can be described as an optimization problem, where the cost criterion to be minimized is the Root Mean Square Error (RMSE) between the measured and estimated currents, as given bellow:

$$RMSE = \sqrt{\frac{1}{S}\sum_{i=1}^{S}\left[g_i\left(I_{meas}, V_{meas}, \gamma\right)\right]^2} \tag{2.2}$$

where:

$$g\left(I_{meas}, V_{meas}, \gamma\right) = I_{meas} - \left(I_{ph} - I_0\left(\exp\left(\frac{q\left(V_{meas} + R_s I_{meas}\right)}{nk_B T}\right) - 1\right) - \frac{V_{meas} + R_s I_{meas}}{R_{sh}}\right) \tag{2.3}$$

I_{meas} and V_{meas} are the measured current and voltage of the PV module, obtained from a set of experimental outdoor measurements, using an (I-V) curve tracer. The symbol γ defines the vector of ODM parameters, and S is the size of experimental (I-V) curve data used during the ODM parameters identification stage. The ODM parameters identification scheme is highlighted in Figure (2.2) [49].

The used temperature during this stage denotes the PV module temperature. The main goal of this stage is to find the optimal ODM parameters $\left[I_{ph}^*, I_0^*, n^*, R_s^*, R_{sh}^*\right]$ that minimize the cost criterion and gives the lowest RMSE value.
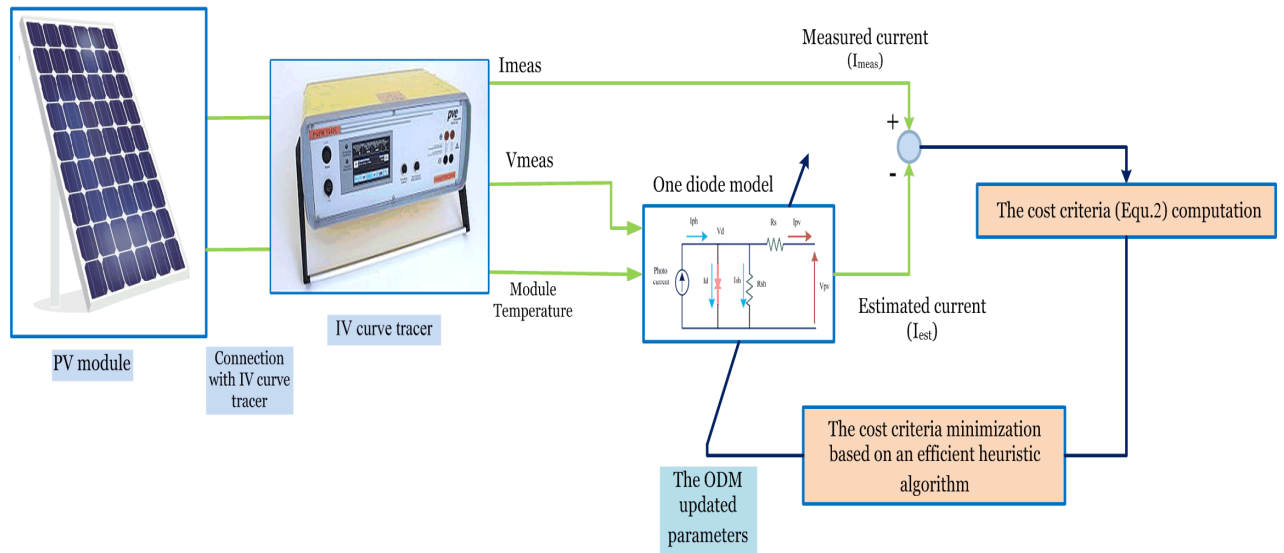
Figure 2.2: The ODM parameters identification scheme.

2.3.2. The developed strategies of cost criterion minimization:

In this thesis, two heuristic algorithms, namely Artificial Bee Colony (ABC) and best-so-far ABC, have been used to minimize the cost criterion of equation (2.2).

2.3.2.1. The Artificial Bee Colony (ABC) Algorithm:

ABC algorithm is a heuristic optimization algorithm, inspired from the foraging behavior of honey bees. In this algorithm, the bees are subdivided into three categories such as: employed, onlooker and scout bees [18]. The aim of the employed bees is to exploit the nectar sources explored before and share the information with the waiting bees (onlooker bees) in the hive. This information is related to the quality of food source sites exploited by the employed bees. Based on this information, the onlooker bees inside the hive decide on the food source to exploit. The goal of the scout bees is to search randomly around the hive in order to find a new food source site [18, 19].

The basic steps of the ABC algorithm are summarized in the following points [18]:

a. Initializing phase:

- Set the algorithm parameters (SN, LN, D, MCN, limit$=SN \times D$, cycle =1). SN and LN denote respectively the number of employed and onlooker bees, MCN is the Maximum Cycle Number, D is the solution dimension, and it denotes the number of parameters to be identified. Finally, limit is the number of unsuccessful trials to produce a better food source.

- Randomly generate the initial positions of food sources (initial solutions) using the following equation:

$$x_{ij} = x_j^{min} + rand\,[0,1]\left(x_j^{max} - x_j^{min}\right)$$

(2.4)

Where: $i = 1,...,SN$, $j = 1,...,D$, x_j^{min} and x_j^{max} are the lower and upper limits of the solution position X in the dimension j, and rand $[0,1]$ is a random number between 0 and 1.

- Evaluate the fitness of the generated positions.

While cycle $\leq$ MCN do:

b. Employed bees phase:

- For all employed bees (i=1… SN) do:

▪ Generate new food source according to equation (2.5):

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right)$$

(2.5)

Where v_{ij} is the new solution, x_{ij} is the previous solution, ϕ_{ij} is a random number between [-1,1], The indexes k and j are randomly selected from $\{1,2,3,...,SN / k \neq i\}$ and $\{1,2,3,...,D\}$, respectively.

▪ Evaluate the fitness value of the generated positions and compare the old and the new solutions, select then the best one.

c. Onlooker bees phase:

- For all onlooker bees (i=1… LN) do:

▪ Evaluate the probability p_i given by:

$$p_i = \frac{fit(x_i)}{\sum_{n=1}^{SN} fit(x_n)}$$

(2.6)

where $fit(x_i)$ is the fitness value of the position x_i.

- Assign to each onlooker bee a food source using the roulette wheel and the probability p_i.

- According to equation (2.7), generate a new food source for each onlooker bee (i=1,2,…,LN):

$$v_{ij} = x_{ij} + \phi_{ij}\left(x_{ij} - x_{kj}\right) \qquad (2.7)$$

Where v_{ij} is the new solution, x_{ij} is the old solution, ϕ_{ij} is a random number between [-1,1]. The indexes k and j are randomly selected from $\{1,2,3,...,SN/k \neq i\}$ and $\{1,2,3,...,D\}$, respectively.

- Evaluate the fitness value of the new generated positions and compare the old and the new solutions, select then the best one.

d. Scout bees phase:

- For all the onlooker bees, which their food sources quality does not improve in the determined number (limit) of cycles:

  - The food sources are abandoned and their onlooker bees become scout bees, which randomly generate new food source positions based on equation (2.8).

$$x_{ij} = x_j^{min} + \text{rand}\,[0,1]\left(x_j^{max} - x_j^{min}\right) \qquad (2.8)$$

- Evaluate the fitness of the generated positions.

- Save the best solution as the new solution.

- cycle =cycle+1.

End while.

2.3.2.2. <u>The best-so-far ABC Algorithm:</u>

Even the high accuracy of the ABC algorithm in solving a lot of optimization problems in different areas [51], the main weakness of this algorithm is its low convergence speed in some situations [20]. In order to solve this issue, the best-so-far ABC algorithm has been proposed to improve both exploration and exploitation steps [20].

The different steps of the best-so-far ABC algorithm are given below [20]:

a. Initializing phase:

  - Set the algorithm parameters (SN, LN, D, MCN, limit= $SN \times D$, cycle =1).

  - Randomly generate the initial positions of the food sources (initial solutions) using equation (2.4).

  - Evaluate the fitness of the generated positions.

 While cycle $\leq$ MCN do:

b. Employed bees phase:

  - For all employed bees (i=1… SN) do:

    ▪ Generate new food source according to equation (2.5):
    ▪ Evaluate the fitness value of the generated positions and compare the old and the new solutions, select then the best one.

  - Find the best-so-far food source x_b (the best food source position from all employed bees).

  - For all employed bees, determine the fitness f_b of the best-so-far food source x_b (from all dimensions).

c. Onlooker bees phase:

  - For all onlooker bees (i=1… LN) do:

    ▪ Evaluate the probability p_i given in equation (2.6).
    ▪ Assign to each onlooker bee a food source using the roulette wheel and the probability p_i .
    ▪ According to equation (2.9), generate a new food source for each onlooker bee (i=1,2,…,LN):

$$v_{id} = x_{ij} + \phi \, f_b \left(x_{ij} - x_{bj} \right)$$
(2.9)

  where v_{id} is the ith onlooker bee's new candidate solution in dimension d.

$d = 1, ..., D$, x_{ij} is the ith selected solution of the jth selected dimension, and x_{bj} is j^{th} best-so-far selected food source, f_b is the best-so-far food source's fitness value.

- Evaluate the fitness value of the generated positions and compare the old and the new solutions, select then the best one.

d. Scout bees phase:

- For all the onlooker bees, which their food sources quality does not improve in the determined number (limit) of cycles:

  ▪ The food sources are abandoned and their onlooker bees become scout bees, which randomly generate new food source positions based on equation (2.10).

$$v_{ij} = x_{ij} + \phi_{ij}\left[w_{max} - \frac{cycle}{MCN}\left(w_{max} - w_{min} \right) \right]x_{ij} \qquad (2.10)$$

where: x_{ij} is the abandoned food source position. v_{ij} is the new food source position of the scout bee. ϕ_{ij} is a random number between [-1,+1]. w_{max} and w_{min} are the maximum and the minimum percentage of the position adjustment for the scout bees. The values of w_{max} and w_{min} are fixed to 1 and 0.2, respectively.

- Evaluate the fitness of the generated positions.

- Save the best solution as the new solution.

- cycle =cycle+1.

End while.

2.3.3. <u>Parameters identification results:</u>

2.3.3.1. <u>Identification results using the ABC algorithm:</u>

In order to check the effectiveness of the identified parameters, obtained by using the ABC algorithm, outdoor measurements of (I-V) curves from three different PV modules (such as Isofoton106/12, SILIKEN (SLK60P6L)

and PHHOTOWATT-poly (PW1650)) have been used. The electrical characteristics of these PV modules are summarized in table (2.1).

Table 2.1: The electrical characteristics of the three PV modules

	ISOFOTON 106/12	SILIKEN (SLK60P6L)	PHHOTOWATT-poly (PW1650)
P_{mpp} (W)	106	220	155
V_{oc} (V)	21.6	36.7	43
I_{sc} (A)	6.54	8.10	4.8
V_{mpp} (V)	17.4	29.2	34
I_{mpp} (A)	6.10	7.54	4.6
αI_{sc} (%/°C)	0.060	0.062	0.0015
βV_{oc} (%/°C)	-0.36	-0.356	-0.158

Table (2.2) summarizes the upper and lower limits for the three PV modules parameters.

Table 2.2: Upper and lower variation limits of the three PV modules

	ISOFOTON 106W-1V2	SILIKEN (SLK60P6L)	PHHOTOWATT-POLY(PW1650)
R_S(Ω)	[0 – 1]	[0 - 1]	[0 - 1]
R_p (Ω)	[0 - 400]	[0 - 400]	[0 - 400]
I_{ph0} (A)	[0 - 10]	[0 - 10]	[0 - 10]
I_{sat} (A)	$[10^{-7} – 10^{-4}]$	$[10^{-7} – 10^{-4}]$	$[10^{-7} – 10^{-4}]$
n	[0 - 75]	[70 - 130]	[60 - 160]

The adjustable parameters of ABC algorithm have been chosen as depicted in table (2.3).

Table 2.3: The adjustable parameters of the ABC algorithm

ABC parameters	Values
Colony size	150
Number of employed bees	75
Number of onlooker bees	75
The limit of the scout bees	150 X 5
Maximum number of iteration	10000

The comparison between the measured (I-V) characteristics and the estimated using the identified parameters, for each PV module and for different

meteorological conditions, are illustrated in Figures (2.3-2.5). Finally, the convergence rate of the ABC algorithm during the identification process is illustrated in Figure (2.6).
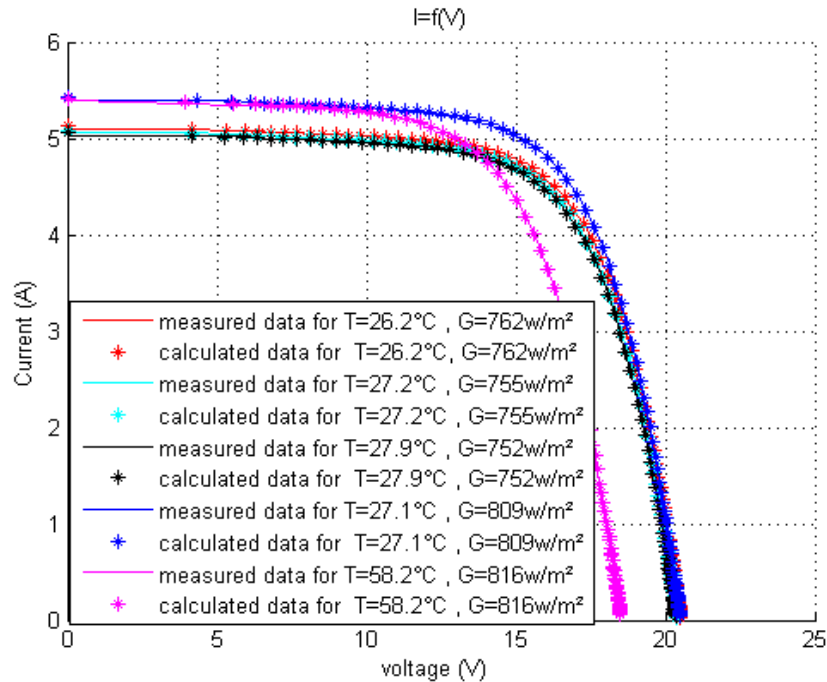
Figure 2.3: Measured and calculated (I-V) curve of Isofoton106/12 PV module for different meteorological conditions.
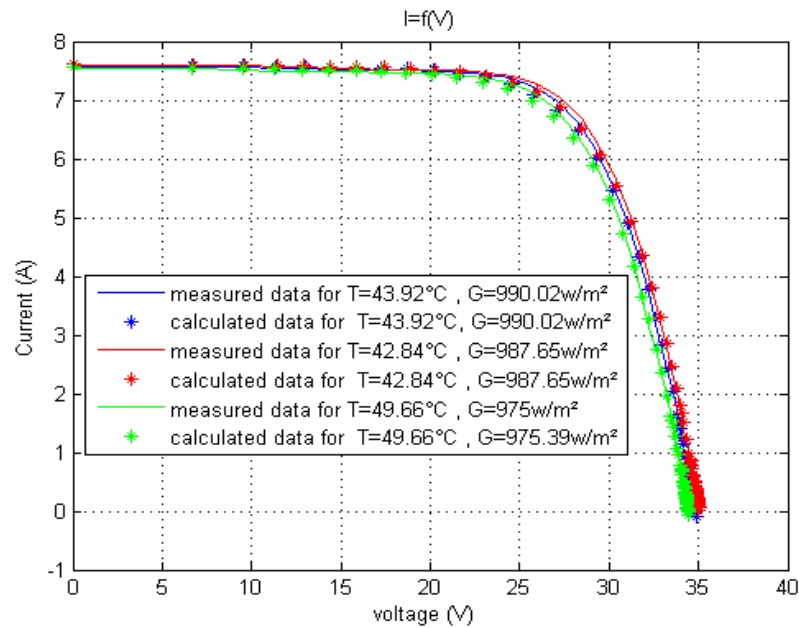
Figure 2.4: Measured and calculated (I-V) curve of SILIKEN (SLK60P6L) PV module for different meteorological conditions.
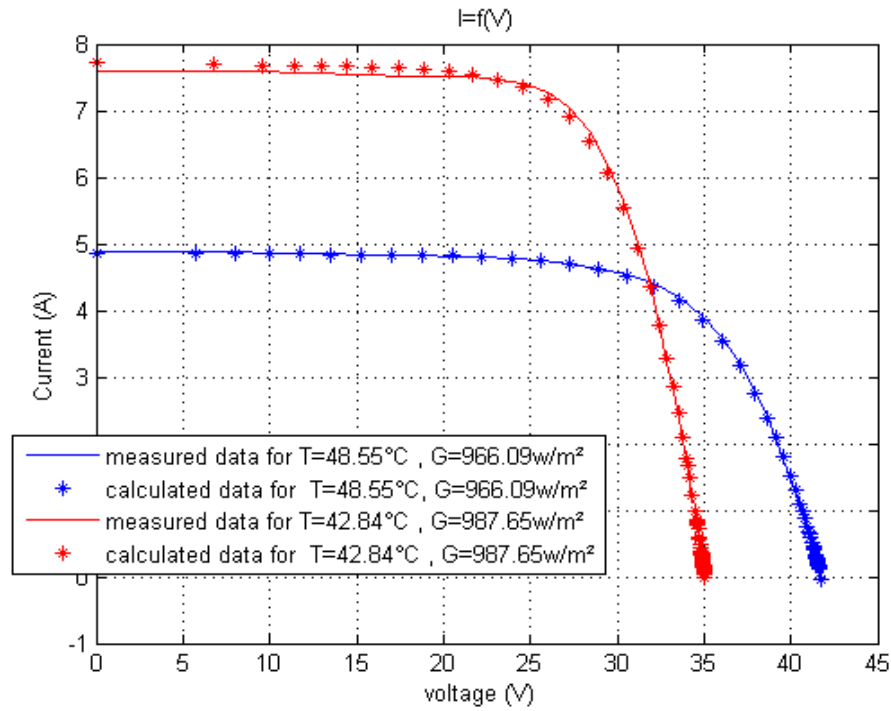
Figure 2.5: Measured and calculated (I-V) curve of PHHOTOWATT-POLY
(PW1650) PV module for different meteorological conditions.
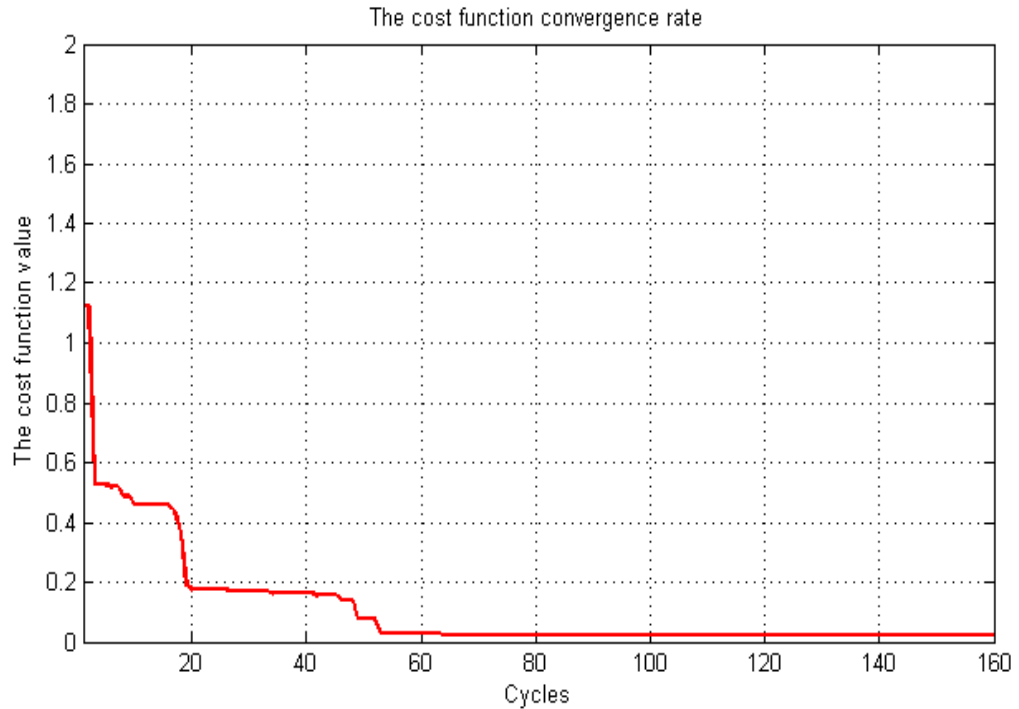
Figure 2.6: The convergence rate of the ABC algorithm during
the extraction process of Isofoton106/12 PV module.

The identified parameters of all PV modules are summarized in table (2.4).

Table 2.4: The final identified parameters of the three PV modules.

PV Modules	$I_{Ph}(A)$	$I_{sat}(A)$	$R_s(\Omega)$	$R_{sh}(\Omega)$	n	RMSE
Isofoton 106/12	6.68	1.43e-5	0.12	143	61.913	0.015
SILIKEN	7.67	1.19e-5	0.21	400	95.50	0.069
PHHOTOWATT	5.09	4.61e-5	0.25	387	130.35	0.029

In order to check the performance of the proposed ABC algorithm, a comparative study with two others optimization algorithms found in the literature [52, 53] has been carried out using Isofoton 106W/12V PV module. The algorithms are: Differential Evolution (DE) [53] and Particle Swarm Optimization (PSO) [52] algorithms. In order to compute the average root mean square error (RMSE) value and the standard deviation STD value, the three algorithms have been executed 30 times.

The obtained results, given in table (2.5), show that the best results are obtained with the ABC algorithm.

Table 2.5: The comparative study of the three optimization algorithms

item	ABC	DE	PSO
$I_{ph0}(A)$	6.73	6.71	6.73
$I_{sat}(A)$	1.38E-005	1.62E-005	1e-5
n	61.76	62.53	60.24
$R_s(\Omega)$	0.12	0.12	0.13
$R_{sh}(\Omega)$	103	120	95.50
RMSE	0.015	0.018	0.018
STD	8.46E-005	1.48E-004	3.90E-004

2.3.3.2. Identification results using the best-so-far ABC algorithm:

Hereafter, the effectiveness of PV module parameters identification based on the best-so-far ABC algorithm is experimentally assessed. Towards this end, real experimental (I-V) curve measurements from individuals PV modules are used. These modules are part of two installed grid connected PV systems, Algerian and Spanish PV systems. Moreover, these PV modules are of two different technologies: the Spanish PV module is "KANEKA GEA060" PV module, type amorphous silicon (a-Si) technology, while the Algerian PV module is "Isofoton 106/12" PV module, type mono-crystalline technology.

Table (2.6) summarizes the various irradiance and temperatures conditions at which the real measured (I-V) curves have been collected.

Table 2.6: The experimental weather conditions obtained from
outdoor measurements of (I-V) curves.

PV module	Temperature (°C)	Irradiance (W/m²)
Isofoton 106-12	27.2	755
KANEKA GEA060	25	1000

The electrical characteristics given in the manufacturer datasheet, and which are obtained under the standard test conditions (STC) of the Spanish PV module are summarized in table (2.7). Parameters variation limits of both Algerian and Spanish PV modules are summarized in table (2.8).

Table 2.7: The electrical parameters of
KANEKA GEA060 module at STC condition

Parameters	KANEKA GEA060
P_{mpp} (W)	60
V_{oc} (V)	92
I_{sc} (A)	1.19
V_{mpp} (V)	64
I_{mpp} (A)	0.90
αI_{sc} (%/°C)	-0.305
βV_{oc} (%/°C)	0.0752

Table 2.8: The PV module parameters' variation limits

	Isofoton 106-12	KANEKA GEA060
Iph [A]	[0-10]	[0-10]
I0 [A]	$[10^{-7}\text{-}10^{-4}]$	[0-1]
n	[0-75]	[0-300]
Rs [Ω]	[0-1]	[0-10]
Rsh[Ω]	[0-400]	[0-1000]

The adjustable parameters of the best-so-far ABC algorithm are set to the following values: food source number = 150, number of employed bees = 75,

number of onlooker bees = 75, the solution dimension = 5 (i.e., number of parameters to be identified) and the maximum cycle number = 7500. The optimal values of the parameters [I_{ph}, I_0, n, R_s, R_{sh}], determined by the best-so-far ABC algorithm, which gives the best fitting of the measured (I-V) curves, are summarized in table (2.9).

The optimal identified parameters are used to estimate the (I-V) curve by using the same irradiance and temperature conditions as the experiments. Then, the estimated (I-V) curves are compared with experimental (I-V) curves measurements to assess the parameter identification performance. Figures (2.7(a-b)) show the comparison between the real measured (I-V) curve and the estimated using the identified parameters, for the used PV modules under different meteorological conditions.

Table 2.9: The identification results

	Iph [A]	I0 [A]	n	Rs [Ω]	Rsh [Ω]	RMSE
Isofoton 106W-12V	6.54	1.11e-05	59.90	0.1474	202.6	0.014
KANEKA GEA060	1.28	6.15e-07	259.24	4.52	990.93	0.0058
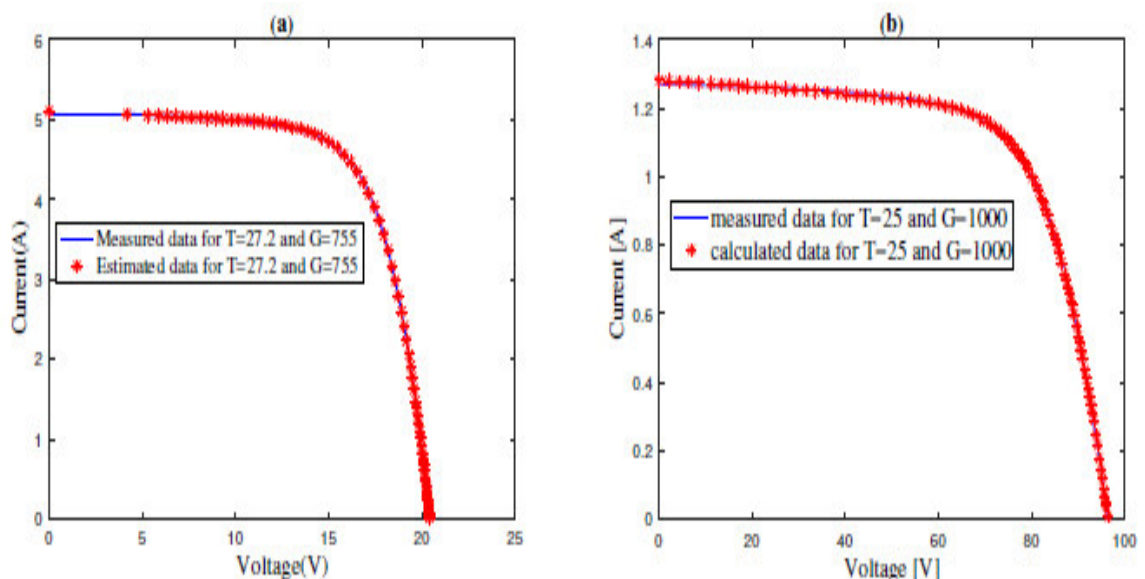
Figure 2.7: (a) Measured and predicted (I-V) curves for the Isofoton 106/12 PV module. (b) Measured and predicted (I-V) curve for the KANEKA GEA060 PV module.

It can be seen that the obtained parameters by the used optimization algorithm fit well with the experimental (I-V) curves, for both amorphous silicon and mono-crystalline technologies.

Furthermore, the lower value of the RMSE (see table (2.9)) indicates a good prediction performance of the best-so-far ABC algorithm to determine the best fitting parameters values, with an acceptable convergence rate, as shown in figures (2.8 (a-b)).
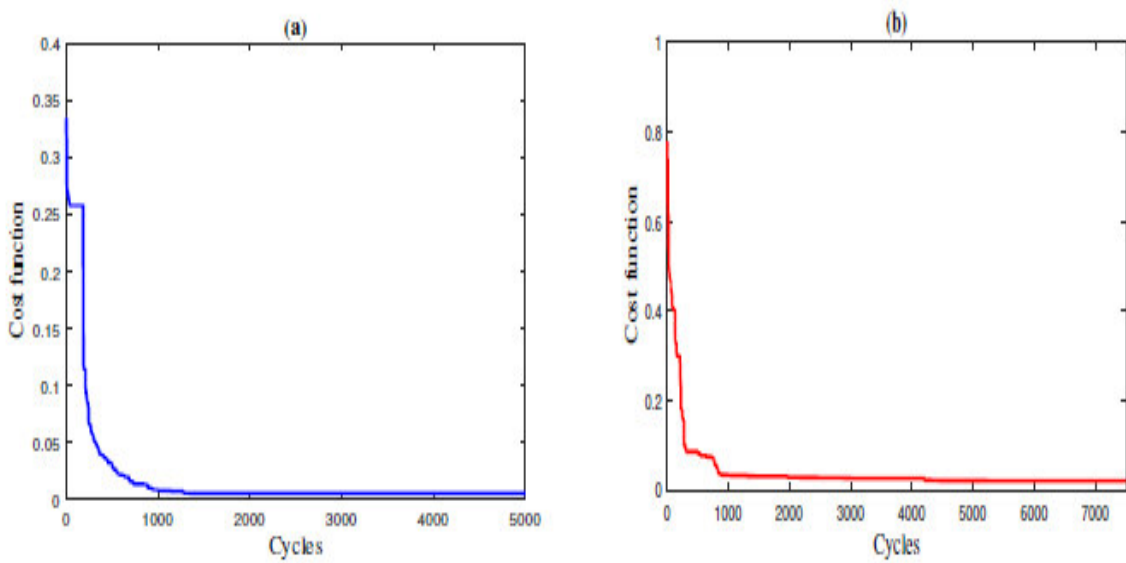
Figure 2.8: Convergence rate of the best-so-far ABC algorithm for (a) KANEKA GEA060 and (b) Isofoton 106-12V PV module.

2.4. The Developed approach of MPP estimation:

2.4.1. Fundamental principle:

Based on the identified parameters, the values of I_{mpp}, V_{mpp} and P_{mpp} can be estimated using equations (2.11-2.13) [54].

$$V_{mpp_0} = \frac{V_{mpp}}{1 + C_T(T_c - T_0)} + V_T \frac{T_0}{T_c} \ln\left(\frac{G_0}{G_{eff}}\right) - I_{mpp} R_s \left(\frac{G_0}{G_{eff}} - 1\right) \tag{2.11}$$

$$I_{mpp_0} = I_{mpp} \frac{G_0}{G_{eff}} \tag{2.12}$$

$$P_{mpp} = I_{mpp} \times V_{mpp} \tag{2.13}$$

where:

G_0 is the nominal irradiance (G_0=1000 W/m²), T_0 is the nominal temperature (T_0=25°C), R_s is the series resistance. G_{eff} is the effective irradiance, C_T is the temperature coefficient of power (typical value for monocristalin silicium is -0.0044 k^{-1}), T_c is the cell temperature, and it is given by the NOCT relationship, given hereafter:

$$T_c\left(G_{eff}, T_{amb}\right) = T_{amb} + \left(NOCT - T_{ambN}\right)\frac{G_{eff}}{G_N} \tag{2.14}$$

T_{amb} is the ambient temperature, $NOCT = 48°C$ is the Nominal Operating Cell Temperature and it is given in the PV module data-sheet, $G_N = 800 \; W/m^2$ and $T_{ambN} = 20 \; °C$ are the test conditions for the NOCT evaluation, V_T is the thermal voltage and it is expressed as follows:

$$V_T = \frac{nkT}{q} \tag{2.15}$$

V_{mpp_0} and I_{mpp_0} are the nominal maximum voltage and current respectively (the maximum power point coordinates), and they are given within the PV module data-sheet.

2.4.2. MPP estimation results:

2.4.2.1. MPP estimation results using the ABC-based identified parameters:

In order to assess the accuracy of the ABC-based identified parameters, MPP model given in equations (2.11-2.13) has been plotted against real daily production for two conditions: a clear sky condition and cloudy sky condition, of an actual Algerian PV system. This system is formed by two parallel strings; each string is composed of fifteen series connected Isofoton 106-12 PV modules.

Figures (2.9-2.11) show the time evolution of the simulated and measured current, voltage and power, respectively, for a clear sky day profile. While, time evolution of the same quantities for a cloudy day profile is shown in figures (2.12-2.14). These results show clearly the efficiency of the proposed MPP estimation strategy for both clear sky and cloudy sky conditions.
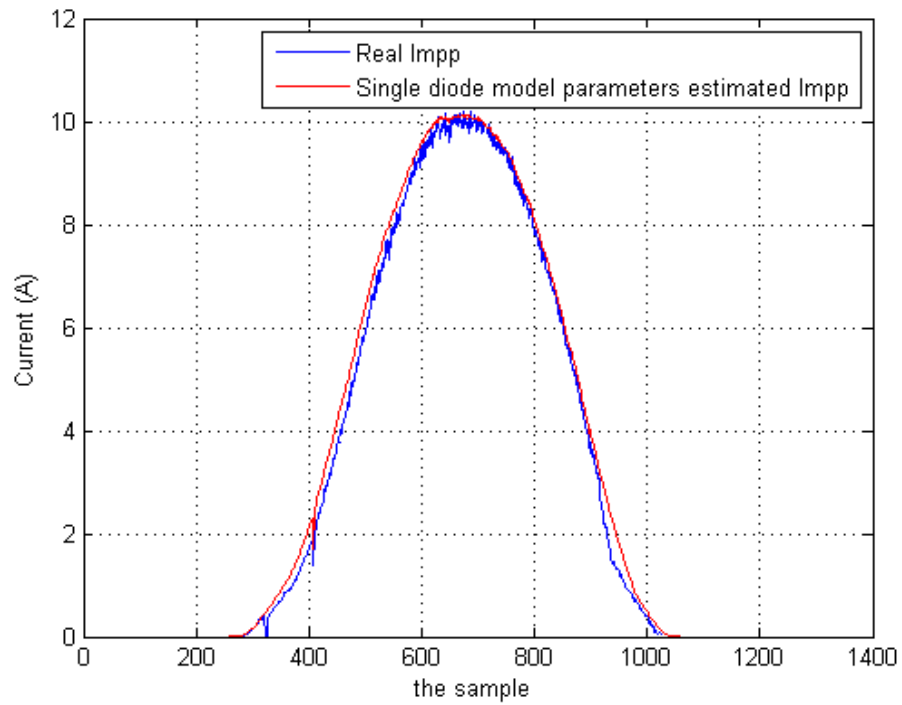
Figure 2.9: Comparison between measured and estimated I_{mpp}

for a clear sky day profile.
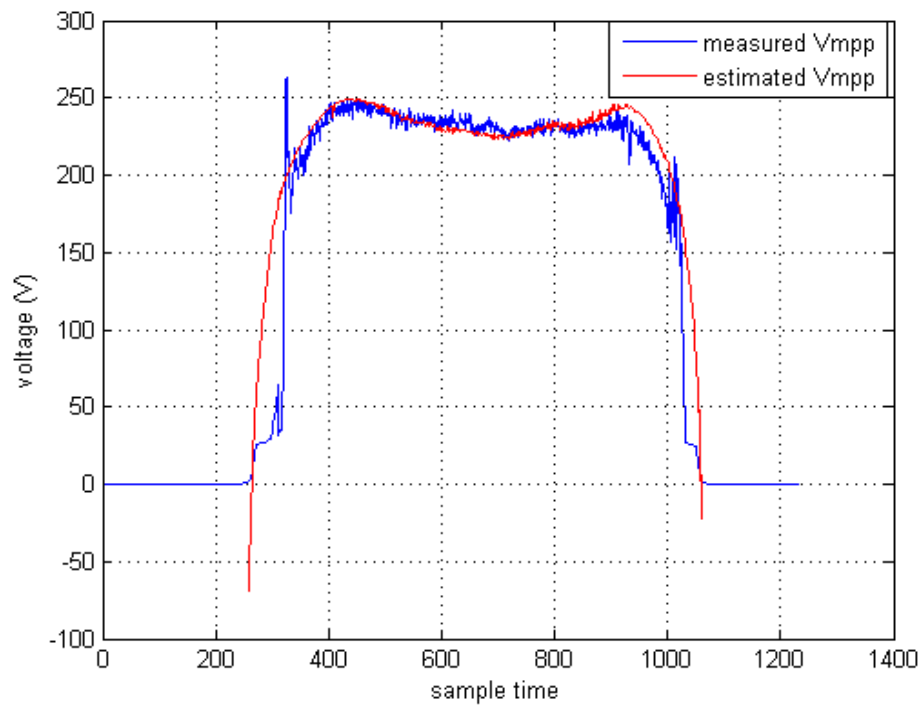
Figure 2.10: Comparaison between measured and estimated V_{mpp}

for clear sky day profile.
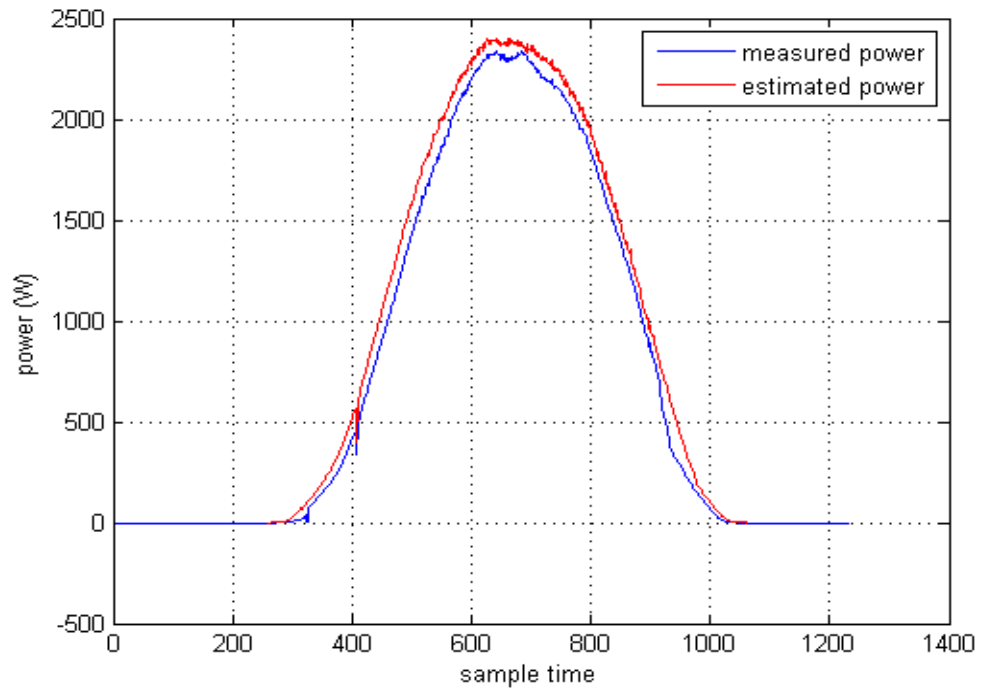
Figure 2.11: Comparaison between measured and estimated P_{mpp} for clear sky day profile.
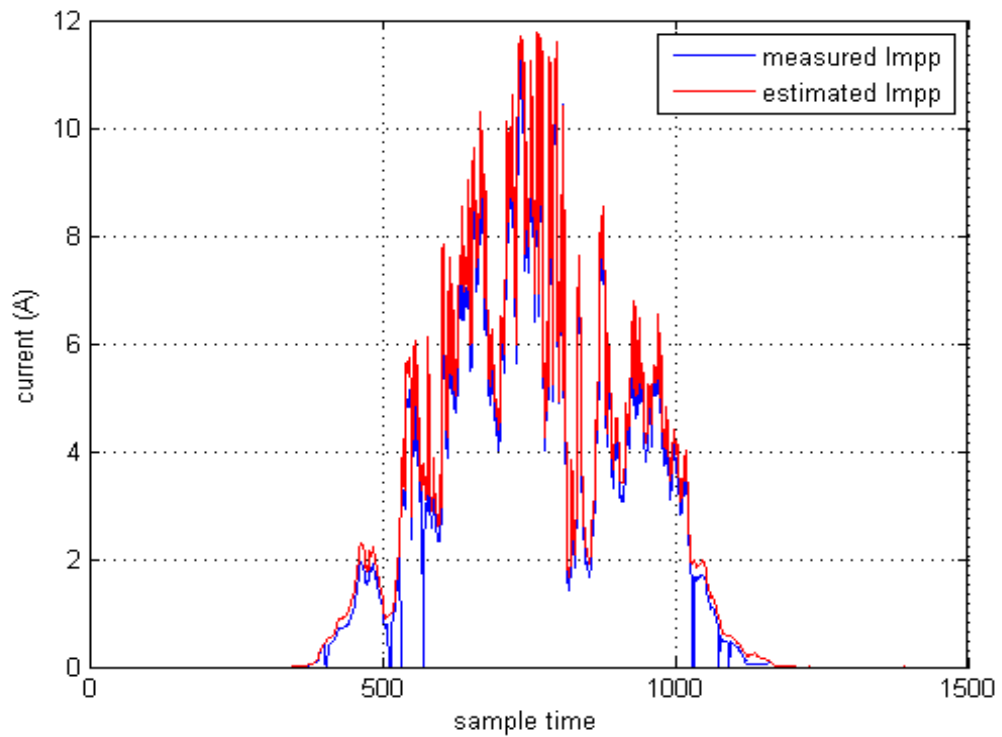
Figure 2.12: Comparaison between measured and estimated I_{mpp} for a cloudy sky day profile.
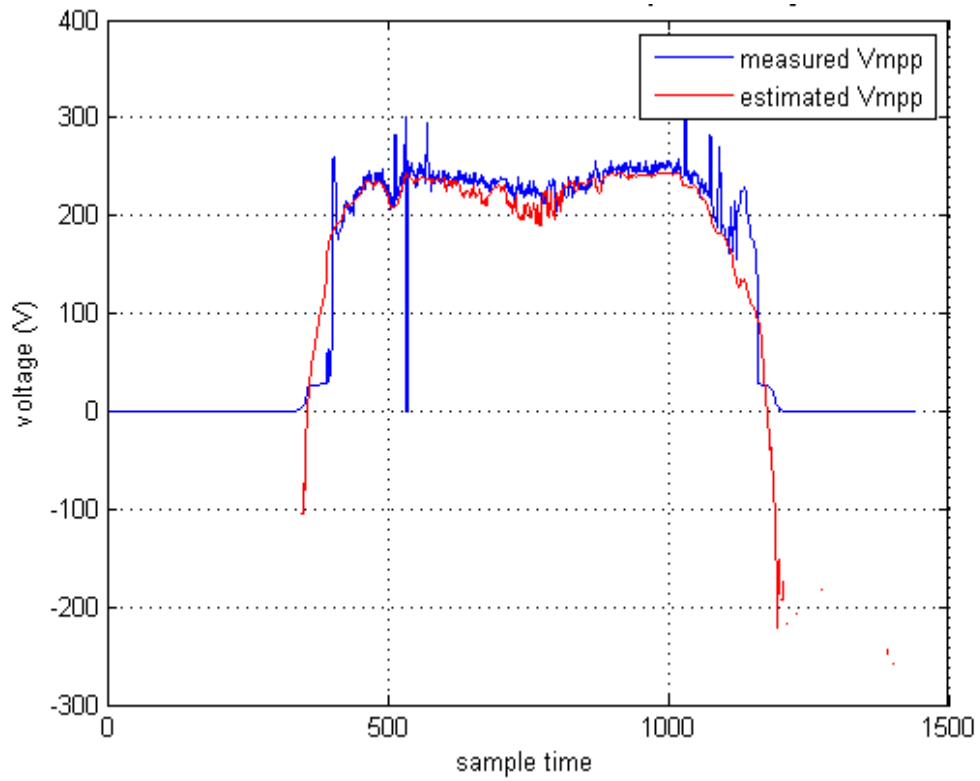
Figure 2.13: Comparaison between measured and estimated V_{mpp}
for a cloudy sky day profile.
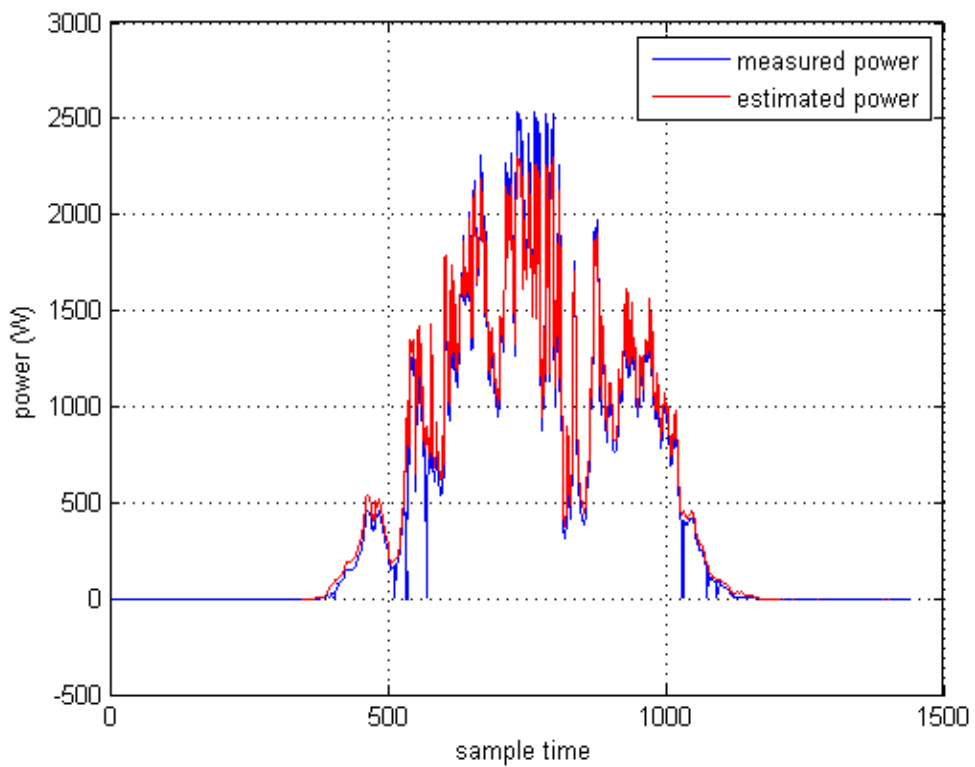
Figure 2.14: Comparaison between measured and estimated P_{mpp}
for a cloudy day profile.

2.4.2.2. <u>MPP estimation results using the best-so-far ABC-based identified parameters:</u>

The identified parameters based on the best-so-far ABC algorithm are used to estimate the dynamic evolution of MPP coordinates of current and power under real operating conditions for the Algerian and Spanish PV systems.

The dynamic MPP estimation approach has been validated experimentally using real daily weather conditions collected from two sites (Algiers province of Algeria and Jaén province of Spain). To assess the proposed MPP estimation approach under different operating conditions, two daily profiles are also selected for both the Algerian and Spanish sites, clear and cloudy sky conditions (see Figures (2.15-2.16)). The measured versus estimated MPP current and power of the Algerian GCPV system under clear sky day condition are shown in Figures (2.17(a-b)), and under cloudy sky day condition are shown in Figures (2.17(c-d)). For Jaén PV array, the measured and estimated MPP current and power under clear sky day condition are presented in Figure (2.18(a-b)), and under cloudy sky day condition are shown in Figure (2.18(c-d)). These Figures show that the observed data are well-fitted by the MPP estimation strategy for the two studied systems.
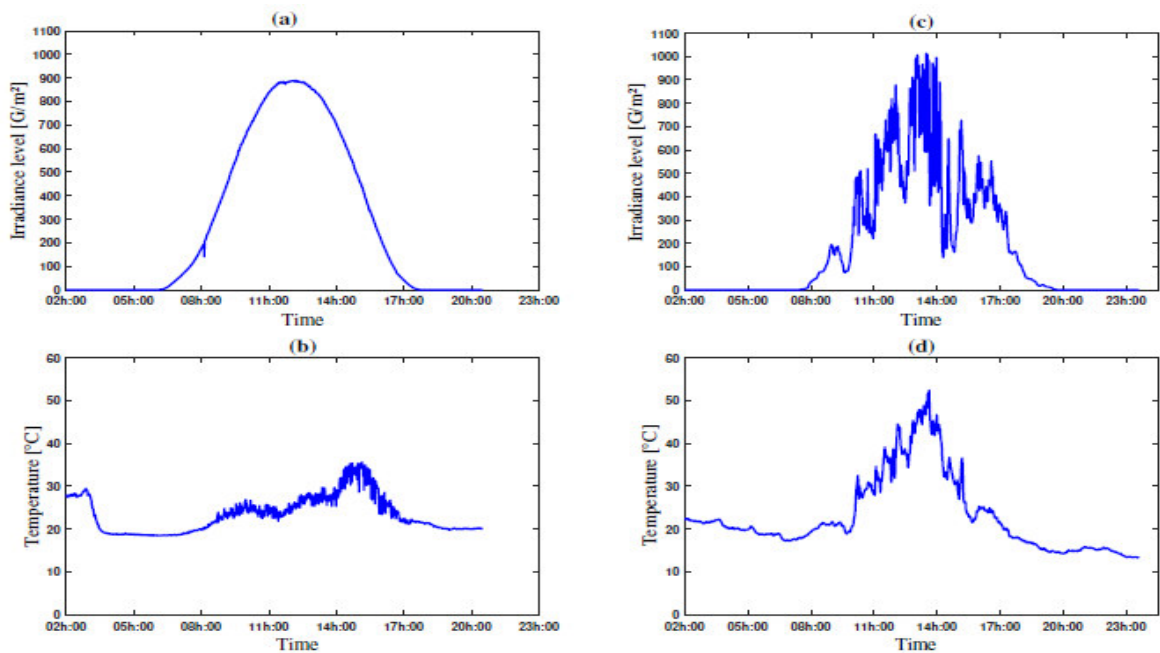
Figure 2.15: The Algerian PV system (a) clear day profile of irradiation level, (b) clear day profile of temperature level, (c) cloudy day profile of irradiation level and (d) cloudy day profile of temperature level.
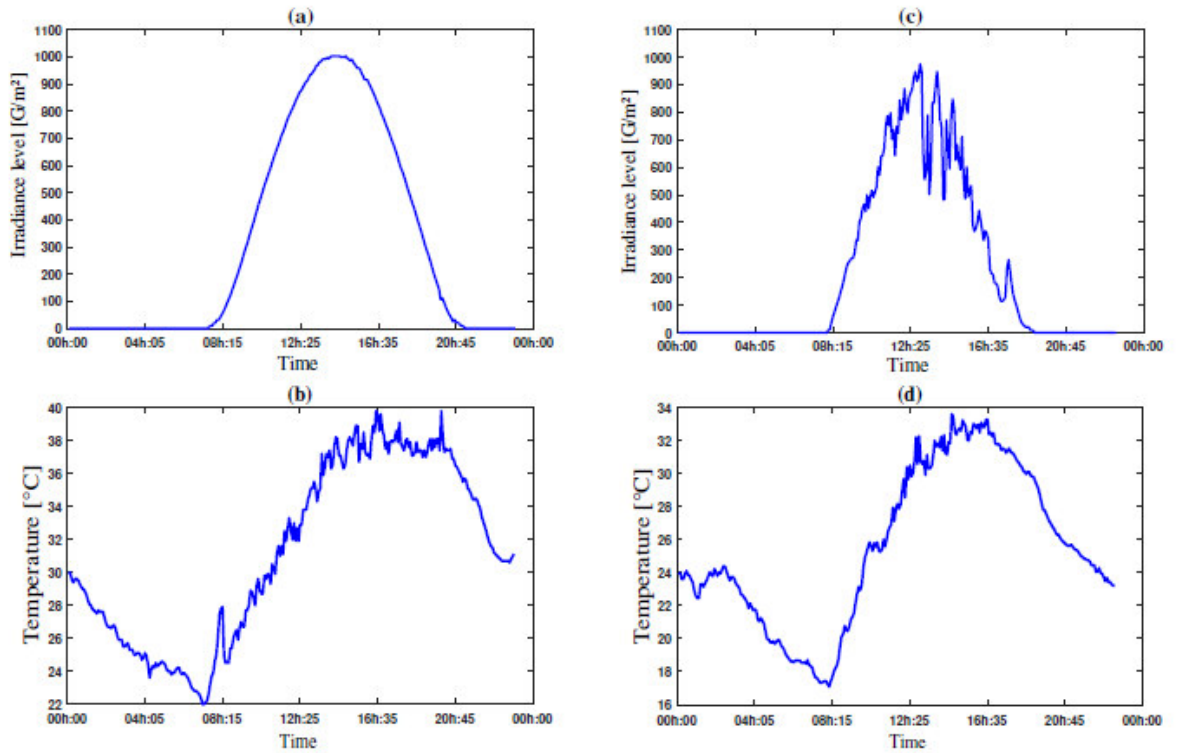
Figure 2.16: the Spanish PV system (a) clear day profile of irradiation level, (b) clear day profile of temperature level, (c) cloudy day profile of irradiation level and (d) cloudy day profile of temperature level.
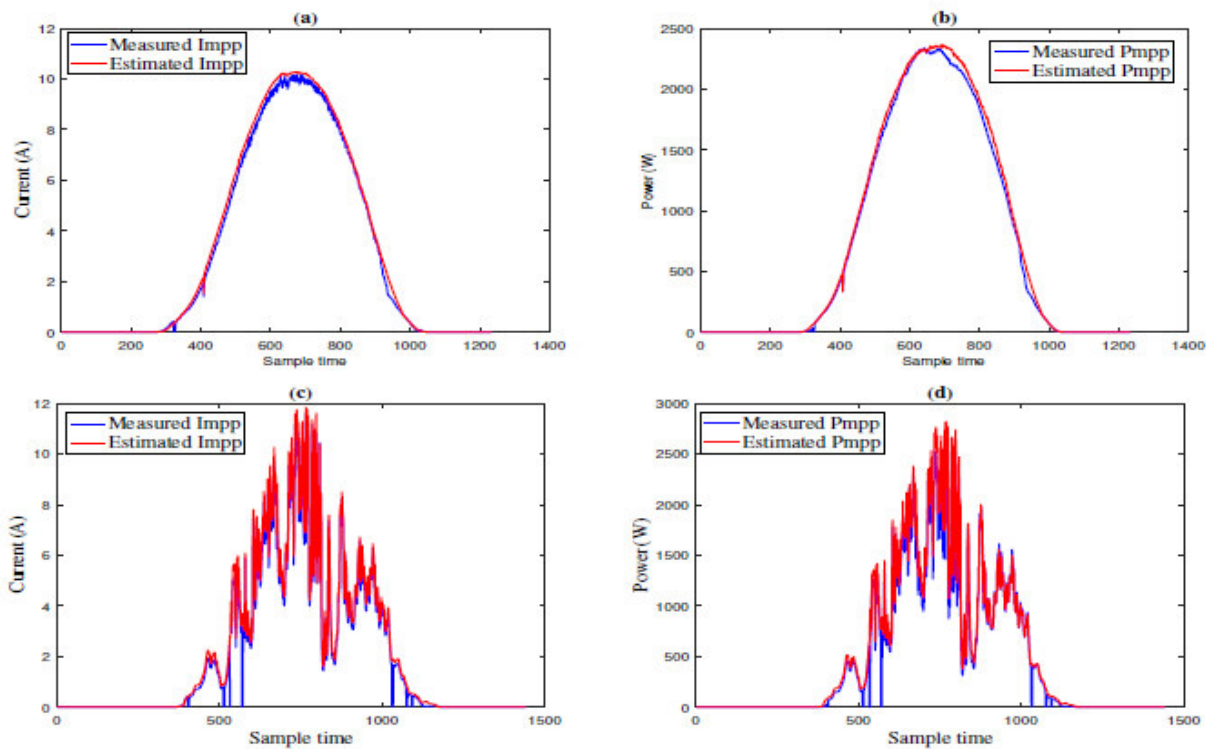
Figure 2.17: Measured versus estimated MPP current and power under clear sky day condition (a-b) and cloudy sky day condition (c-d) of the Algerian PV system.
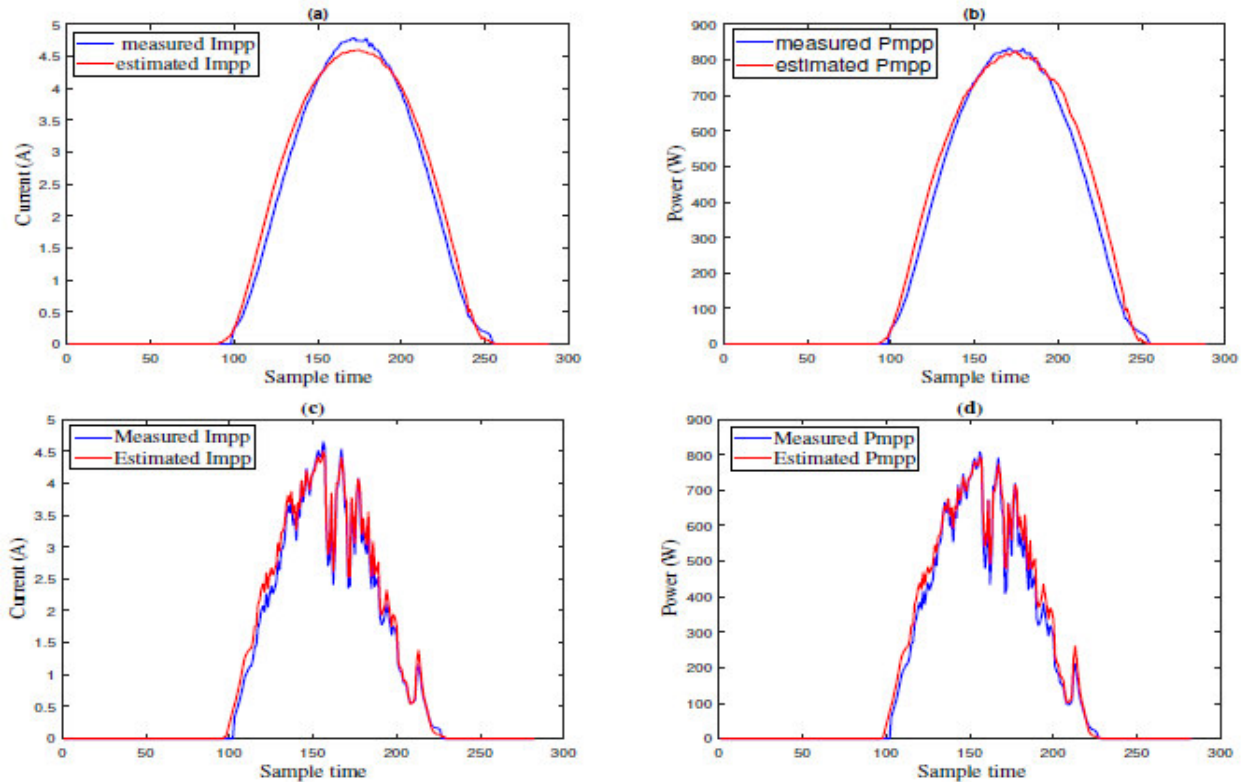
Figure 2.18: Measured versus estimated MPP current and power under clear sky day condition (a-b) and cloudy sky day condition (c-d) of the Spanish PV system.

2.5. Conclusion:

This chapter is mainly devoted to clarify the developed strategy of PV array modeling and validation. It reports the ODM description, the PV module parameters identification using ABC and best-so-far ABC algorithms, and finally the proposed approach to estimate the MPP coordinates. .

The proposed parameters identification algorithms have been tested using several PV modules of different technologies. The identified parameters of each module have been introduced into the characteristic equation, and then tested against real (I-V) curves measurements. The obtained results show clearly the effectiveness of the proposed algorithms to extract the PV module parameters.

Finally, an efficient approach of MPP estimation, based on the identified parameters, has been developed. MPP estimation results have clearly demonstrated the effectiveness of this strategy.

CHAPTER 3

FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING PROBABILISTIC NEURAL NETWORKS

3.1. Introduction:

In this chapter, a novel procedure for fault detection and diagnosis in the DC side of PV systems, based on the probabilistic neural network (PNN) [9, 10] classifier, is proposed. The suggested procedure consists of four main stages: (i) PV module parameters extraction, (ii) PV array simulation and experimental validation (iii) the elaboration of a relevant database of both healthy and faulty operations, and (iiii) network construction, training and testing. In the first stage, the unknown electrical parameters of the one diode model (ODM) will be accurately identified using the best-so-far ABC algorithm. Then, based on these parameters the PV array will be simulated and experimentally validated by using a PSIM^{TM}/Matlab^{TM} co-simulation. Finally, an efficient fault detection and diagnosis procedure based on PNN classifier will be implemented. Four operating cases are tested in a GCPV system of 9.54 kWp: Healthy system, three modules short-circuited in one string, ten modules short-circuited in one string, and a string disconnected from the array. Moreover, the PNN method will be compared, under real operating conditions, with the feed forward back-propagation ANN classifiers method, for noiseless and noisy data.

3.2. Description of the PV system and the faults detection and diagnosis strategy:

3.2.1. PV system description:

PV plant under study is an actual Algerian GCPV system, located in Algiers (Algeria) (Latitude: 36°43'N, Longitude: 3°15'E). This PV plant is a 9.54 kWp arranged in 3 sub-arrays of 3.18 kWp each one. Each sub-array contains 30 Isofoton 106W-12V PV modules, mounted in 2 parallel strings of 15 PV modules in series, and connected to a 2.5 kW (IG30 Fronius) single phase inverter. Tilted and horizontal irradiances are measured using a Kipp & Zonen CM11 thermoelectric pyranometer, while the PV module temperature is measured using K-type thermocouple.

Meteorological variables as well as electrical variables measurements are performed by a data logger (Agilent 34970) as depicted in Figure (3.1).
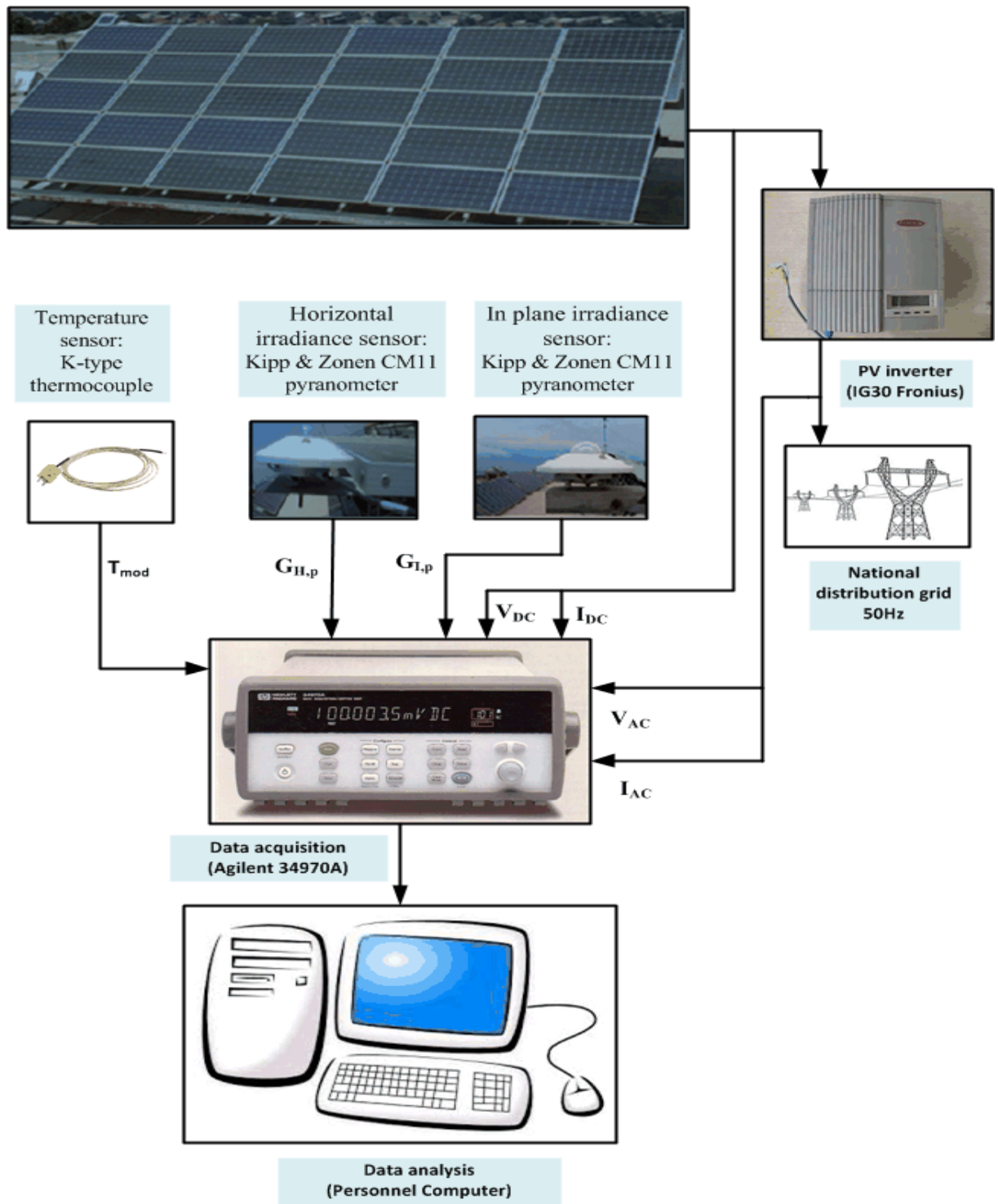
Figure 3.1: PV plant under study and the monitoring system.

3.2.2. Faults detection and diagnosis strategy:

The main objective of this work is to design an efficient and reliable procedure, based on PNN classifier, to detect faults in a photovoltaic system and to diagnose their origin. However, using this type of method, to deal with the classification problem, requires the availability of a high-quality database that describes very well the process for each class. Practically speaking, obtaining such a database cannot often be guaranteed, especially in PV systems. In fact, operating a PV system under some types of failures can make the system completely insecure and cause catastrophic damages and safety hazards. Therefore, the best way to deal with this concern is to have a trusted simulation model that mimics the actual behavior of a PV system under healthy and faulty states. The flowchart, given in Figure (3.2), summarizes the adopted steps followed to construct the fault detection and diagnosis strategy, namely: PV module parameters extraction, model validation, database elaboration and finally fault detection and diagnosis.

3.2.2.1. PV module parameters extraction:

In this work, the same strategy of PV module parameters extraction, which has been developed in chapter 2 based on the best-so-far ABC algorithm, is used to extract the ODM five electrical parameters. This choice is essentially due to its capacity to converge regardless the initial conditions, to accurate global solutions, high convergence speed and its simplicity in terms of implementation [21, 55, 56].

3.2.2.2. Model validation:

The extracted ODM parameters are, subsequently, used to simulate the real PV system for normal conditions using PSIM^{TM}/Matlab^{TM} Co-simulation. Indeed, the physical model of solar module, included in PSIM^{TM} software, is used to simulate the physical behavior of the actual PV system. This can be reached by introducing the ODM identified parameters into this model, and then simulate the whole PV system for real daily profile measurements of temperatures and irradiances. On the other hand, data processing and involved calculations are performed in Matlab^{TM}/Simulink^{TM} environment. Finally, the simulated power (P_{sim}) and the measured one (P_{meas}) are then compared.
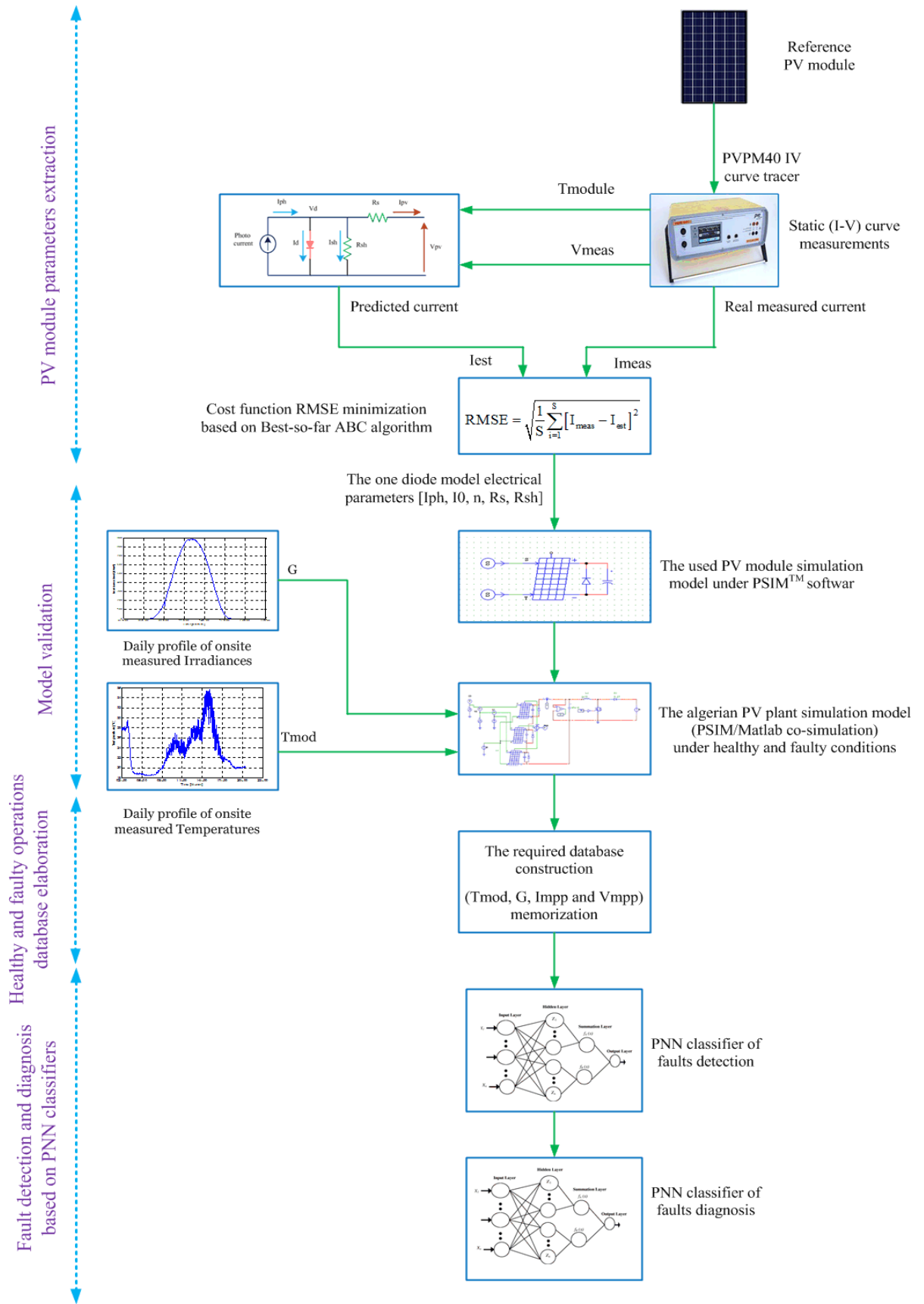
Figure 3.2: Flowchart of the followed steps of the proposed faults detection and diagnosis strategy.

3.2.2.3. <u>Database elaboration:</u>

This stage consists of elaborating a high-quality database that accurately describes the system behavior, in normal and faulty conditions. To this end, the validated PV system model is used to generate healthy and faulty samples, by intentionally introducing the desired faults, with real daily profiles of temperature and irradiance. The recorded samples, corresponding to each operating condition, include four attributes: module Temperature "T", tilted Irradiance "G", current at MPP "I_{mpp}" and voltage at MPP "V_{mpp}".

3.2.2.4. <u>Fault detection and diagnosis based on PNN classifier:</u>

Based on the elaborated data base, the last step consists of constructing two probabilistic neural networks (PNN) classifiers: the first one is dedicated to fault detection and the second is responsible for diagnosing the origin of faults. The methodology of PNN construction, training and test are explained in details hereafter.

3.3. <u>The proposed neuronal strategy of faults detection and diagnosis:</u>

3.3.1. <u>Probabilistic neural network:</u>

The Probabilistic Neural Network (PNN), as a learning based method, has been considered as a powerful classification technique. Similarly to the other neural network strategies, the PNN uses a training set to extract pattern statistics, and a testing set to check the classification accuracy. Its architecture is similar to that of the back propagation one [9], except of using the exponential function, as an activation function, instead of using the sigmoid function [9]. As shown in Figure (3.3), the probabilistic neural network consists of: input units, pattern units, summation units and output units [9, 10].

The aim of the input units is to directly distribute the input vector X to the pattern units, while the number of neurons in these units corresponds to the input vector size.

Besides, the pattern units, described by Figure (3.4), consider the use of the same number of neurons as the input units. The pattern units are responsible for the dot multiplication of the elements of the input vector and their respective weights $Z_i = X \bullet W_i$, given that the weight coefficient W_i of each pattern unit is set to

the same value X_i of the training set. The result, Z_i , will be then introduced to a nonlinear activation function, given by the following equation [9]:

$$Q(Z_i) = \exp\left(\frac{(Z_i - 1)}{\sigma^2}\right)$$

(3.1)

In the case of normalized input and weight vectors, the activation function (equation (3.1)) will result in [9]:

$$Q(X) = \exp\left(\frac{-(W_i - X)^t (W_i - X)}{(2\sigma^2)}\right)$$

(3.2)

where σ denotes the smoothing parameter.

The outputs of the pattern units will be then transmitted to the summation units, in which the number of neurons matches the number of classes. The outputs of each summation unit are given by [9]:

$$f_A(X) = \frac{1}{(2\pi)^{\frac{P}{2}}\sigma^P} \frac{1}{m} \sum_{i=1}^{m} \exp\left[\frac{-(W_i - X_{Ai})^t (W_i - X_{Ai})}{2\sigma^2}\right]$$

(3.3)

where: $f_A(X)$ is the Probability Density Function (PDF) of the input vector X; P is the number of patterns; m is the learning set size; X_{Ai} describes the corresponding i^{th} training pattern of class A. The output units define the decision blocks. PNN comprises only one neuron in its output layer. This layer receives, from the summation layer, the probability density functions (PDF) values of each class and predicts the adequate class of the new sample. In other terms, this last layer outputs the adequate target that corresponds to the highest probability density function in the summation units. The efficiency of PNN classifier depends on the used PDF accuracy, which is defined by the best choice of its smoothing parameter value [9].

Compared with the conventional back-propagation network, PNN has several important advantages such as [9]:

- Its training stage requires only one single pass (neither iterations nor weights computation).
- It can support both erroneous and noisy samples.
- It has only one adjustable parameter (the smoothing parameter σ).
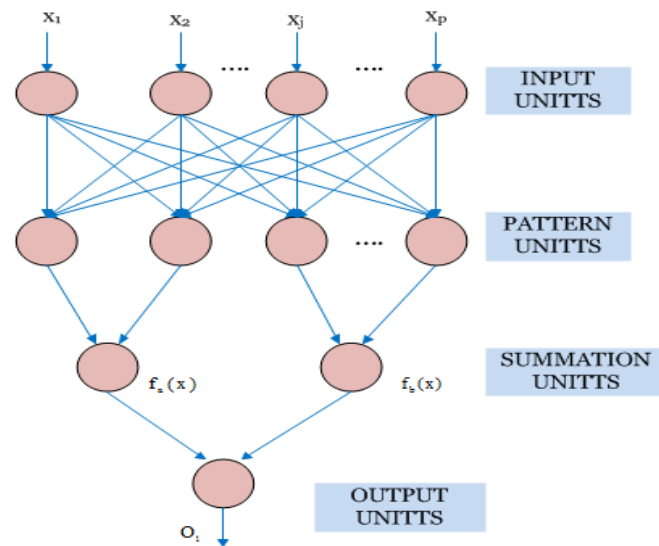
Figure 3.3: PNN structure for classification problem of two dimensions [9].
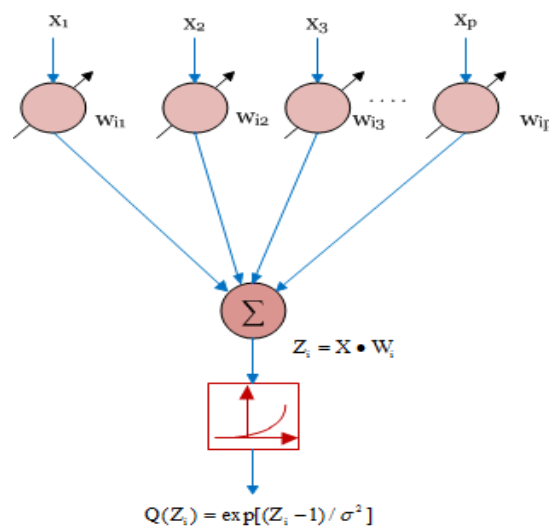
Figure 3.4: PNN Pattern unit [9].

3.3.2. PNN based method for fault detection and diagnosis:

As stated before, the objective of this work is to detect and diagnose faults in the DC side of a GCPV system based on PNN classifier. Using PNN as fault detection and diagnosis approach requires four essential steps: (i) the elaboration of a relevant database that contains the necessary data to train and test the network, (ii) the network construction, (iii) train the network using the learning set, and finally (iiii) the test and evaluation of the network efficiency based on the test set.

3.3.2.1. The elaboration of the relevant database:

To be able to obtain a pertinent database for both healthy and faulty systems, the physical model of the PV module implemented in PSIM^{TM} software, is used to simulate, under healthy and faulty conditions, the real operation of the previously described Algerian GCPV system. The employed PV array comprises two parallel strings of fifteen PV modules in series (Isofoton 106W-12V).

During this simulation, the physical model of the PV array requires the introduction of the ODM parameters' values, which are obtained using the best-so-far ABC algorithm. This can be achieved by using the developed parameters extraction strategy of chapter 2.

The proposed strategy for simulating the photovoltaic system using the PSIM^{TM} software is characterized by its physical nature, which makes it possible to easily simulate the behavior of several faults that usually happen in the DC side of PV systems, such as: shading faults, short circuit faults, open circuit faults, line-line faults, …etc.

Moreover, the used simulation model requires much reduced time to simulate the behavior of PV system whatever its physical configuration, and even for a very high number of meteorological data.

In this work, the simulations of four operating cases have been considered: healthy system; three modules short-circuited in a string; ten modules short-circuited in a string, and finally a completely disconnected string in an array, as depicted in Figure (3.5). To finish with, the final database is constructed considering four attributes [T, G, I$_{mpp}$, and V$_{mpp}$] for each operating case. Figure (3.6) shows the flowchart of PNN dataset elaboration.

3.3.2.2. The networks construction:

Since the proposed work aims both fault detection and diagnosis, two PNN classifiers are then constructed. Indeed, the occurrence of any fault in the PV system is detected by the first network, while the diagnosis of its type is carried out by the second one.

The detection network, shown in Figure (3.7), has four inputs (T, G, I$_{mpp}$ and V$_{mpp}$) and two outputs (healthy state, faulty state), while the diagnosis network, shown in Figure (3.8), has the same inputs as the first one, and three outputs (fault #1, fault #2 and fault #3) corresponding to the three faulty operating cases.

PNN classifiers of fault detection and diagnosis (Figures (3.7-3.8)) comprise only one neuron in their output layers. This neuron outputs the class that corresponds to the highest probability density function in the summation units.

In order to avoid any conflict, the detection and diagnostic networks are connected in series so that the diagnostic network will not be activated as long as the faulty alarm does not occur from the detection network. This concept is explicitly explained by the flowchart of Figure (3.9), given below.
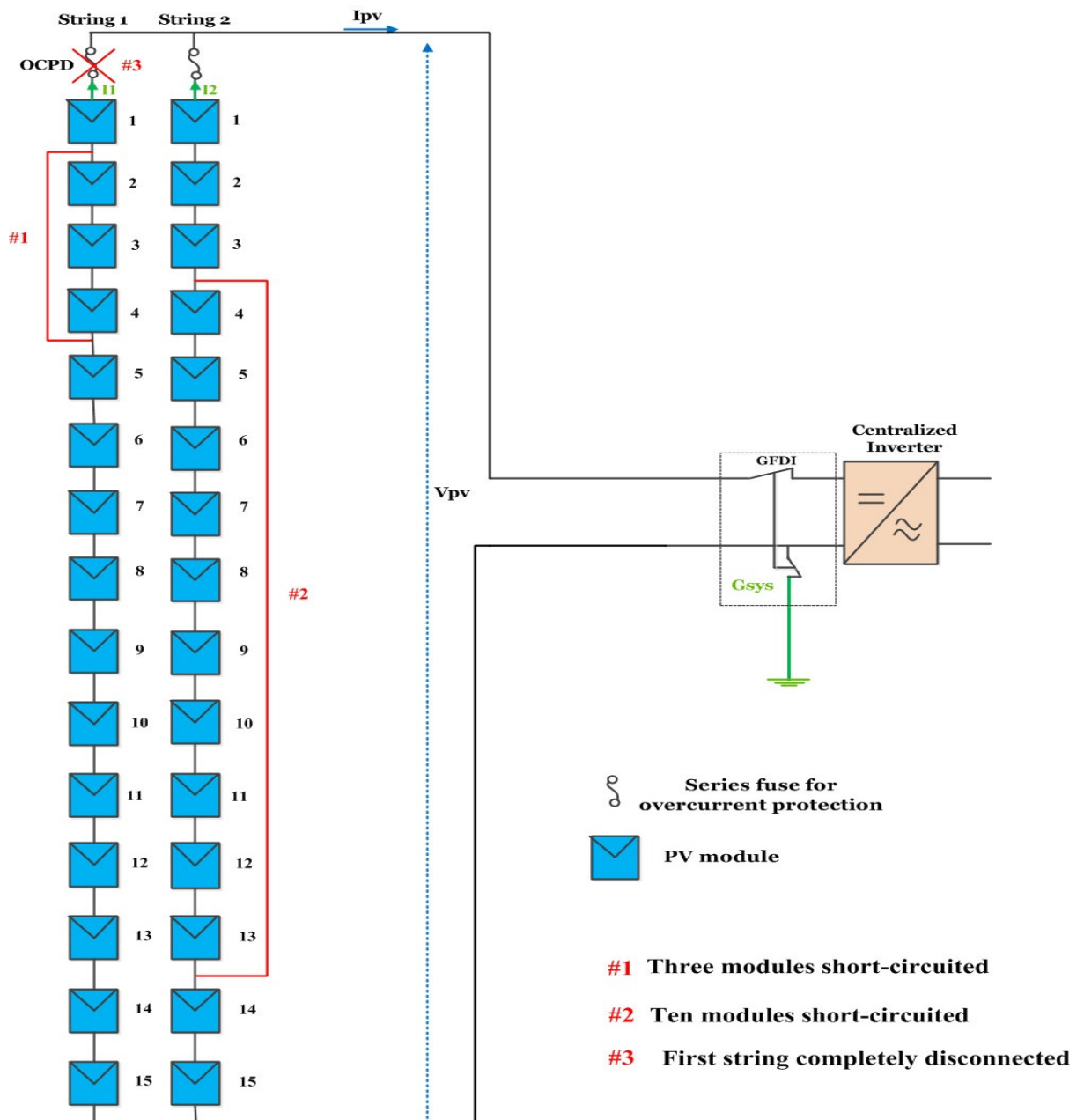
Figure 3.5: The tested open circuit and short circuit failures
of the studied PV system.
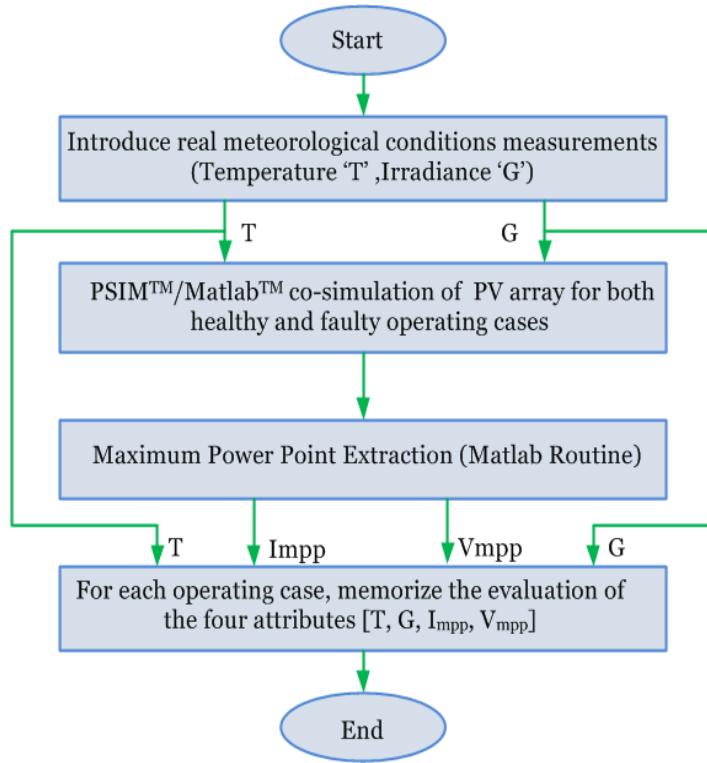
Figure 3.6: The dataset elaboration flowchart.
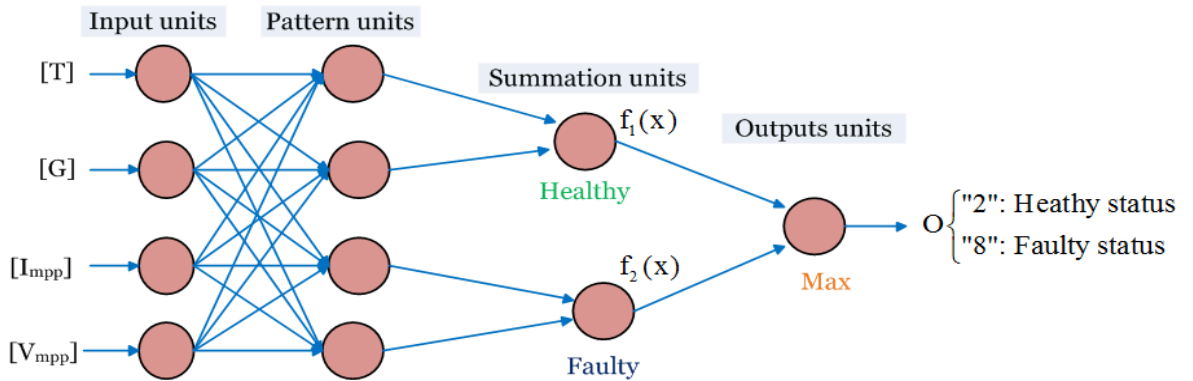
Figure 3.7: PNN detection network.
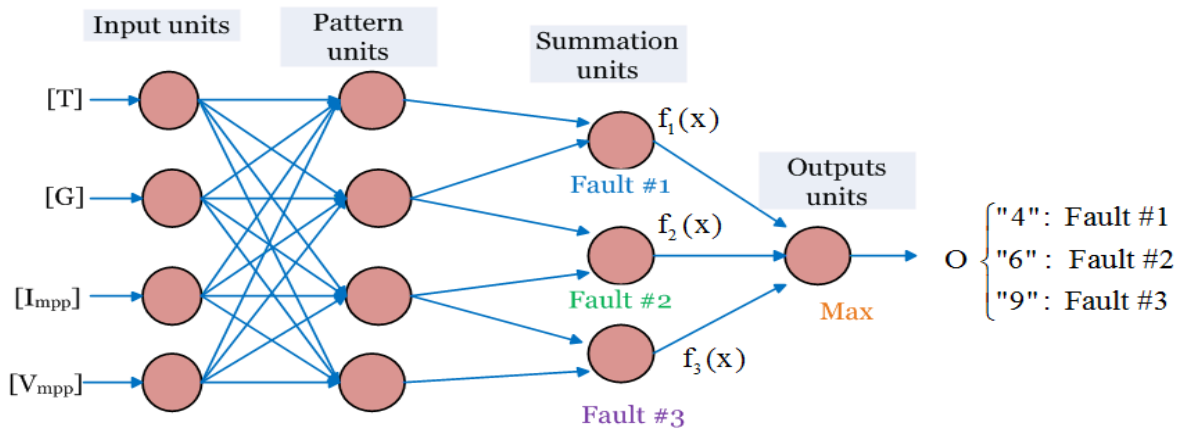
Figure 3.8: PNN diagnosis network.
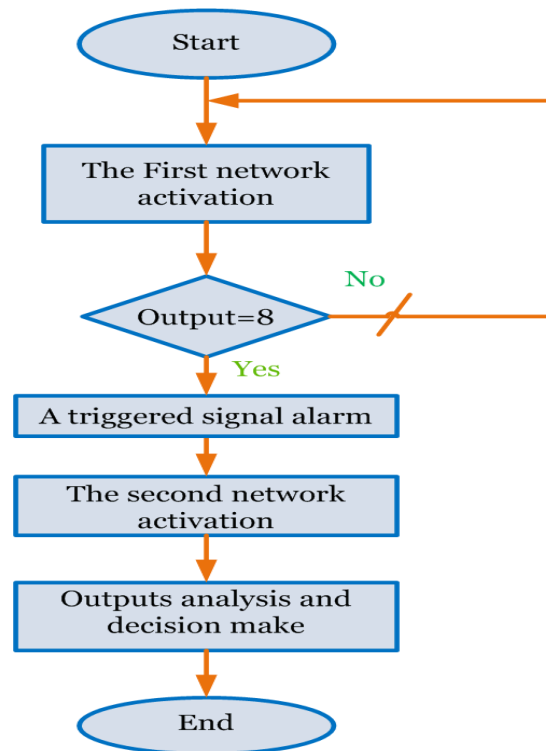
Figure 3.9: Flowchart of the two classifiers' connection strategy.

3.3.2.3. The learning stage:

Now, the detection and diagnostic networks are formed by using the learning data set previously obtained. It should be noted here that, on one hand, the input data (T, G, I_{mpp} and V_{mpp}), are of continuous type which are directly processed by the two PNNs. On the other hand, the outputs (healthy state, faulty state, fault #1, fault #2 and fault #3) are nominal variables that are not supported by this type of networks. For this reason, particular codes, arbitrary chosen, have been assigned to the PNN outputs as mentioned below:

- Healthy system: code "2", and which will be used for the faults detection network to describe the healthy operating case.
- Faulty system: code "8", and which will be used for the faults detection network to describe the occurrence of possible DC side faults.
- Fault #1: code "4", and which will be used for the diagnosis network to describe the occurrence of short circuit of three PV modules in one string.
- Fault #2: code "6", and which will be used for the diagnosis network to describe the occurrence of short circuit of ten PV modules in one string.

- Fault #3: code "9", and which will be used for the diagnosis network to describe the occurrence of a complete string disconnection from the PV array.

Finally, to train both detection and diagnosis networks, learning set values and the codes of their corresponding classes are used as input and target vectors, respectively.

3.3.2.4. The testing stage

After having learned both detection and diagnosis networks, the next stage is to test their effectiveness while using the testing set. Similarly to the training stage, the input testing set to the two PNN classifiers are the four attributes (T, G, I_{mpp} and V_{mpp}), while their outputs denote their corresponding estimated classes.

In order to assess the networks efficiency, a performance test of the classifiers has been carried out by using the following four standard metrics [57]:

$$\text{Classification Accuracy} = 100 \times \frac{\text{TP}_{\text{all-classes}}}{\text{d}} \qquad (3.4)$$

$$\text{Sensitivity(class(x))} = 100 \times \frac{\text{TP}_{\text{class(x)}}}{\text{TP}_{\text{class(x)}} + \text{FN}_{\text{class(x)}}} \qquad (3.5)$$

$$\text{Specificity(class(x))} = 100 \times \frac{\text{TN}_{\text{class(x)}}}{\text{TN}_{\text{class(x)}} + \text{FP}_{\text{class(x)}}} \qquad (3.6)$$

$$\text{Positive Predictivity(class(x))} = 100 \times \frac{\text{TP}_{\text{class(x)}}}{\text{TP}_{\text{class(x)}} + \text{FP}_{\text{class(x)}}} \qquad (3.7)$$

where:

TP is the True Positive classifications number, which denotes the number of samples which are supposed to be classified in class "x" and they are really classified into this class. FN is the False Negative classifications number, which denotes the number of samples from class "x" and are classified into another class than class "x". TN is the True Negative classifications number, which denotes the number of samples which are not supposed to be classified in class "x" and are really classified, according to the classifier, into another class than class "x". FP is the False Positive classifications number, which denotes the number of samples which are not supposed to be classified into class "x", while they really

are classified, according to the classifiers, into this class, and 'd' defines the size of the testing set.

In addition to these performance evaluation standard metrics, the confusion matrix is another metric usually used to evaluate the classifier performance. In fact, confusion matrix (also called contingency table) is a clean and unambiguous way to depict the classification results. As an example of a binary classification problem (Problem with two possible classes), the confusion matrix is composed of two rows and two columns, as can be seen in table (3.1).

Table 3.1: The confusion matrix of a binary classification problem

	The first class (Predicted)	**The second class (Predicted)**
The first class (Actual)	True Positive (TP)	False Negative (FN)
The second class (Actual)	False Positive (FP)	True Negative (TN)

3.4. Simulation and experimental results:

In order to verify the performance and effectiveness of the proposed approach, described previously in section (3.3.2), a simulation study and an experimental evaluation are carried out, and the main results are highlighted. This section is subdivided into three sub-sections: PV module parameters extraction results, model validation results, and finally, PNN based fault detection and diagnosis results.

3.4.1. PV module parameters extraction results:

In the first stage, the efficiency of PV module parameters extraction method based on the best-so-far ABC algorithm was experimentally validated. To this end, experimental (I-V) curve measurements, using an I-V curve tracer (PVPM40), were recorded and compared with the prediction model (Equation (2.1). the module temperature and the irradiance level used in this particular test were 27.2 °C and 755 W/m^2, respectively.

The best-so-far ABC algorithm adjustable parameters are summarized in Table (3.2), while the optimal values of the electrical parameters [I_{ph}, I_0, n, R_s, R_{sh}], determined by the best-so-far ABC algorithm, are summarized in Table (3.3).

Table 3.2: The adjustable parameters of the best-so-far ABC algorithm

The adjustable Parameters	Food source number	Number of employed bees	Number of onlooker bees	The solution dimension	Maximum cycles number
The selected value	150	75	75	5	2000

Table 3.3: Isofoton 106W-12V PV module's extracted parameters

The electrical parameter	I_{ph} [A]	I_0 [A]	n	Rs [Ω]	Rsh [Ω]	RMSE
The identified value	6.54	1.11e-05	1.66	0.1474	202.6	0.014

The simulation result of the prediction model (Equ.3.2), obtained with the optimal parameters, and the experimental curves are shown in Figure (3.10).

It is clearly shown the good agreement between the experimental curve and the predicted one when simulated with the extracted parameters.

Also, the smallest value of the RMSE, given in table (3.3), reveals the high accuracy of the best-so-far ABC algorithm to extract the best fitting parameters.

In Figure (3.11), it is shown the best-so-far ABC algorithm convergence rate for each cycle until reaching the convergence condition (smallest RMSE value). While in Figure (3.12), it is depicted the absolute error between measured current (I_{meas}) and estimated current (I_{est}), computed by the following equation.

$$current\_mismatch = \left| I_{meas} - I_{est} \right| \tag{3.8}$$

According to Figure (3.12), the highest value of the absolute error does not exceed 0.06 which confirm the high accuracy of the best-so-far ABC algorithm to extract the best fitting parameters of the ODM.
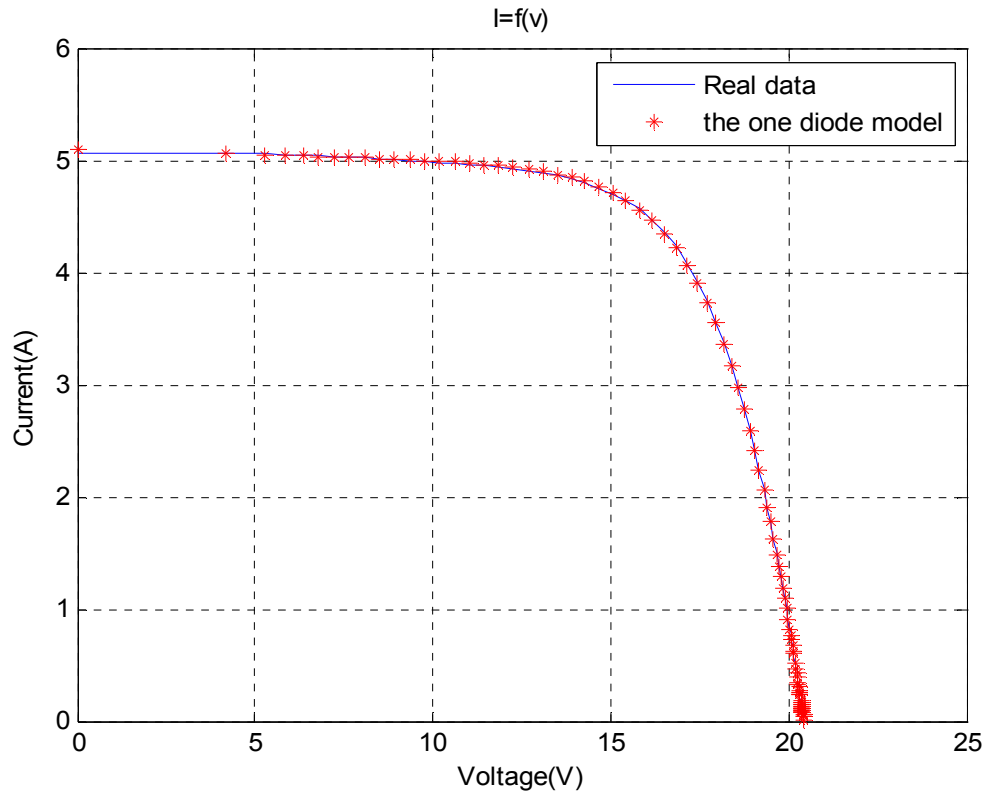
Figure 3.10: The real measured (I-V) curve against the estimated one.
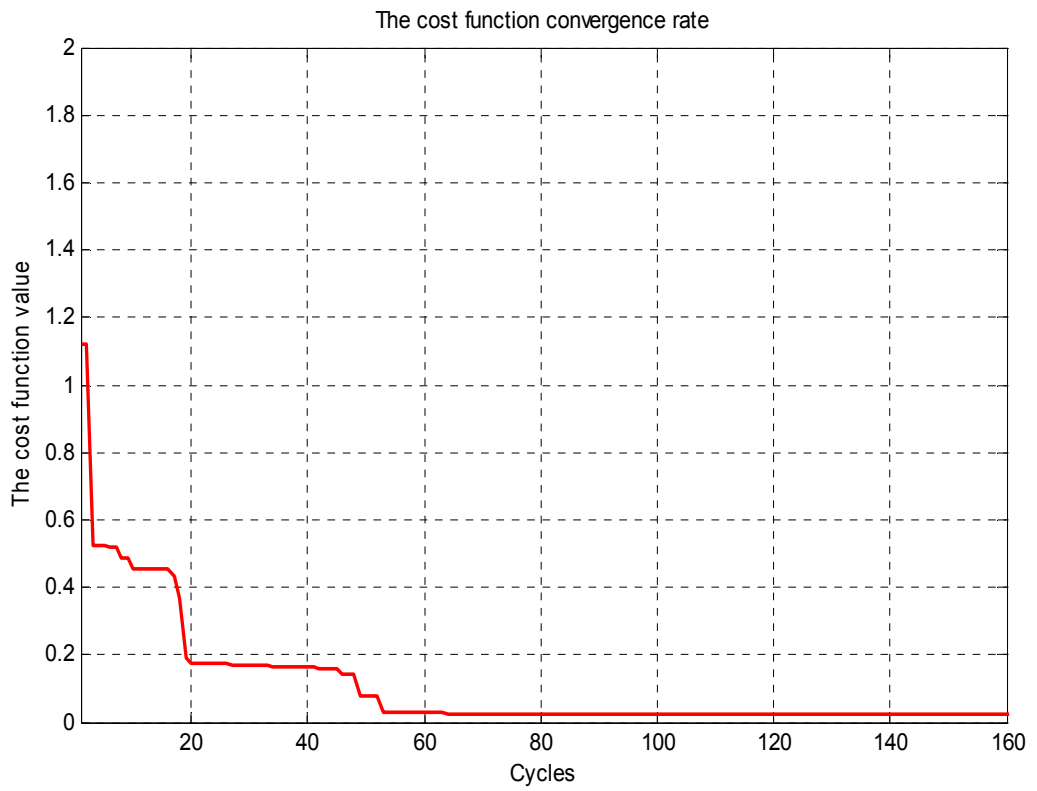
Figure 3.11: The best-so-far ABC algorithm convergence rate.
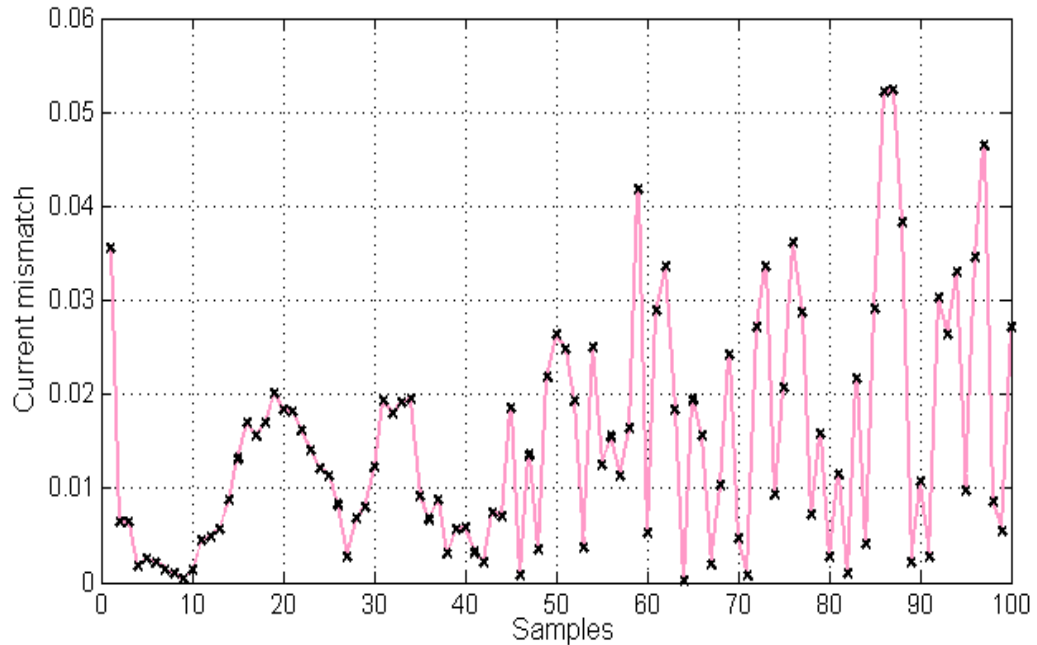
Figure 3.12: The evolution of the absolute error between

the measured and estimated current.

3.4.2. Model validation results:

The extracted parameters, calculated in the previous section, are now introduced into the whole PV plant that models the Algerian GCPV system, in normal operating conditions. The inputs to this model are the actual measured daily profiles of temperature and irradiance. In this work, a co-simulation methodology was adopted by combining PSIM^{TM} and Matlab^{TM} software, where the physical system (PV generator and MPPT) is implemented in PSIM^{TM}, while data processing and the involved calculations are implemented in Matlab/Simulink^{TM}. The simulated output power ($P_{mpp-sim}$) is then compared to the output measured power ($P_{mpp-mea}$), under the same temperature and irradiance inputs, as shown in Figure (3.13).

As this stage is crucial for the fault detection and diagnosis step, detailed in the next section, the results given in Figure (3.13) must be quantified in terms of discrepancy between measured and simulated powers. To this end, the Mean Absolute Error (MAE), given by equation (3.9), was considered to evaluate the daily mean gap between $P_{mpp-sim}$ and $P_{mpp-meas}$. The calculated MAE for this particular day, characterized by fast-moving clouds, is found to be about 8 watts.

$$MAE_{Pmpp} = \frac{1}{L}\sum_{d=1}^{L}\left|P_{mpp\_meas} - P_{mpp\_sim}\right| \tag{3.9}$$

where: L denotes the length of MPP power vector.
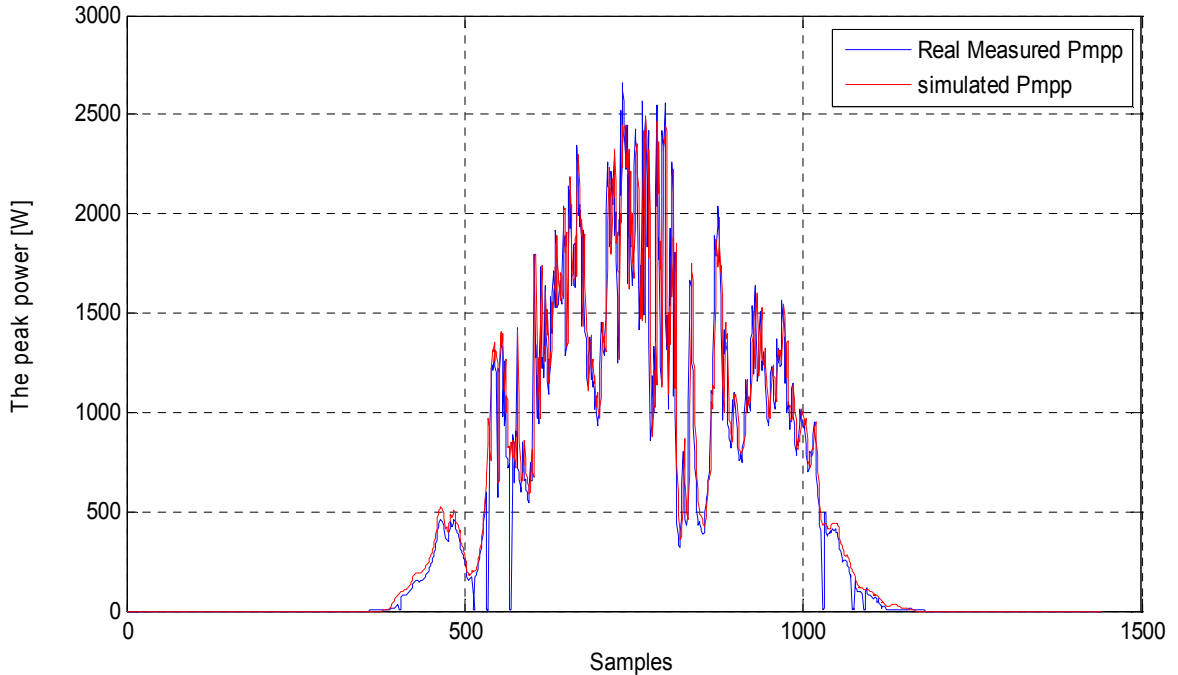
Figure 3.13: Real measured against simulated peak power.

3.4.3. Fault detection and diagnosis results

In this subsection, the effectiveness of the proposed fault detection and diagnosis approach is presented in detail. At first, the PNN construction, training and validation are conducted with noiseless data, obtained from the simulation model, in normal and faulty operations. In the second step, the constructed PNNs are tested against noisy data, obtained by adding a predefined noise to the simulated data. In addition, a comparison study with classical ANN classifier for both noiseless and noisy data is also carried out in this subsection.

3.4.3.1. Noiseless data case:

Once the simulation model has been validated, the previously described DC side failures have been simulated under measured daily profiles of temperature and irradiance. This simulation aims the elaboration of a high-quality database for both healthy and faulty systems as described in section (3.3). The formed database contains 2960 data for each operating case, which means a total of 11840 data for the four simulated scenarios. For the training phase, we have

chosen 75% (8880 data) while the test set contains the remaining 25 % (2960 data). The details of the total database construction are given bellow:

- 740 data are associated with each of the four attributes (T, G, I_{mpp} and V_{mpp}) for each operating case. Thus, a total of 2960 = 740 x 4 data set is associated with each operating case. More precisely, the vector of each attribute is subdivided into two sub-vectors: a first one comprises 556 samples (75% of 740) and it is stated as the classifier training set, while the second sub-vector includes the remaining 184 samples (25% of 740), and it is used as the classifier testing set.

- The test dataset comprises 184 x 4 =736 data for each attribute, organized in the following manner:
  - The test set of healthy system includes 184 samples.
  - The test set of fault #1 includes 184 samples.
  - The test set of fault #2 includes 184 samples.
  - The test set of fault #3 includes 184 samples.

  On the other hand, this work proposes the use of two classifiers (two different networks). The first PNN is responsible of failures detection, while the second classifier aims to diagnose accurately the type of the occurred fault. Therefore, the performance of the detection network is assessed by using the test set in the following manner:
  - Healthy system test set includes 184 samples for each attribute.
  - Faulty system test set includes 184 x 3=552 samples (Denoted by the three faulty cases: Fault #1, Fault #2 and Fault #3) for each attribute.

- While the performance of the diagnosis network is assessed by using the test set in the following manner:
  - Fault #1 test set includes 184 samples for each attribute.
  - Fault #2 test set includes 184 samples for each attribute.
  - Fault #3 test set includes 184 samples for each attribute.

It is worth mentioning here that the measured data from the real operating system are considered as noiseless data as the quantification of error measurements and sensor drifts is a very difficult task.

For comparison purpose, the performances of the proposed fault detection and diagnosis in PV systems are compared with those obtained with the classical ANN classifier. Toward this end, two ANNs have been constructed, trained and

tested with the same dataset as the PNN. The number of the hidden layers in the ANN networks and the smoothing parameter, σ, for the PNN networks are obtained by trial and error tests. The main parameters used in this work, for each network, are summarized in table (3.4). Classifiers efficiency (total efficiency) and classification sensitivity results for ANN and PNN classifiers are summarized in Table (3.5).

Table 3.4: ANN and PNN networks adjustable parameters

	Feed Forward Back-Propagation Network (ANN)	Probabilistic Neural Network (PNN)
Detection network	The weights adjustment's method: back propagation algorithm. The number of neurons in the hidden layer= 25. The activation function: tangent sigmoid. The training phase's maximum number of epochs = 100. The training phase's performance goal = 10^{-5}.	The smoothing parameter $\sigma = 1.1$. The activation function: exponential
Diagnosis network	The weights adjustment's method: back propagation algorithm. The number of neurons in the hidden layer = 40. The activation function: tangent sigmoid. The training phase's maximum number of epochs = 100. The training phase's performance goal = 10^{-5}.	The smoothing parameter $\sigma = 1.1$. The activation function: exponential

Table 3.5: Classification accuracy and sensitivity results for noiseless data

Detection networks tested with noiseless data			
The classifier type	Classification accuracy %	Sensitivity %	
		Healthy operating case	Faulty operating case
ANN	90.35	100	78.14
PNN	100	100	100
Diagnosis networks tested with noiseless data			
The classifier type	Classification accuracy %	Sensitivity %		
		Fault 1	Fault 2	Fault 3
ANN	73.18	86.96	100	32.61
PNN	100	100	100	100

While the confusion matrixes of the two networks (ANN and PNN for detection and diagnosis) are summarized in tables (3.6-3.7) respectively.

Table 3.6: Confusion matrixes of ANN networks for noiseless data

Detection network type ANN tested on the noiseless data		
The real classes	The predicted classes	
	Healthy case	Faulty case
Healthy case	184	0
Faulty case	71	481
Diagnosis network type ANN tested on the noiseless test data		
The real classes	The predicted classes	
	Fault 1	Fault 2	Fault 3
Fault 1	160	24	0
Fault 2	0	184	0
Fault 3	119	5	60

Table 3.7: Confusion matrixes of PNN networks for noiseless data

Detection network type PNN tested on noiseless data		
The real classes	The predicted classes	
	Healthy case	Faulty case
Healthy case	184	0
Faulty case	0	552
Diagnosis network type PNN tested on the noiseless test data		
The real classes	The predicted classes	
	Fault 1	Fault 2	Fault 3
Fault 1	184	0	0
Fault 2	0	184	0
Fault 3	0	0	184

For more clarity of the obtained classification results, the confusion matrixes have been displayed graphically for the two networks (PNN and ANN). Figures (3.14-3.15) are the simulation results of the ANN classifiers in detection and diagnosis stages respectively, while Figures (3.16-3.17) are the results of the PNN

classifiers. The simulation results in terms of Specificity and Positive Predictivity for noiseless data are summarized in table (3.8) (ANN networks) and table (3.9) (PNN networks), respectively.
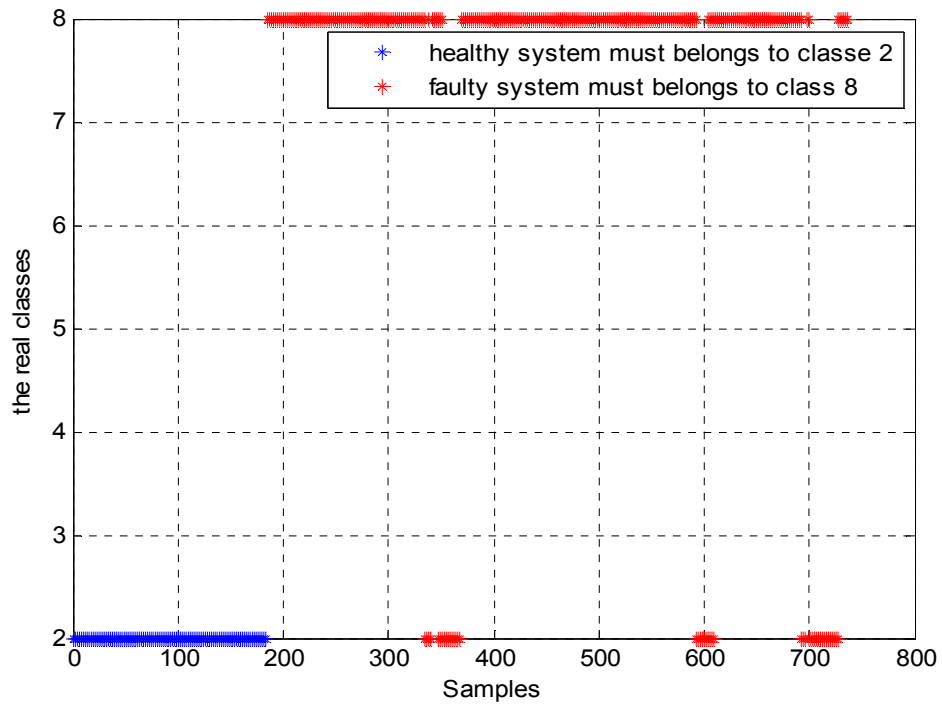
Figure 3.14: Fault detection results based on ANN classifier
for the case of noiseless data.

Figure 3.15: Fault diagnosis results based on ANN classifier
for the case of noiseless data.

Figure 3.16: Fault detection results based on PNN classifier

for the case of noiseless data.

Figure 3.17: Fault diagnosis results based on PNN classifier

for noiseless data

Table 3.8: Classification Specificity and Positive Predictivity for ANN networks with noiseless data

Detection network type ANN tested on noiseless data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Healthy case	87.14	72.16
Faulty case	100	100
Diagnosis network type ANN tested on the noiseless test data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Fault1	67.66	57.35
Fault2	92.12	86.38
Fault3	100	100

Table 3.9: Classification Specificity and Positive Predictivity for PNN networks with noiseless data

Detection network type PNN tested on the noiseless test data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Healthy case	100 %	100 %
Faulty case	100 %	100 %
Diagnosis network type PNN tested on the noiseless test data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Fault1	100 %	100 %
Fault2	100 %	100 %
Fault3	100 %	100 %

According to the results given above, the high classification abilities of the proposed method, based on PNN classifiers, are clearly demonstrated, whether at the detection stage or at the diagnosis stage. In fact, all the evaluated metrics (classification accuracy, sensitivity, specificity and positive predictivity) give a 100% success rate for the classification of the input samples in their corresponding classes (see table (3.5) and table (3.9)). Furthermore, these high performances, to correctly classify the incoming samples in their right classes, are highlighted by the confusion matrixes given in table (3.7) and represented graphically in Figures (3.16-3.17). However, the evaluation metrics for ANN classifiers (Table (3.5)) have given lower percentage rate in both detection

and diagnosis stages. For instance, a very low percentage rate is obtained for sensitivity metric (32.61%) in case of fault#3. This luck to classify some samples in their right classes, observed in ANN classifiers, is clearly observed in the confusion matrixes plots where data belonging to certain classes are wrongly classified in another classes (Figures (3.14-3.15)). Thus, with the case of noiseless data, the PNN classifiers have given better results than ANN classifiers regarding fault detection and diagnosis.

3.4.3.2. Noisy data case

In practical cases, the measured quantities from the PV system are subjected to various types of disturbances and noises, caused mainly by the tolerances of the measuring devices and sensor drifts.

However, since these disturbances are difficult to be quantified from a practical point of view and in order to test the performances of the proposed PNN with noisy data, we have considered the measured data as noiseless data where a predefined noise function is added to each attribute to form the noisy data.

This predefined noise function is a simple Matlab^{TM} routine given by the following equation:

$$Disturbance = \alpha + \beta \times randn(1, N) \tag{3.10}$$

Where:

α is the mean of the disturbance signal; β is the standard deviation of the disturbance signal and which defines its magnitude; "randn" is an implemented Matlab function used to generate a normal distributed number, and N is the size of the disturbance signal.

Subsequently, the two networks (ANN and PNN), previously constructed in subsection (3.3.2.2), are tested on noisy data as follows:

- The disturbance signal is added to the corresponding attributes (T, G, I_{mpp} and V_{mpp}) by using equation (3.10).

- The added noisy signals are characterized by their means equal to zero and their standard deviations equal to the disturbance magnitude.

- The chosen magnitudes of each disturbed attribute are: $\beta_T = 4°C$, $\beta_G = 5$ W/m², $\beta_{Impp} = 2$ A, and $\beta_{Vmpp} = 5$ V.

In order to test the networks efficiency with noisy data, the used learning set is composed of noiseless data, while the test set includes only the noisy attributes.

By following the same assessment steps, described in subsection (3.4.3.1), the simulation results in terms of Classification Accuracy (Total Efficiency), Classification Sensitivity, Confusion Matrixes, Specificity and Positive Predictivity for the case of noisy data are summarized in Tables (3.10-3.14) and graphically represented in Figures (3.18-3.21).

Table 3.10: Classification accuracy and Sensitivity results with noisy data

Detection networks tested on the noisy test data			
The classifier type	**Classification accuracy %**	**Sensitivity %**	
		Healthy operating case	**Faulty operating case**
ANN	58.02	66.85	55.07
PNN	82.34	61.96	89.13
Diagnosis networks tested on the noisy data			
The classifier type	**Classification accuracy %**	**Sensitivity %**	
		Fault 1	**Fault 2**	**Fault 3**
ANN	76.63	69.02	91.85	69.02
PNN	98.19	100	94.57	100

Table 3.11: Confusions matrixes of ANN networks with noisy data

Detection network type ANN tested on the noisy test data		
The real classes	**The predicted classes**	
	Healthy case	**Faulty case**
Healthy case	123	61
Faulty case	248	304
Diagnosis network type ANN tested on the noisy test data		
The real classes	**The predicted classes**	
	Fault 1	**Fault 2**	**Fault 3**
Fault 1	127	57	0
Fault 2	15	169	0
Fault 3	15	42	127

Table 3.12: Confusions matrixes of PNN networks for the case of noisy data.

Detection network type PNN with noisy data		
The real classes	**The predicted classes**	
	Healthy case	**Faulty case**
Healthy case	114	70
Faulty case	60	492
Diagnosis network type PNN with noisy data		
The real classes	**The predicted classes**	
	Fault 1	**Fault 2**	**Fault 3**
Fault 1	184	0	0
Fault 2	10	174	0
Fault 3	0	0	184
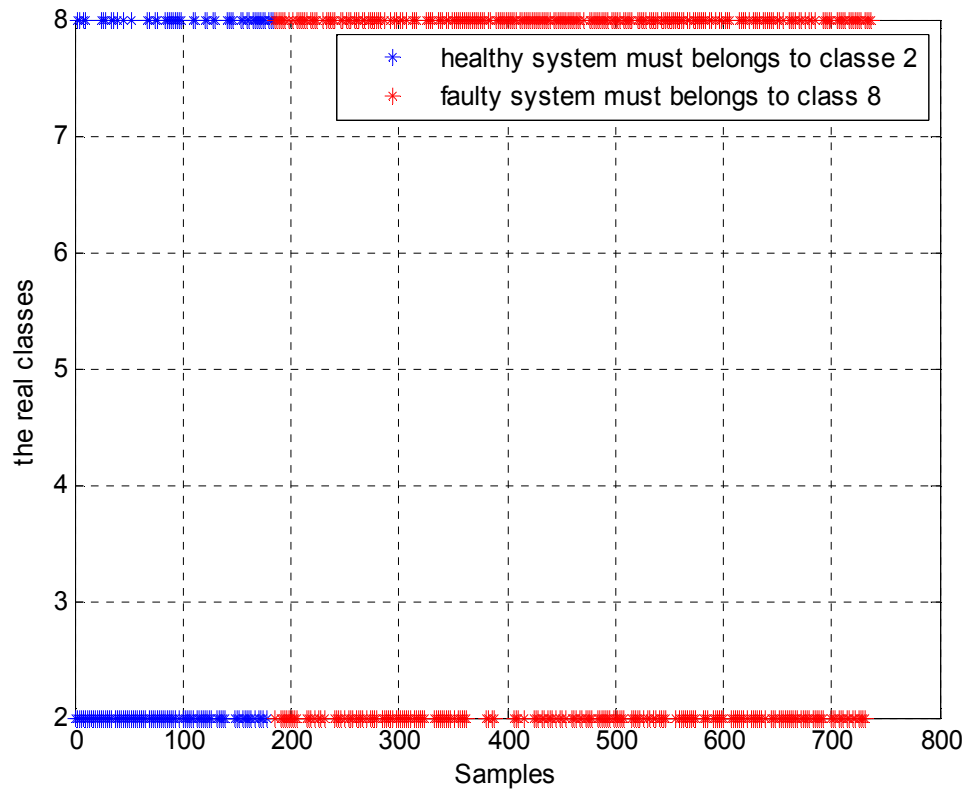
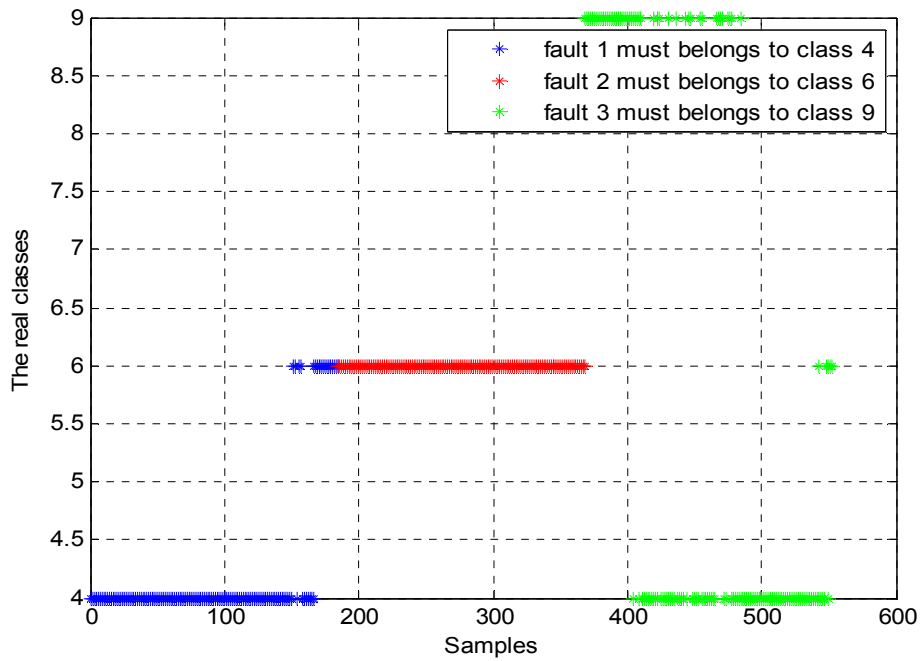Figure 3.18: Fault detection results based on ANN classifier for the case of noisy data.
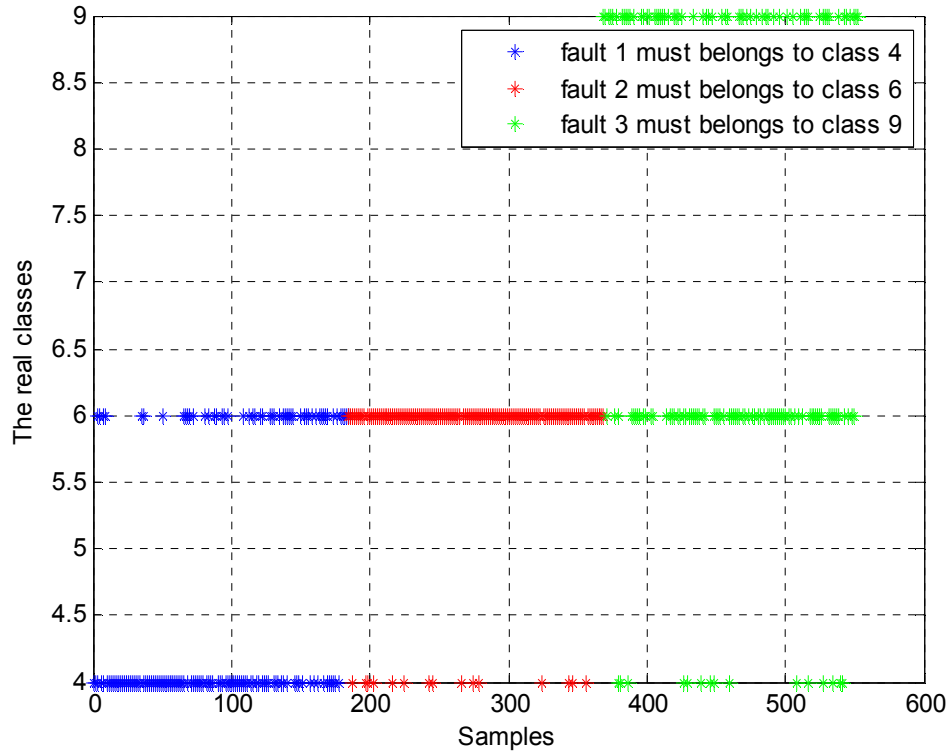
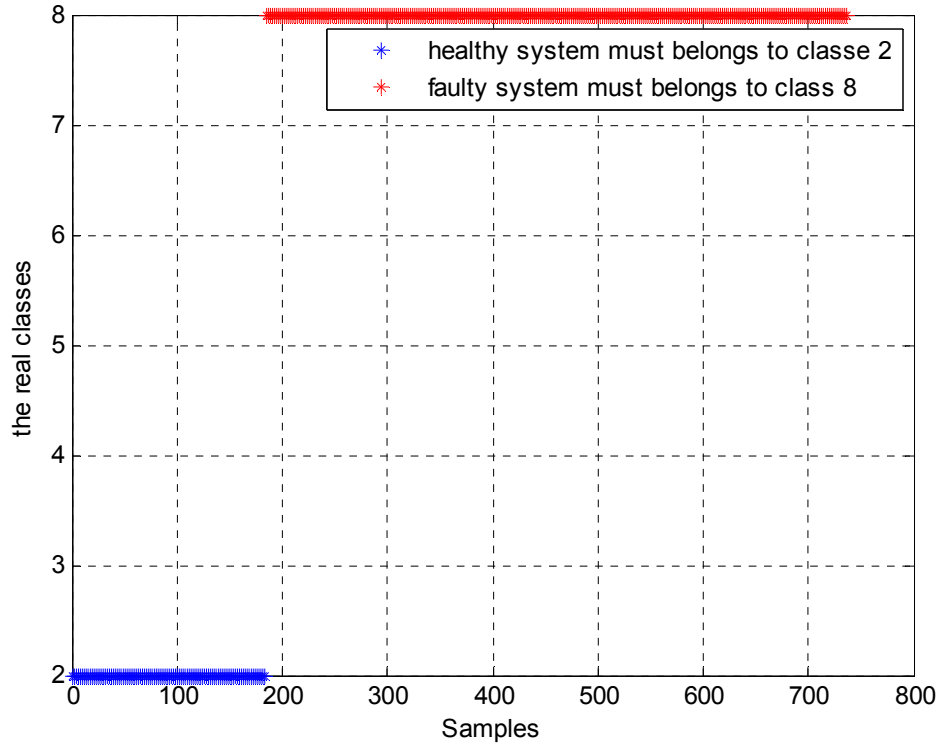Figure 3.19: fault diagnosis results based on ANN classifier for the case of noisy data.

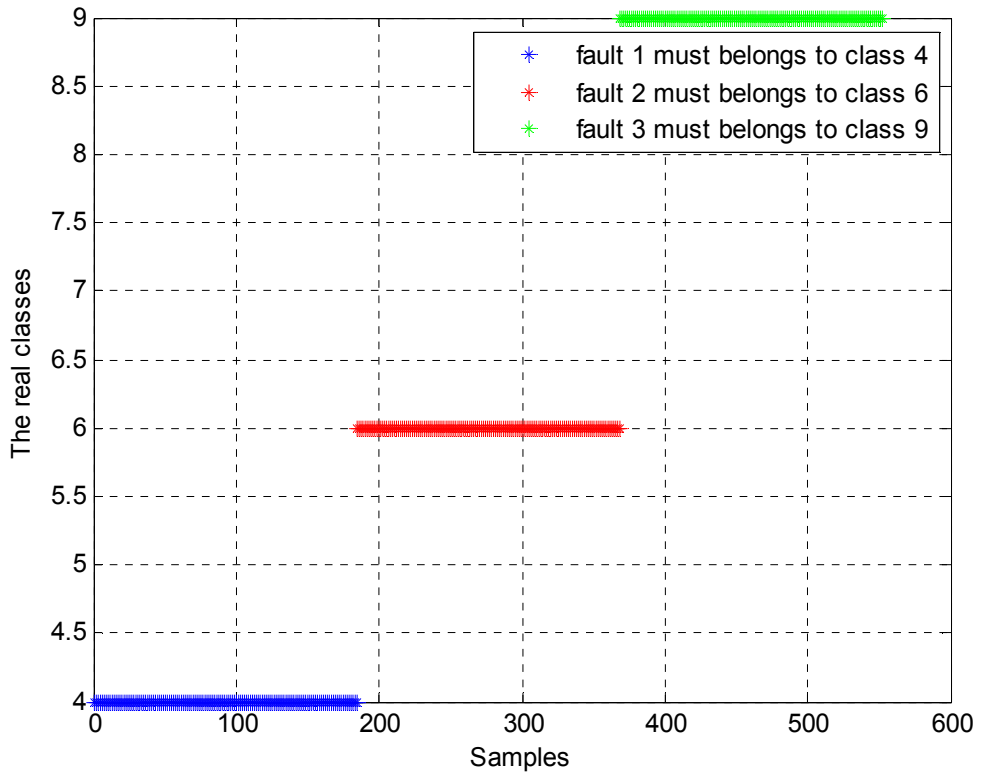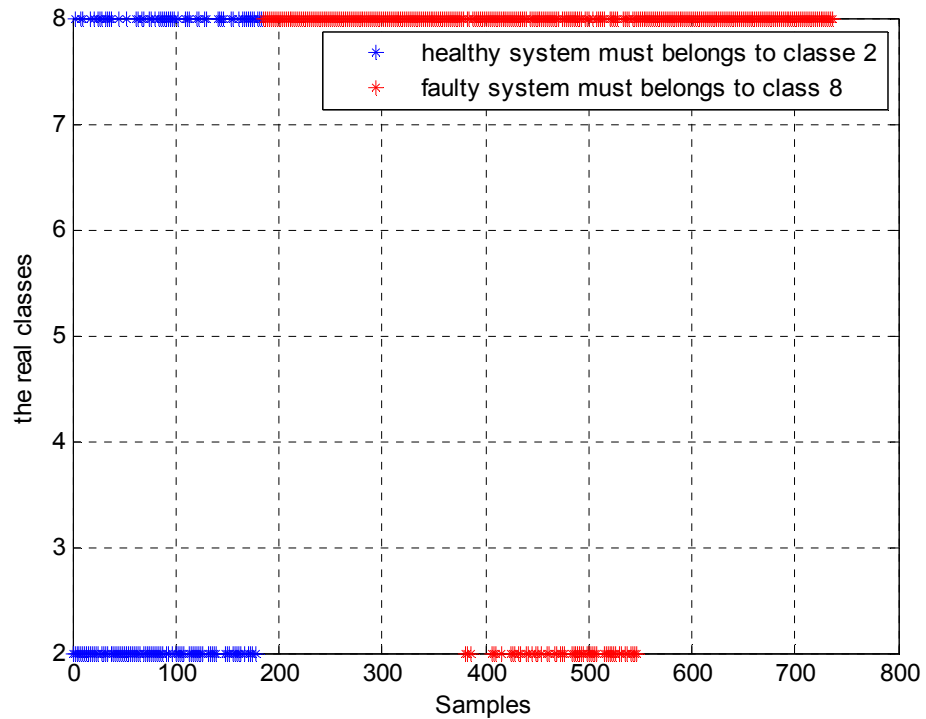Figure 3.20: fault detection results based on PNN classifier for the case of noisy data.
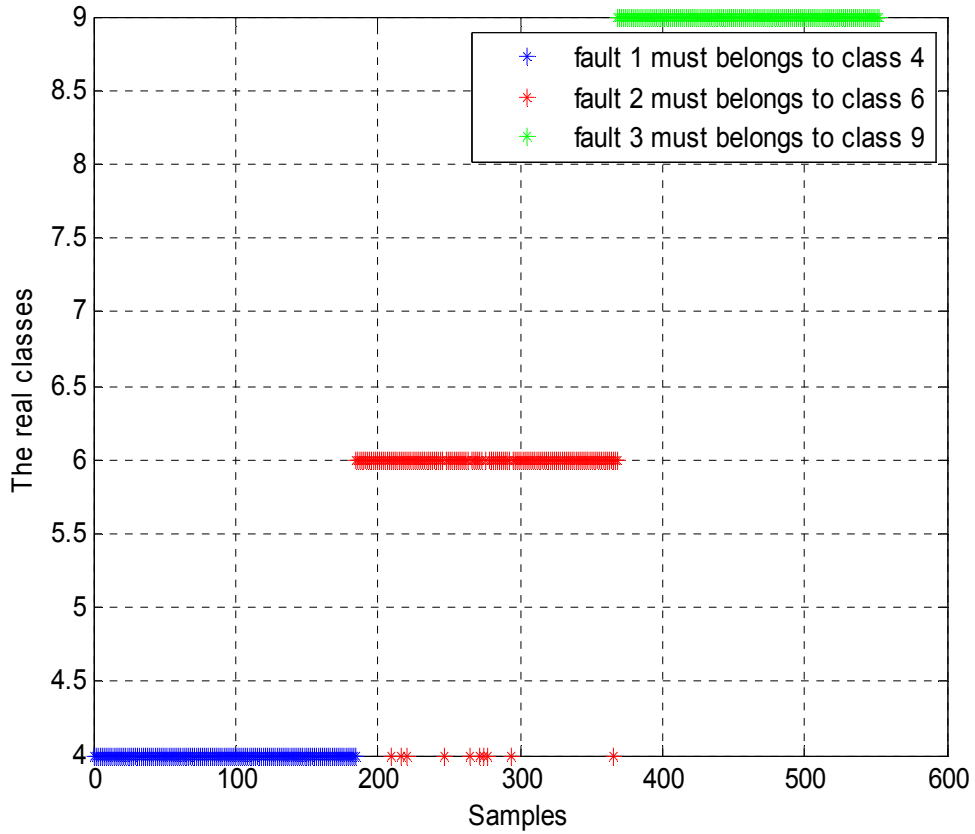
Figure 3.21: Fault diagnosis results based on PNN classifier
for the case of noisy data.

Table 3.13: Classification specificity and Positive Predictivity
for ANN networks under noisy data

Detection network type ANN tested on the noisy test data		
classes	Specificity (%)	Positive Predictivity (%)
Healthy case	55.07	33.15
Faulty case	66.85	83.29
Diagnosis network type ANN tested on the noisy test data		
classes	Specificity (%)	Positive Predictivity (%)
Fault1	91.85	80.89
Fault2	73.10	63.06
Fault3	100	100

Table 3.14: Classification specificity and Positive Predictivity
for PNN networks under noisy data.

Detection network type PNN tested on the noisy test data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Healthy case	89.13	65.52
Faulty case	61.96	87.54
Diagnosis network type PNN tested on the noisy test data		
classes	**Specificity (%)**	**Positive Predictivity (%)**
Fault1	97.28	94.85
Fault2	100	100
Fault3	100	100

By closely analyzing the above results, it can be noted that even the exaggerated magnitudes of the noisy attributes (β_T = 4°C, β_G = 5W/m², β_{Impp} = 2A, and β_{Vmpp} =5V), the two PNN classifiers still have a motivated results in both detection and diagnosis stages.

In addition, according to table (3.10) and table (3.14), the main performance metrics have not undergone a serious decline even under this test condition. For instance, the recorded classification efficiency, for the proposed classifiers, is 82.34% for detection network, and 98.19% for diagnosis network.

In contrast, the test of the ANN classifiers under noisy data has led to a significant decrease of almost all the performance metric indicators. For instance, the recorded classification efficiency is 58.02% for the detection network. Moreover, a very low percentage rate is obtained for the positive predictivity metric (33.15%) in the case of the healthy state (Table (3.13)). This significant decrease of the performances for ANN classifiers is also clearly observed in the confusion matrixes plots where several amounts of data belonging to certain class are wrongly classified in another classes (Figures (3.18-3.19)).

3.4.3.3. PNN classifiers' efficiency under real operating conditions:

a. PNN classifier' efficiency under noiseless data:

As mentioned in the previous sections, the on-site measurements have been considered as noiseless data in this work. Thus, in this subsection a real measured daily profile of temperature and irradiance is used to assess the detection and diagnosis capabilities of the proposed PNN classifiers when the prescribed faults occur in a particular time of day. For clarity purposes, the measurements of a clear day are taken as a test pattern, where faults are intentionally introduced in the following manner:

From 2h:00 to 8h:00, the PV system is in normal working conditions (no faults), then a short-circuit of three PV modules (Fault#1) in the same string has been introduced at 8h:01. The system remains in that faulty condition until 9h:30. At this time, the fault is cleared and the system comes back to its normal state. At 11h:00, the system was forced to work under a condition of ten (10) short-circuited PV modules (Fault#2). The system keep working in this state until 12h:30 where the fault is cleared and the PV system comes back to its normal condition. After that, from 14h:00 to 15h:30 the first string of the PV array is completely disconnected (Fault#3). Thereafter, this fault is removed and the system continues working in normal conditions. These details are graphically shown in Figures (3.22-3.23), for the monitored irradiance and module temperature, and Figures (3.24-3.26), for I_{mpp}, V_{mpp} and P_{mpp}.
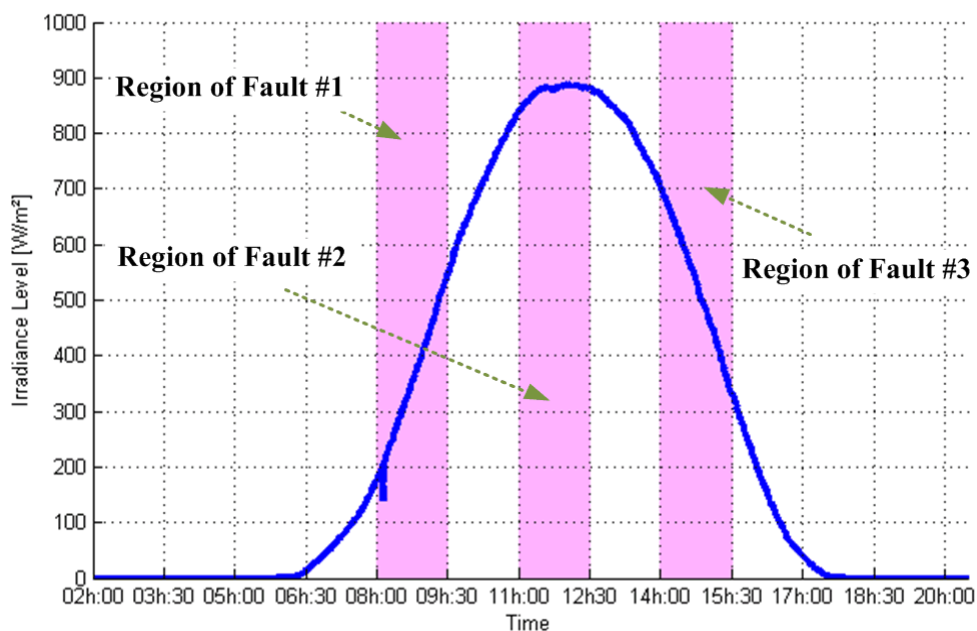
Figure 3.22: The used daily profile of onsite irradiance level.
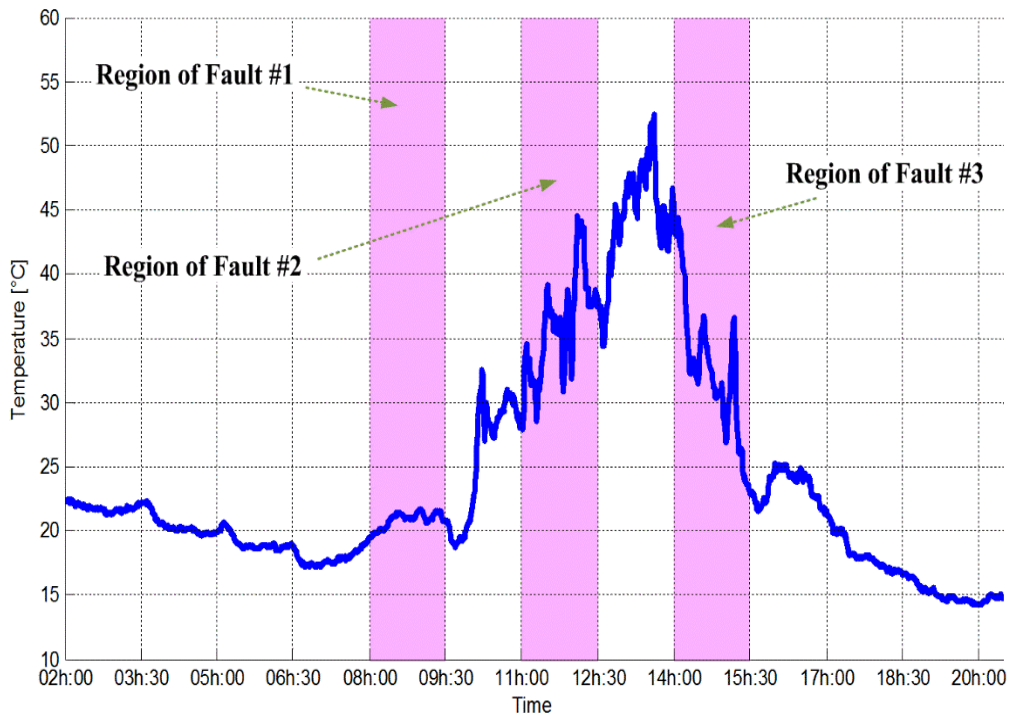
Figure 3.23: The used daily profile of onsite temperature level.
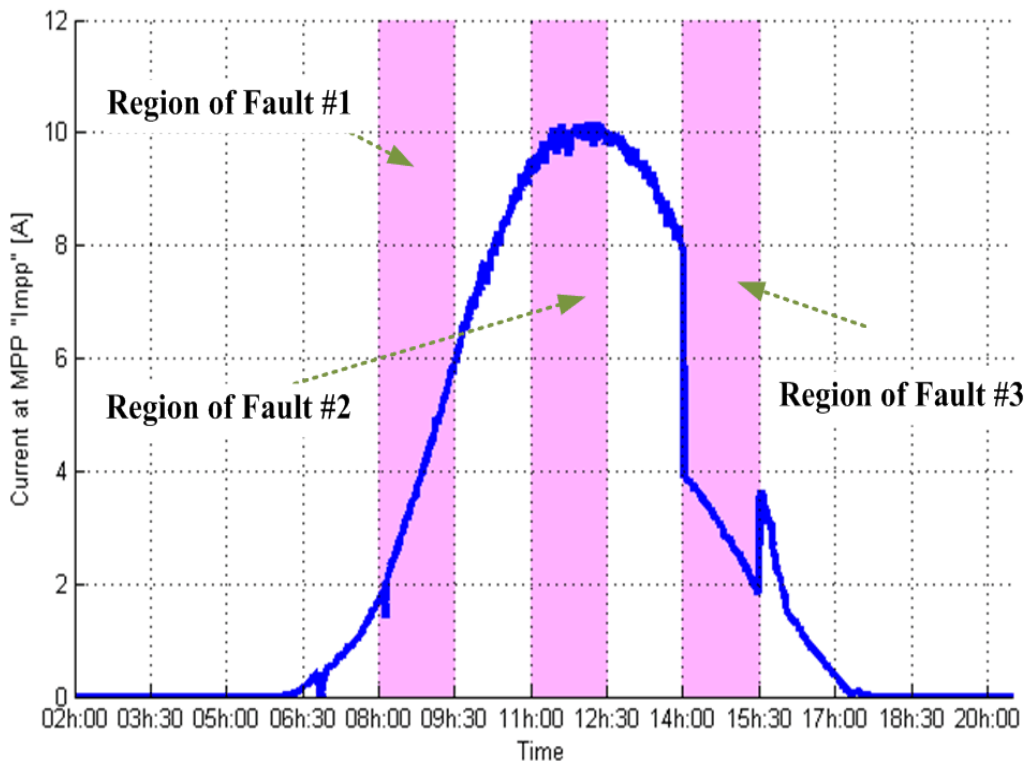
Figure 3.24: The used daily profile of real measured current at MPP "I_{mpp}".
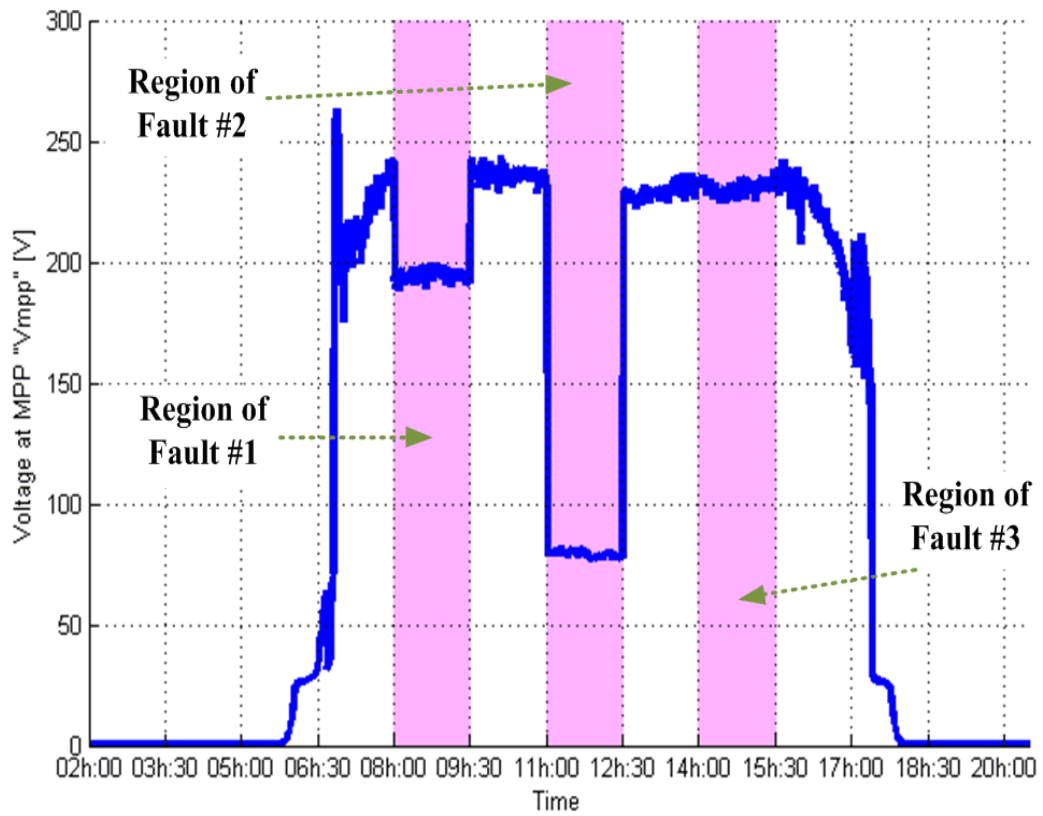
Figure 3.25: The used daily profile of real measured voltage at MPP "V_{mpp}".
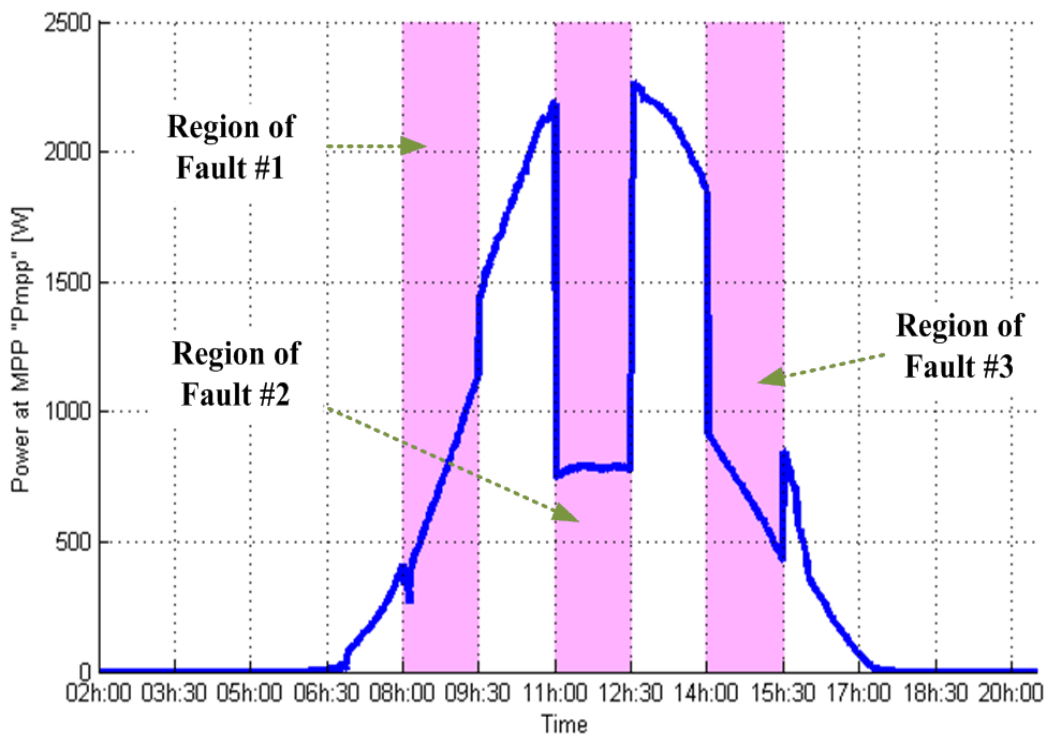
Figure 3.26: The used daily profile of real measured power at MPP "P_{mpp}".

The main results, obtained when using the previous patterns, are shown in Figure (3.27) (detection) and in Figure (3.28) (diagnosis).

It is clearly shown that, before the occurrence of any fault, the detection PNN network (Figure (3.27)) generates 'Class 2', indicating healthy operating state. While the diagnosis PNN generates, in this particular case, "class 0" which denotes that is not yet activated. Therefore, the entire photovoltaic array is considered to be in a normal operating state.

In case of faulty operation, the detection PNN generates "Class 8" indicating the occurrence of a fault (see Figure (3.27)), while the diagnosis PNN generates "Class 4, Class 6 and Class 9", which are the corresponding classes for the considered faults diagnosed at the time of their appearance (Figure (3.28)).
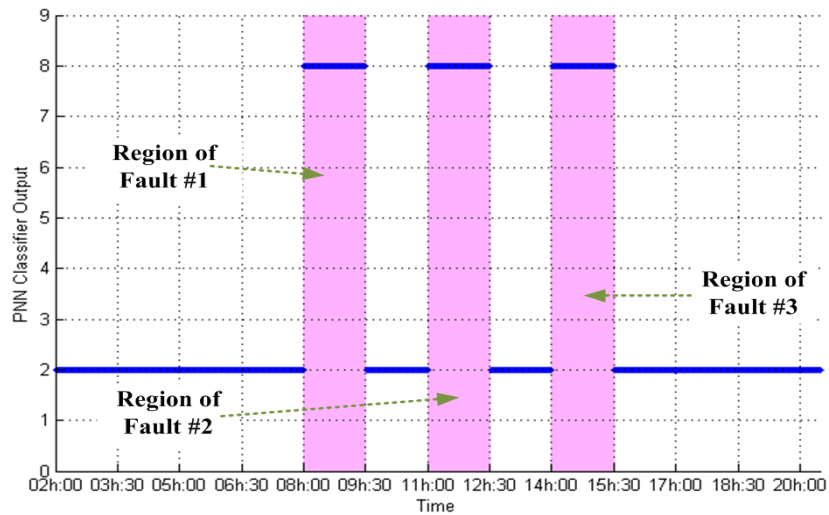
Figure3.27: Fault detection results of PNN classifier in the presence of multiple faults and tested under noiseless data.
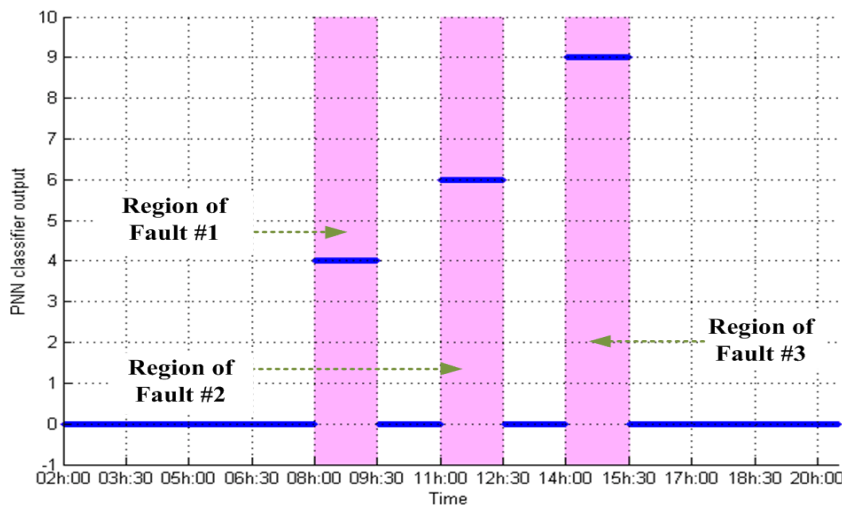
Figure 3.28: Fault diagnosis results of PNN classifier in the presence of multiple faults and tested under noiseless data.

b- <u>PNN classifiers' efficiency under noisy data.</u>

In this last test, the effectiveness of the proposed method is carried out by superimposing the predefined noise levels (Equation (3.10)) to the monitored data. The same patterns, described above, have been used in this test. Fault detection and diagnosis results are summarized in Figures (3.29-3.30), respectively.
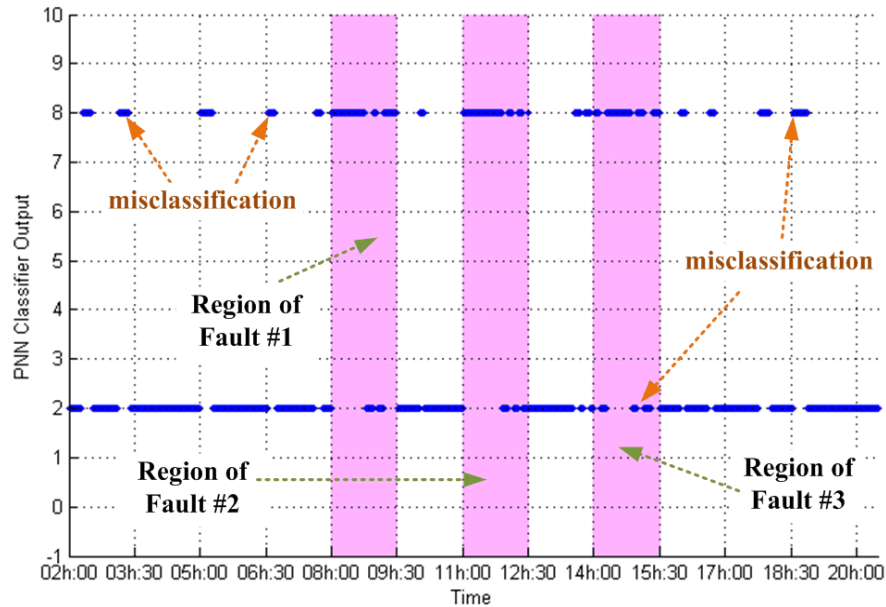
Figure 3.29: Fault detection results of PNN classifier in the presence of multiple faults and tested under noisy data.
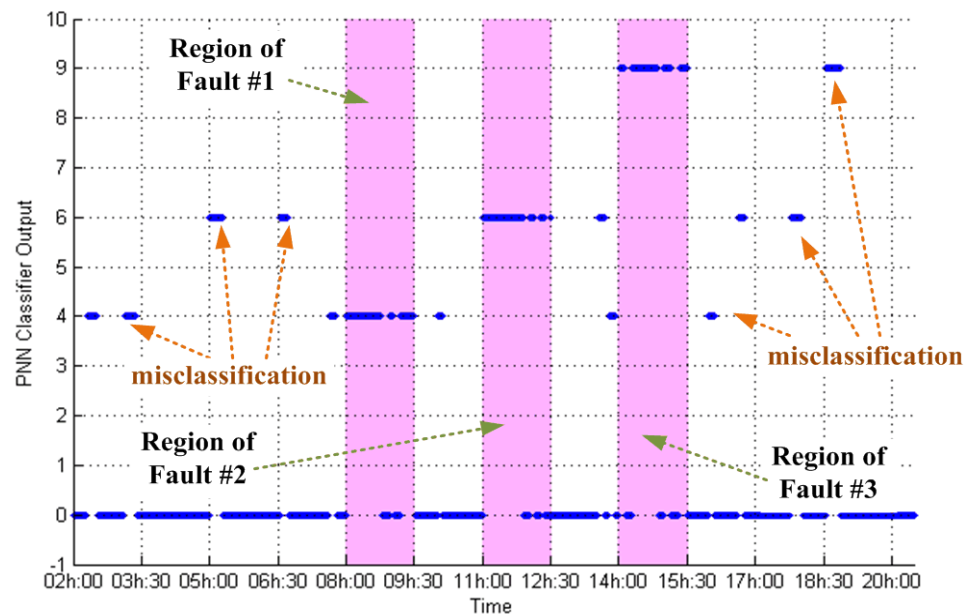
Figure 3.30: Fault diagnosis results of PNN classifier in the presence of multiple faults and tested under noisy data.

In this particular test, the results shown in figures (3.29-3.30) include some misclassifications mainly due to the exaggerated magnitudes of the superimposed noises. Obviously, such noise levels can lead to highly distorted attributes. But, despite the use of noise signals of a such magnitude, fault detection and diagnosis results based on PNN classifiers are very motivating even with the appearance of these misclassifications. In fact, the recorded PNN classifier efficiency is of 82.34% for the detection network and 98.19% for the diagnosis network.

3.5. Conclusion:

This chapter has proposed an enhanced machine learning based approach for fault detection and diagnosis of PV systems. The efficiency of this method has been tested and validated using experimental measurements and simulated data. In the other side, the proposed strategy has been tested and compared with the conventional feed-forward ANN classifier for both noisy and noiseless data. Finally, this method has been evaluated under real operating conditions, using real experimental measurements, collected from the tested PV system, in order to evaluate its performance and robustness. Results have shown clearly the high efficiency of this method to detect and diagnose DC side anomalies occurrences for both noiseless and noisy data.

However, the efficiency of this approach to detect and diagnose short-circuit faults is guaranteed from at least two short-circuited PV modules. In fact, short-circuit of one PV module cannot be detected using this approach. This is due to its high similarity with healthy system for a large scale PV plant.

CHAPTER 4

FAULTS DETECTION AND DIAGNOSIS OF PHOTOVOLTAIC SYSTEMS USING STATISTICAL APPROACHES

4.1. <u>Introduction:</u>

In this chapter, two statistical-based methods of faults detection and diagnosis in PV systems are developed. The first method denotes the improved ratio-based strategy [17]. It consists on the evaluation of three coefficients: currents coefficient α , voltage coefficient β and power coefficient γ , and it allows the detection and diagnosis of short-circuits and open-circuits faults.

In the other side, the second statistical-based method applies two control charts, EWMA [11, 12, 58] and Shewhart [59-61] charts, to detect and diagnose DC side faults. This last strategy combines the ODM flexibility, and the control charts aptitude in detecting small changes in PV systems, to elaborate an efficient strategy of faults detection and diagnosis.

The detailed explanations of these two methods will be given in the next sections.

4.2. <u>The improved ratio-based method:</u>

This section proposes an efficient strategy to detect and diagnose short-circuits and open-circuits faults in PV systems, based on the evaluation of three coefficients. The proposed method consists fundamentally on two steps: an offline step based on a simulation model, and an online step at which a comparison between the real measured coefficients against those obtained in the offline step is performed. The simulation model of the PV array has been validated using real experimental data of a daily profile from the Algerian PV system described previously. The effectiveness of the proposed method has been evaluated also based on the PSIM™/Matlab™ Co-simulation approach of four operating cases: healthy operating case, one short-circuited module in a string, five short-circuited modules in a string, and finally a completely disconnected string in the array operating case. Simulation results have demonstrated the ability of the proposed

method to detect and diagnose short-circuits and open-circuits faulty operations under any meteorological condition.

4.2.1. Fundamental principle:

In order to detect and diagnose both short-circuits faults and open-circuits faults, three coefficients for each type of fault have been introduced: current coefficient α, voltage coefficient β and finally power coefficient γ.

The expressions given bellow define each coefficient.

$$\alpha_{faulty\_case}(T,G) = \frac{I_{mpp\_healthy\_case}(T,G)}{I_{mpp\_faulty\_case}(T,G)} \qquad (4.1)$$

$$\beta_{faulty\_case}(T,G) = \frac{V_{mpp\_healthy\_case}(T,G)}{V_{mpp\_faulty\_case}(T,G)} \qquad (4.2)$$

$$\gamma_{faulty\_case}(T,G) = \frac{P_{mpp\_healthy\_case}(T,G)}{P_{mpp\_faulty\_case}(T,G)} \qquad (4.3)$$

where:

$I_{mpp\_healthy\_case}$, $V_{mpp\_healthy\_case}$ and $P_{mpp\_healthy\_case}$ are the non-faulty case's MPP coordinates; $I_{mpp\_faulty\_case}$, $V_{mpp\_faulty\_case}$ and $P_{mpp\_faulty\_case}$ are the MPP coordinates of the faulty cases (short-circuits and open-circuits faults).

The proposed approach consists of two main steps: an offline step that aims to extract the lower and higher boundaries of each one of the three coefficients, for each category of fault, and an online step in which we calculate the three real onsite coefficients and compare them against the variation boundaries of each type of fault.

Hereafter the detailed procedure of this method is presented:

a- The offline step

As described before, the aim of this step is to extract, for each faulty case, the variation boundaries of each coefficient. Toward this end, three stages should be accomplished, as follows:

- Simulate both healthy and faulty operating cases, under few meteorological conditions, using a PSIM^{TM}/Matlab^{TM} Co-simulation. This step is devoted to extract few MPP coordinates for each simulated case.

- Calculate the three coefficients (α, β and γ) of each faulty case (short-circuits and open-circuits faults) based on equations (4.1-4.3).

- For each type of fault, extract the variation boundaries of the three coefficients by adding an offset of $\pm 2\%$ for the three obtained coefficients.

b- <u>The online step</u>:

During the real operation of PV system, both meteorological conditions and MPP values could be measured and monitored using the different sensors. These measurements will be used to calculate the three real onsite coefficients, such as:

$$\alpha_{measured}(T,G) = \frac{I_{mpp\_healthy\_case}(T,G)}{I_{mpp\_real\_case}(T,G)} \tag{4.4}$$

$$\beta_{measured}(T,G) = \frac{V_{mpp\_healthy\_case}(T,G)}{V_{mpp\_real\_case}(T,G)} \tag{4.5}$$

$$\gamma_{measured}(T,G) = \frac{P_{mpp\_healthy\_case}(T,G)}{P_{mpp\_real\_case}(T,G)} \tag{4.6}$$

Where:

$I_{mpp\_healthy\_case}$, $V_{mpp\_healthy\_case}$ and $P_{mpp\_healthy\_case}$ describe the MPP current, Voltage and Power of the healthy case, and which could be obtained directly based on our proposed MPP estimation strategy given in chapter 2.

$I_{mpp\_real\_case}$, $V_{mpp\_real\_case}$ and $P_{mpp\_real\_case}$ denote the real measured MPP current, voltage and power, respectively.

Finally, the real onsite coefficients are compared against the variations boundaries of each faulty case, stored before from the offline step.

To finish with, the fault occurrence will be identified by using the real onsite power coefficients $\gamma_{measured}$ in such away if its value exceeds 1.02, a DC side's fault alarm will be triggered. In the other hand, the type of faults will be diagnosed based on the comparison between the three real onsite coefficients and the variation boundaries of each simulated faulty case.

4.2.2. Simulation results:

The effectiveness of the proposed improved ratio-based approach has been tested based on PSIM™/Matlab™ Co-simulation of the previously described Algerian GCPV system.

Four case studies: healthy PV array, one short-circuited PV module in a string, five short-circuited PV modules in a string and a completely disconnected string in the array; are suggested in this work, to show the performance of the proposed procedure to detect and diagnose the previously described faults. The four scenarios are simulated using real daily profile of irradiance and temperature.

The MPP data points obtained from the simulation results have been subdivided into two sets; the first set includes few measurements, and it has been reserved to the offline step, where the upper and lower variation boundaries of the three coefficients have been extracted for each fault type, while the second set includes the remaining measurements, and it has been considered as the test set of the proposed fault detection and diagnosis method. Simulations results of the offline step for the three faulty cases are shown in Figures (4.1-4.3).
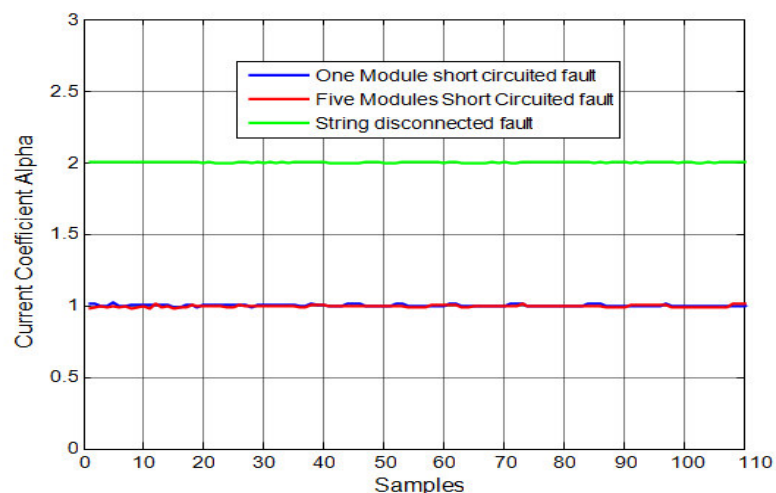
Figure 4.1: Current coefficient α evaluated under the different faulty operations.
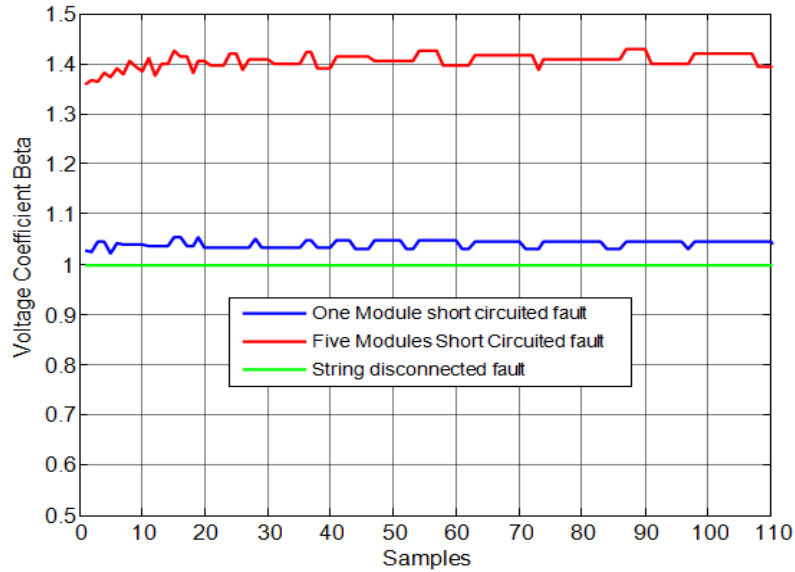
Figure 4.2: Voltage coefficient β evaluated under the different faulty operations.
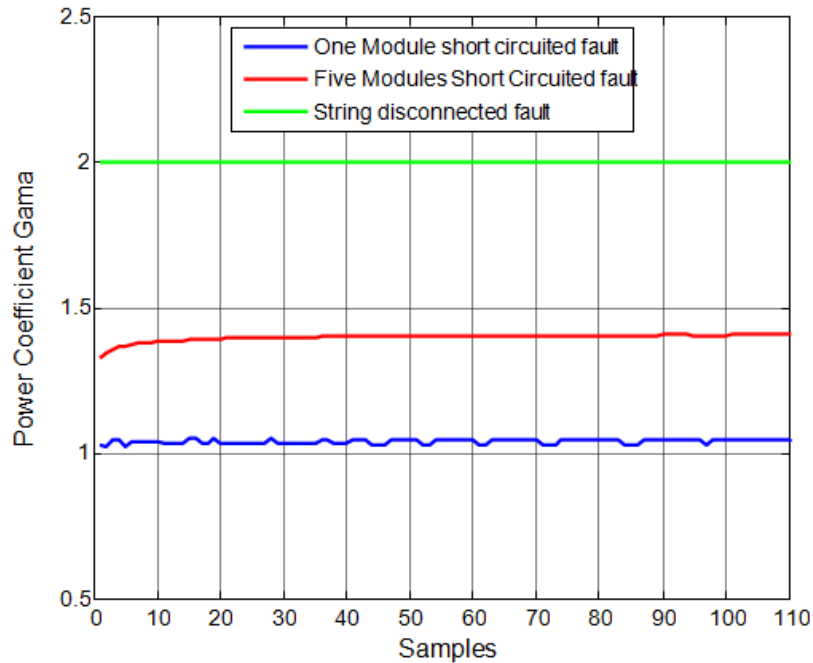
Figure 4.3: Power coefficient Υ evaluated under the different faulty operations.

As described previously, the next stage consists on the online step, where the actual MPP measurements should be compared against the boundary sets of each faulty case, in order to diagnose the type of the occurred faults.

Simulation results of the actual measured coefficients against those obtained in the offline step are shown in figures (4.4-4.6). During this simulation, the variation boundaries of each faulty case have been achieved from their faulty coefficients with an offset of $\pm\, 2\%$.
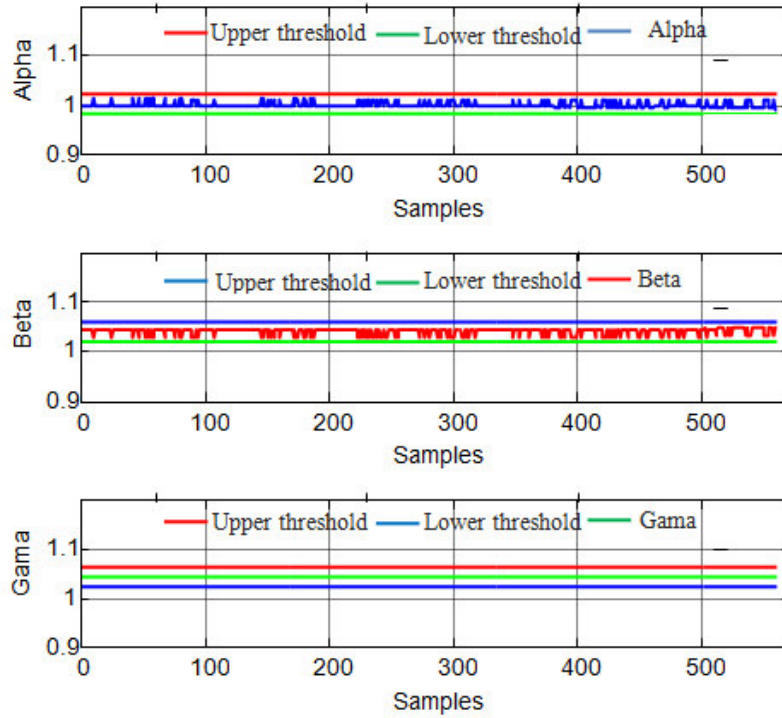
Figure 4.4: Actual measured coefficients against their variations boundaries for one short-circuited PV module faulty case.
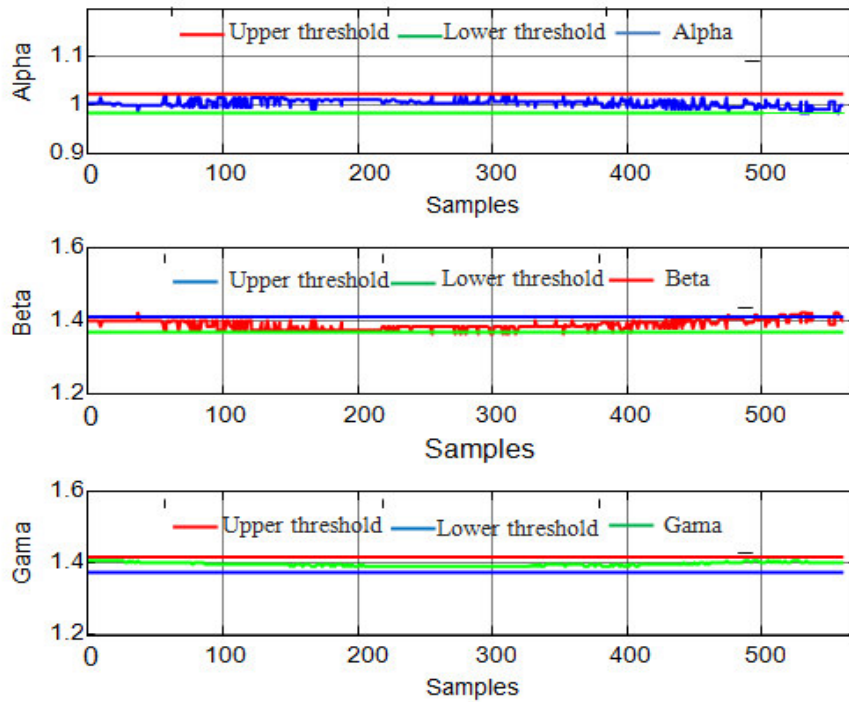
Figure 4.5: Actual measured coefficients against their variations boundaries for five modules short circuited faulty case
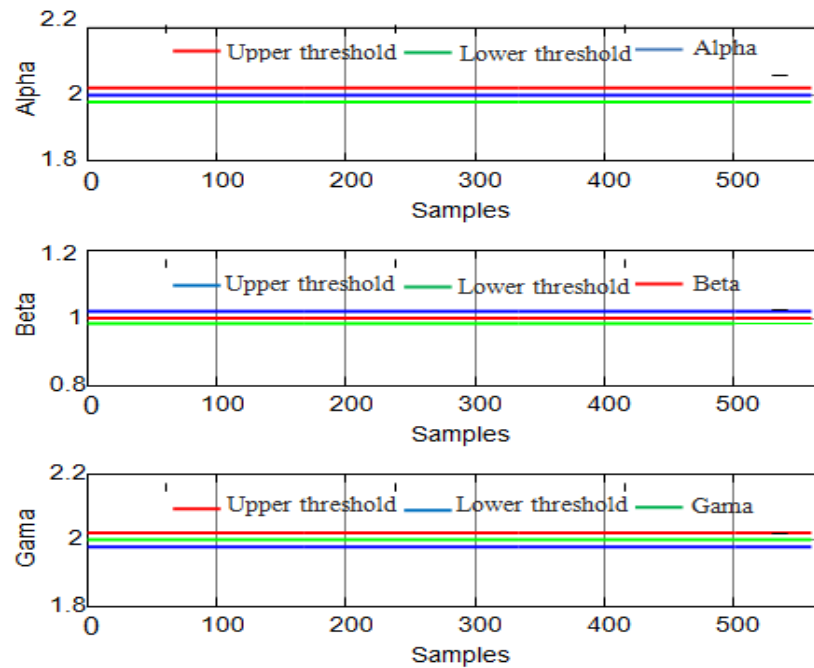
Figure 4.6: Actual measured coefficients against their variations boundaries for the completely disconnected string faulty case.

As can be seen from the simulation results, for all the simulated faulty operations, the three measured coefficients obtained from the online step are inside their boundary sets obtained from the offline step. Therefore, these results show clearly the efficiency of this strategy to detect and diagnose the aforementioned faults.

4.3. <u>Control charts-based method:</u>

Here, an efficient approach for faults detection and diagnosis is suggested [11, 62, 63]. This approach combines the flexibility and simplicity of the ODM with the extended capacity of an exponentially weighted moving average (EWMA) control chart to detect incipient changes in a PV system. The one-diode model, which is easily calibrated due to its limited calibration parameters, is used to predict the healthy PV array's maximum power and its coordinates (current 'I_{mpp}' and voltage 'V_{mpp}') using measured temperatures and irradiances. Residuals, which capture the difference between the measurements and the predictions of the ODM, are generated and used as fault indicators. Then, the EWMA monitoring chart is applied on these residuals to detect fault occurrence and identify its type. Actual data, from the previously described Algerian GCPV system, are used to assess the performance of the proposed approach. Results show that the proposed

approach successfully monitors the DC side of PV systems and detects the DC side faults.

4.3.1. Fundamental principle:

The overarching goal of this strategy is to improve the PV system efficiency by monitoring its DC side in an efficient manner. Traditionally in manufacturing industries, statistical quality control is used for monitoring and controlling product quality. Furthermore, statistical process control charts can provide early warnings of the abnormal changes in the system operations, helping the operators to identify the onset of potential faults, such as short-circuits faults, open-circuits faults, sensor bias and shading faults. These statistical charts include Shewhart, cumulative sum (CUSUM) [64-66], and EWMA charts.

The univariate statistical methods, such as the Shewhart and EWMA charts , have been widely used to monitor industrial processes for many years [11, 62]. These methods are briefly introduced here.

4.3.1.1. Shewhart monitoring chart:

In a Shewhart chart, a sequence of samples (denoted as x_i) is plotted against time. Upper and lower control limits for these samples are established around the process mean (μ) based on the three-sigma rule, i.e., UCL \ LCL = μ_0 ± $3\sigma_0$, where σ_0 is the standard deviation of the fault-free (healthy system) data computed when the process is running under healthy conditions.

Whenever the most recent measured point or a consecutive sequence of points is outside the control limits, an abnormal condition is encountered, and attention is focused on diagnosing the source of the problem. Here, the **Shewhart** chart is used as a **benchmark** for fault detection and diagnosis in the DC side of a PV system. In the next section, EWMA chart and its use in fault detection and diagnosis will be briefly described.

4.3.1.2. EWMA monitoring chart:

In this section, the basic idea of the EWMA chart and its properties are introduced. For a more detailed discussion about its design, implementation and properties, refer to [11, 58]. EWMA is constructed based on the exponential weighting of available observations, a design that provides improved sensitivity

to small changes in the mean of a multivariate process. EWMA charts are able to detect small shifts in the process mean because its statistic is a time weighted average of all previous observations. This chart was firstly introduced by Roberts [12, 58] and it has been extensively used in time series analysis. It is an anomaly-detection technique widely used by scientists and engineers in various disciplines [63, 67, 68].

Assume that $\{x_1, x_2, ..., x_n\}$ are individual observations collected from a monitored process. The expression for EWMA is given as follows:

$$\begin{cases} z_t = \lambda x_t + (1-\lambda) z_{t-1} & \text{if } t > 0 \\ z_0 = \mu_0 \end{cases} \tag{4.7}$$

The starting value, z_0, is usually set to be the mean of the fault-free data, μ_0. z_t is the output of the EWMA chart and x_t is the observation from the monitored process at the current time.

The forgetting parameter, $\lambda \in [0,1]$, determines how fast EWMA forgets the data history.

From equation (4.7), it is easy to see that:

$$z_t = \lambda x_t + (1-\lambda)\overbrace{\left[\lambda x_{t-1} + (1-\lambda) z_{t-2}\right]}^{z_{t-1}}$$

$$= \lambda x_t + \lambda(1-\lambda) x_{t-1} + (1-\lambda)^2 z_{t-2}$$

Using equation (4.7) recursively, we find that the EWMA is a linear combination of the observations:

$$z_n = \lambda x_n + \lambda(1-\lambda) x_{n-1} + \lambda(1-\lambda)^2 x_{n-2} + ... + \lambda(1-\lambda)^{n-1} x_1 + \lambda(1-\lambda)^n \mu_0 \tag{4.8}$$

Equation (4.8) can also be written as follows:

$$z_t = \lambda \sum_{t=0}^{n} (1-\lambda)^{n-t} x_t + (1-\lambda)^n \mu_0 \tag{4.9}$$

Where: $\lambda(1-\lambda)^{n-t}$ is the weight for x_t, which decreases off exponentially for the past observations.

In other words, as time passes, the smoothed statistic z_t becomes the weighted average of a greater and greater number of the past observations x_{t-n}, and the weights assigned to the previous observations are in general proportional to the terms of the geometric progression $\left\{\lambda, \lambda(1-\lambda), \lambda(1-\lambda)^2, \lambda(1-\lambda)^3, ...\right\}$.

A geometric progression is the discrete version of an exponential function, so this is where the name of this method is originated. The weighting for older data point decreases exponentially, giving much more important to the recent observation while still not discarding the older observation entirely. It can be seen that if λ is small, more weight is assigned to the past observations and the chart is efficient at detecting small changes in the process mean. On the other hand, if λ is large, more weight is assigned to the current observation and less weight is assigned to its previous observations. The chart is thus able to detect large shifts [62]. In the special case when $\lambda=1$, EWMA is equal to the most recent observation, x_t, and provides the same results as the Shewhart chart provides.

As λ approaches zero, EWMA approximates the CUSUM criteria, which give equal weight to the historical observations.

Under fault-free conditions, the standard deviation of z_t is defined as [11, 62]:

$$\sigma_{z_t} = \sigma_0 \sqrt{\frac{\lambda}{2-\lambda}\left[1-\left(1-\lambda\right)^{2t}\right]}$$ (4.10)

Where: σ_0 is the standard deviation of the fault-free or preliminary data. Therefore, in such cases, $z_t \sim N\left(\mu_0, \sigma_{z_t}^2\right)$. However, in the presence of a mean shift at the time point, the chart statistic will be defined as follows [11, 62]:

$$z_t = N\left(\mu_0 + \left[1-\left(1-\lambda\right)^{n-\tau+1}\right]\left(\mu_1 - \mu_0\right), \sigma_{z_t}^2\right)$$ (4.11)

It can be seen from equation (4.11) that the mean of the EWMA statistic in the presence of faults is a weighted average of μ_0 and μ_1, and the weight of μ_1 is larger when n is larger. Therefore, the EWMA statistic, z_t, indeed contains useful information about the mean shift.

The upper and lower control limits of the EWMA chart for detecting a mean shift are:

$$UCL \, / \, LCL = \mu_0 \pm L\sigma_{z_t} \tag{4.12}$$

where L is a multiplier of the EWMA standard deviation, σ_{zt}. The parameters L and λ need to be set carefully [11, 62]. L is usually specified in practice to be 3, which corresponds to a false alarm rate of 0.27% implying that 99.73% of the observations should be contained within the control limits in normal conditions. The value of λ is usually set between 0.2 and 0.3 [11].

On the other hand, from equation (4.10), it can be seen that the term $[1 - (1 - \lambda)^{2t}]$ converges to unity as t gets larger. In practice, people often use the asymptotic variance of z_t, $\tilde{\sigma}_{z_t}^2 = \dfrac{\lambda}{2 - \lambda}\sigma_0^2$.

Of course, If z_t is within the [LCL,UCL] interval, then it will be concluded that the process is under control up to time point t. Otherwise, the process is considered out of control.

4.3.1.3. ODM-based EWMA for faults detection and diagnosis:

In general, the model is firstly built and then a fault diagnosis procedure is accordingly performed. The estimation of the residuals, which is crucial in this method, depends on the appropriate system modeling. Once the ODM is built, based on the healthy system data, and validated, it can be used to monitor the future deviations in the system. Here, the advantages of the ODM with those of the EWMA monitoring chart are combined, which should result in an improved fault detection and diagnosis strategy, especially for detecting small changes. Specifically, in this approach, the EWMA chart is employed for fault detection and diagnosis to indicate how well the measurements conform to the model or how large the deviation from the normal model is.

Towards this end, the EWMA chart is applied to monitor residuals obtained from the ODM (see Figure (4.7)).

The differences between the real measured and predicted MPP current, MPP voltage and MPP power (obtained from the simulation model) are the residuals that can be used as indicators to detect and diagnose the possible faults.

$$\tilde{I}_t = I_t - \hat{I}_t, \quad \tilde{V}_t = V_t - \hat{V}_t, \quad \tilde{P}_t = P_t - \hat{P}_t, \quad t \in [1, n] \tag{4.13}$$

where: I_t and $\hat{I}_t$ are the measured and the predicted MPP current, respectively; V_t and $\hat{V}_t$ are the measured and predicted MPP voltage, respectively; P_t and $\hat{P}_t$ are the measured and the predicted peak power, respectively.
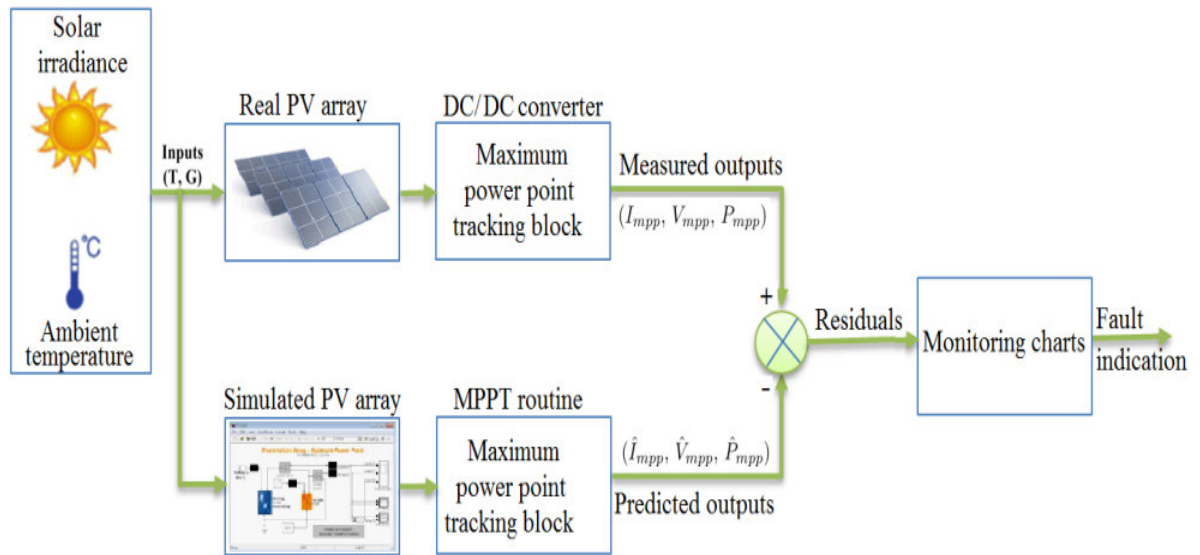
Figure 4.7: The flowchart of the proposed strategy
of faults detection and diagnosis.

In this work, the residuals are used as fault indicators. Indeed, under normal operation, the residuals are close to zero due to measurement noise and errors, while they significantly deviate from zero in the presence of abnormal events. The implementation of the developed monitoring methods comprises two stages: (i) an offline modeling and (ii) an online monitoring. In the offline modeling phase, the ODM is used on the normal operating data (training data), enabling us to obtain a reference model. Then, the fault detection procedure is executed by using the reference simulated model with the EWMA chart in the online monitoring phase.

The ODM-EWMA fault detection and diagnosis algorithm is summarized as follows:

Given:

- Real measurements of irradiance and module temperature.

- Real measurements of MPP (I_{mpp} and V_{mpp}) collected from the PV plant under normal operating condition.

- The PV module electrical characteristics obtained from the manufacturer datasheet.

Build the ODM using the fault-free training data:

- Extract the ODM electrical parameters based on the measured cell temperature and irradiance, collected from the monitored PV system under normal operating conditions, using the best-so-far ABC algorithm.

- The extracted parameters are then used to simulate the PV system behavior (PSIM^{TM}/Matlab^{TM} co-simulation).

- Compute the residuals between the measured and the predicted DC current, DC voltage and DC power (I_{mpp}, V_{mpp} and P_{mpp}) from the constructed model using fault-free data.

- Compute the control limits for the EWMA chart using equation (4.12).

Test the new data:

- Generate residual vectors, $\tilde{I}$, $\tilde{V}$ and $\tilde{P}$, by using the simulation model.

- Compute the EWMA monitoring statistic for the new data using equation (4.7).

Check for faults:

- Declare a fault when the EWMA decision statistic for the new data exceeds the control limits.

To improve the system operations, we want not only to monitor the system in an efficient manner, but also to identify the type of fault that results in any degradation of the PV system, including declines in operation reliability, and profitability, such that we can accordingly respond by making any necessary correction to the system.

Towards this end, the EWMA chart is applied on the residual of output DC power to detect the presence of faults. Then, the type of fault is identified by

analyzing the monitoring results of the EWMA chart when it is applied to the residuals of output DC current and voltage. The fault identification procedure is summarized in Figure (4.8).

The proposed strategy tests at the first stage the DC output power to detect a fault. This choice is mainly due to the fact that faults affect it inevitably. Thus, it is used as the fault indicator in the detection phase. On the other hand, both DC output current and voltage are unsuitable to be used as sensitive indicators in this phase. For example, when a short circuit occurs in one PV module from a string, current indicator value will not be significantly changed from its healthy set point. Meanwhile, a substantial change will appear in the power indicator (peak of power).

Besides, the same situation occurs when a string is completely disconnected. Indeed, the DC output voltage remains unaltered regarding its healthy status in contrast to the peak power, which will decrease significantly. Moreover, the ranking of the current and voltage indicators is unimportant in the fault diagnosis phase.
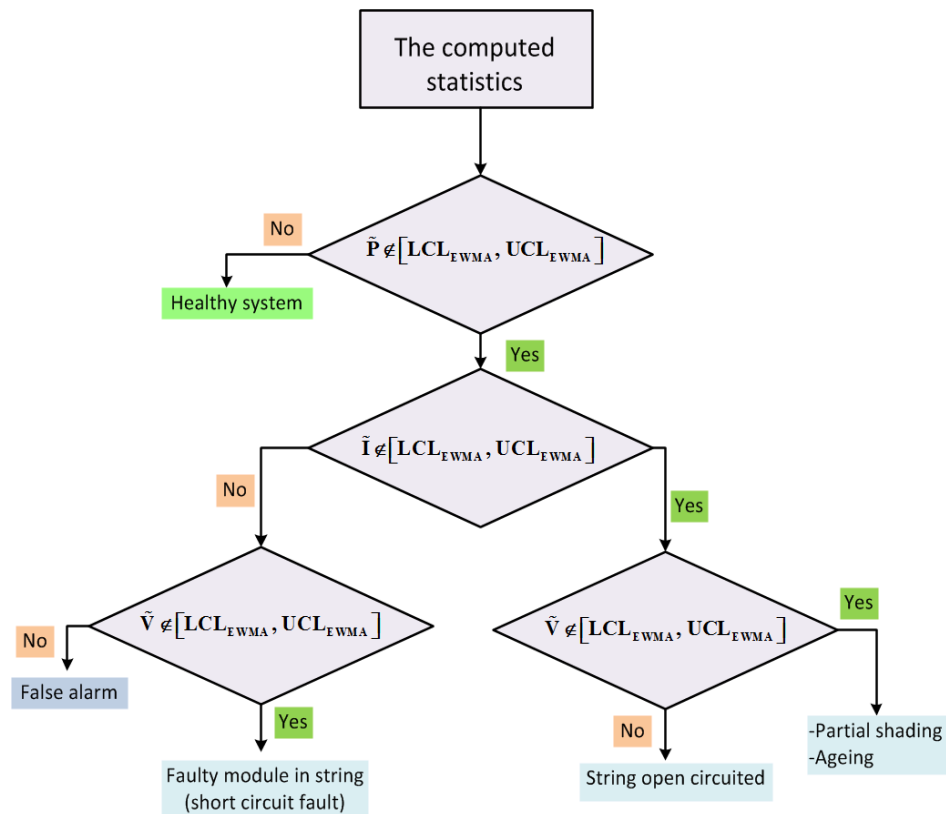
Figure 4.8: The fault identification procedure.

4.3.2. ODM-based monitoring charts results of fault detection and diagnosis:

The proposed fault detection scheme is validated using practical data collected from the previously described Algerian GCPV system. In this section, the ability of the EWMA chart to detect the presence of faults in the data and to identify the type of detected fault is assessed. To this end, three cases, involving different types of faults, were conducted. In the first case, it is assumed that the PV system contains one or more short-circuited PV modules. In the second case, an open-circuit PV string is considered. In the third case, the monitored PV system is exposed to temporary shadowing.

4.3.2.1. Normal operating condition:

Monitoring results from ODM-based Shewhart chart under normal operating conditions are shown in Figure (4.9(a-c)), monitoring results from the EWMA chart under normal operating conditions are presented in Figure (4.10(a-c)). Since the Shewhart plots for current, voltage and power shown in Figure (4.9(a-c)) are based on normal operating data; we expect that almost all the data will lie within the lower and upper control limits. Similarly, the data points in the EWMA charts are also within the confidence limits (see Figure (4.10(a-c))). It can be concluded that the ODM model describes the data well when no faults are presented.
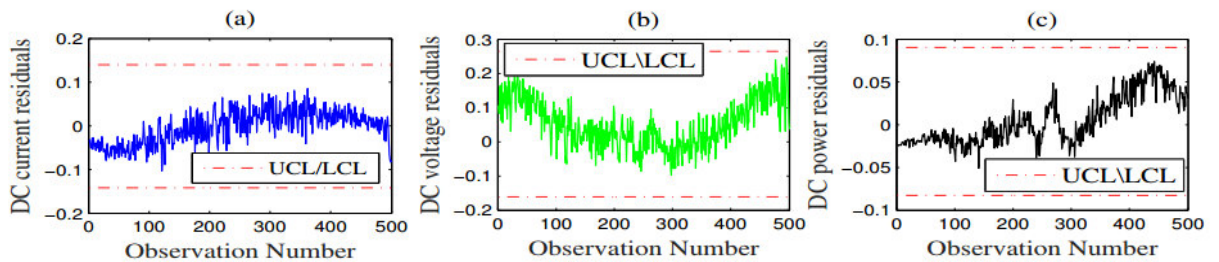
Figure 4.9: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) under normal operating conditions.
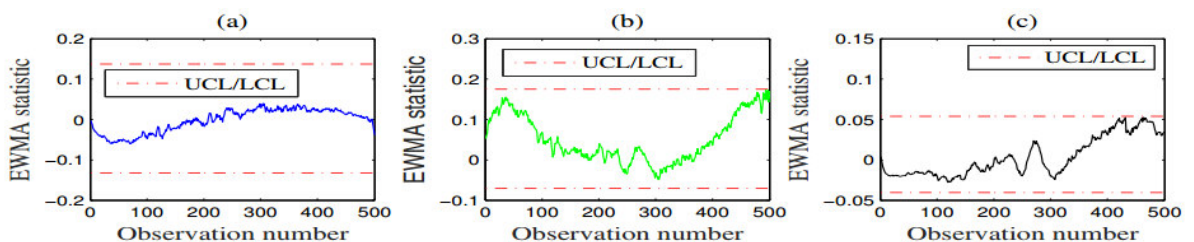
Figure 4.10: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) under normal operating conditions.

4.3.2.2. <u>Case of open-circuit PV string:</u>

In this case, the performance of the two monitoring charts when there is an open-circuit fault is investigated. To do so, an open-circuit fault in a PV array is introduced by disconnecting the second string from the monitored PV system (see fault #1 in Figure (4.11)) between sample times 300-500. To monitor the PV system, the residuals (i.e., $\tilde{I}$, $\tilde{V}$ and $\tilde{P}$) are firstly computed. Then both monitoring charts, Shewhart and EWMA, are used for fault detection and diagnosis.

The Shewhart and EWMA charts based on the MPP residuals of current, voltage and power are presented in Figures (4.12-4.13) ,respectively. The shaded area is the region where the fault is introduced. The plots in Figures (4.12(c)-4.13(c)) indicate that before the occurrence of the fault, both charts are within the lower and upper control limits. The PV system is thus working normally. For this case, the two charts can both give fault signals because the introduced fault is quite large.

Figures (4.12(c)-4.13(c)), show that the Shewhart and EWMA charts based on the output power residuals, $\tilde{P}$, significantly decrease and exceed the lower control limits, indicating that there is a significant power loss. Since one of the two strings of the PV array is disconnected at this fault, a large amount of power (nearly 50% of the rated power) is lost. After detecting the presence of a fault, the monitoring results related to the output DC current and voltage are analyzed to identify the type of fault.

Both Figures (4.12(b)-4.13(b)) are within the lower and upper control limits before and after the fault, which means that the DC voltage is almost the same after the occurrence of this open-circuit fault. The two monitoring charts based on the current residuals are given in Figures (4.12(a)-4.13(a)). These figures show that both charts exceed the lower control limits, indicating the presence of a faulty string (open-circuit fault). Indeed, the current of the faulty string drops to zero when the string is disconnected from the PV array. As a result, the residuals, which indicate the difference between the simulated and measured DC current, immediately decrease after the occurrence of the open-circuit fault. From this case, it can be seen that the open-circuit fault in a PV array increases the power loss, reduces the array current and results in almost the same array voltage as

the normal PV array voltage. These results indicate the efficiency of these charts in detecting and diagnosing open-circuit faults in a PV system.
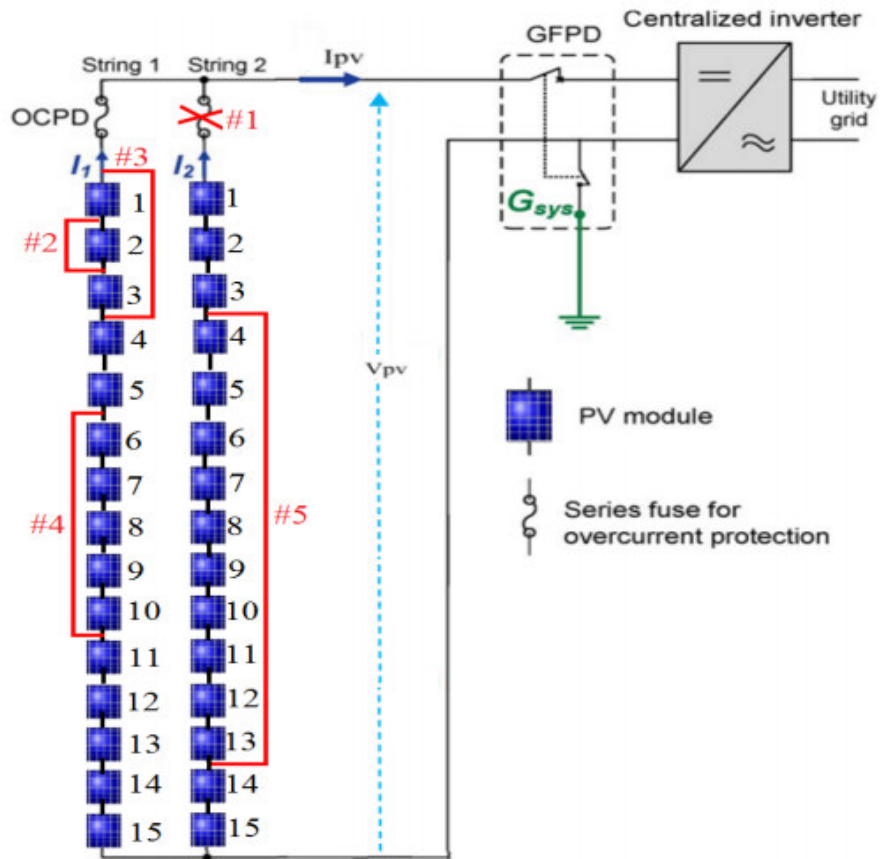
Figure 4.11: Open-circuit and short-circuits faults

Figure 4.12: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of an open-circuit fault.

Figure 4.13: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of an open-circuit fault.

4.3.2.3. <u>Case of a short-circuit fault in a PV string:</u>

In this case, the detection of short-circuited PV modules in the monitored PV system is investigated. Four examples are given in this study (see Figure (4.11), faults #2-#5).

a- <u>One short-circuited PV module:</u>

In the first example, the second module of the first string is short-circuited from the observation number 300 until the end of the testing data (see Figure (4.11), fault #2). The output DC current, voltage and power were monitored using Shewhart and EWMA charts. The two monitoring charts are shown in Figure (4.14(a-c)) and Figure (4.15(a-c)). Figure (4.14) shows that the Shewhart chart cannot detect this fault. In fact, Shewhart chart is insensitive to this fault because it is designed to detect relatively moderate and large faults, while the fault in this case is quite small. This is mainly due to the fact that the Shewhart chart uses only observed data at a particular instant to make a decision about the process performance and it ignores past data. On the other hand, the plot in Figure (4.15(c)) shows clearly the capability of the EWMA monitoring chart in detecting this small fault. From the plots in Figures (4.15(a-b)), it can be seen that the DC current residuals are within the control limits, while the DC voltage residuals exceeds the lower control limit. Thus, we can conclude that the detected fault is related to a faulty module in the string. This case shows clearly the superiority of the EWMA over the Shewhart chart in detecting small faults.

Figure 4.14: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of one short-circuited module.

Figure 4.15: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of one short-circuited module.

b- <u>Three short-circuited PV modules:</u>

In the second example, three modules have been short circuited in the first string (see Figure (4.11), fault #3). The monitoring results of the Shewhart and EWMA charts are shown in Figures (4.16-4.17), respectively. The performance of the Shewhart chart when it is applied to the output power residuals is presented in   Figure (4.16(c)), which shows that the Shewhart statistic clearly violates the lower control limit. The Shewhart chart detects this fault (i.e., a power loss) but it misses some data. On the other hand, the plot in Figure (4.17(c)) shows clearly the capability of the EWMA monitoring chart in correctly detecting this moderate fault without missed data. This short-circuit fault degrades the performance of the monitored systems and leads to a significant power loss (i.e., approximately 15% power loss). After detecting the fault based on the output DC power, the two monitoring charts based on residuals of output DC current and voltage, which are shown in Figures (4.16(a-b)-4.17(a-b)), can provide more information about the type of fault. Both Figures (4.16(b)-4.17(b)) show fault signals because the decrease in the output DC voltage in this case is quite large. The output DC current from the array does not change by much. Because the output DC voltage decreases compared to the output DC voltage of the normal array and the output DC current does not change by much, we then conclude that this fault is a short circuit in the PV array.
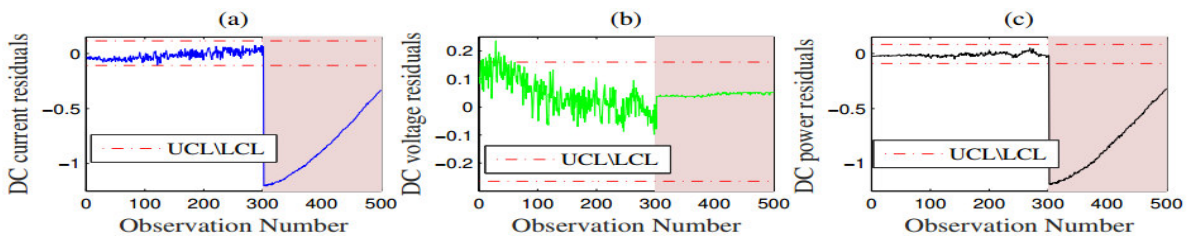

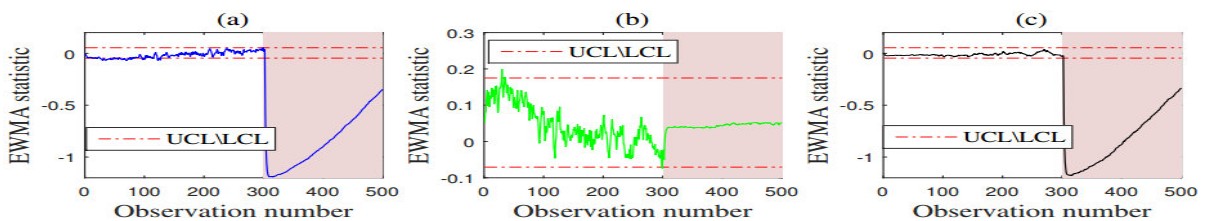Figure 4.16: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of three short-circuited modules in a PV array.

Figure 4.17: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of three short-circuited modules in a PV array.

c- Five short-circuited PV modules:

In the third example, five modules in the first string are short-circuited (Figure (4.11), fault #4). This fault leads to a power loss of 30% compared to the healthy PV array. Both monitoring charts can detect this quite large fault as shown in Figures (4.18(c)-4.19(c)). Similar to the above cases, to identify this fault, we look at the monitoring results related to the array voltage and current (see Figures (4.18(a-b)-4.19(a-b)). In fact, it is a fault that corresponds to a short-circuited PV module, since both charts based on voltage are below the control limits (see Figure (4.18(b)-4.19(b)), and the current does not change by much (see Figure (4.18(a)-4.19(a)). This demonstrates the effectiveness of the proposed strategy in detecting and diagnosing faults related to five short-circuited modules.

Figure 4.18: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of five short-circuited modules in a PV array.
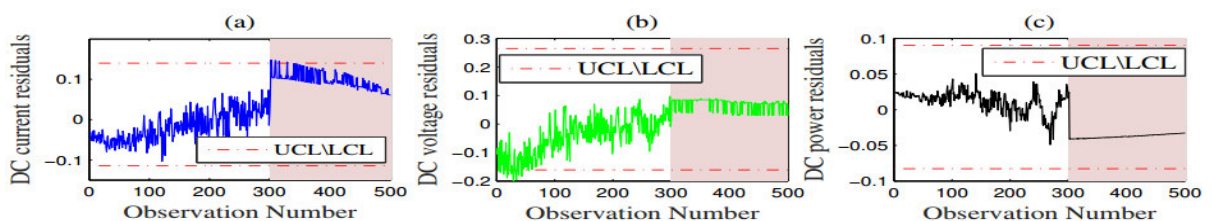

Figure 4.19: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of five short-circuited modules in a PV array.

d- Ten short-circuited PV modules:

In the fourth example, ten PV modules in the second string of the monitored PV array were short-circuited (see Figure (4.11), fault #5). Indeed, the fault resulted in large voltage drops and significant power loss (i.e. nearly 63%). Both monitoring charts can clearly detect and identify this quite large fault (see Figures (4.20(a-c)-4.21(a-c)).

Figure 4.20: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of ten short-circuited modules in a PV array.
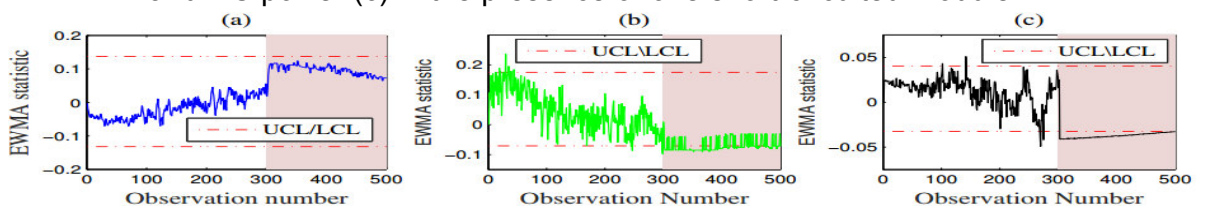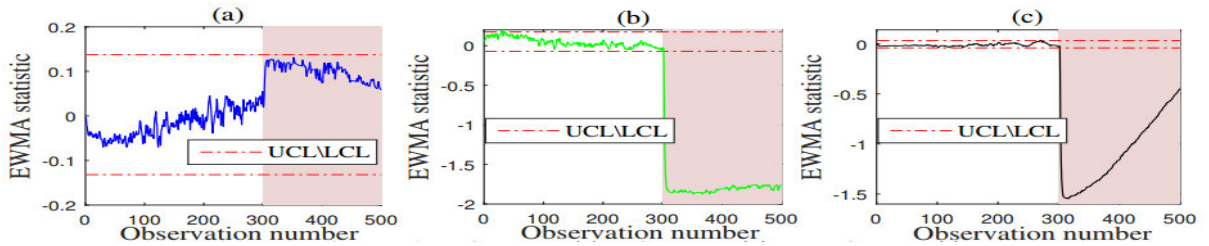

Figure 4.21: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of ten short-circuited modules in a PV array.

4.3.2.4. Case of temporarily shading fault:

The aim of this study is to assess the potential of the proposed ODM-based EWMA method to detect and identify temporarily shading faults in a PV system.

a- Temporarily shading of four PV modules

In this example, the first four PV modules of the second string were temporarily shaded (see Figure (4.22), fault #6), between samples 150 and 250. Figures (4.23-4.24) show that both charts can detect and identify this fault. This type of fault may cause decreases in current and voltage and significant power loss.
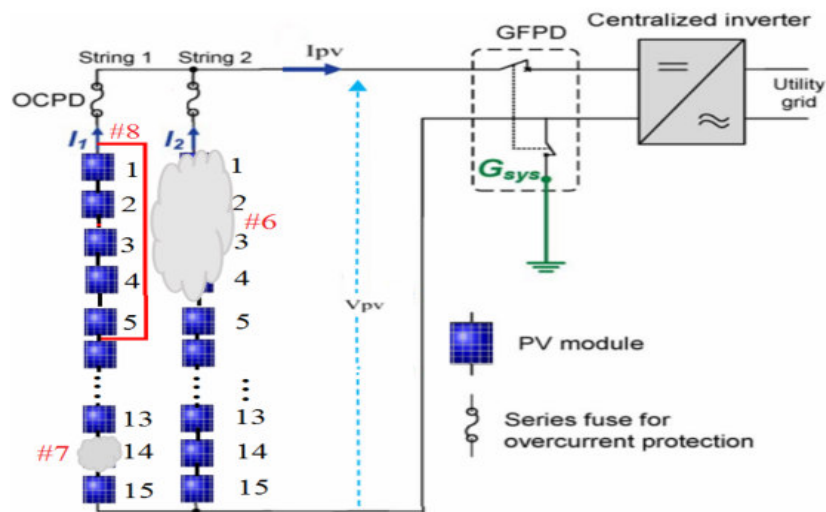
Figure 4.22: Typical faults in a PV array: temporarily shading and faulty modules.
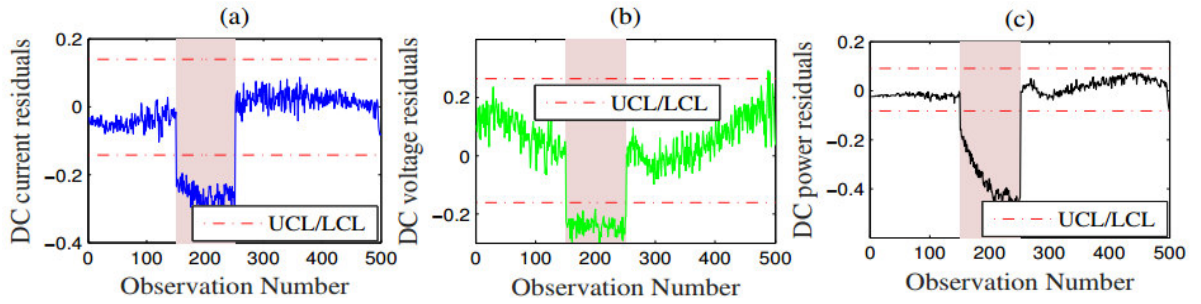
Figure 4.23: Monitoring results of Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of four PV modules temporarily shaded in the PV system.
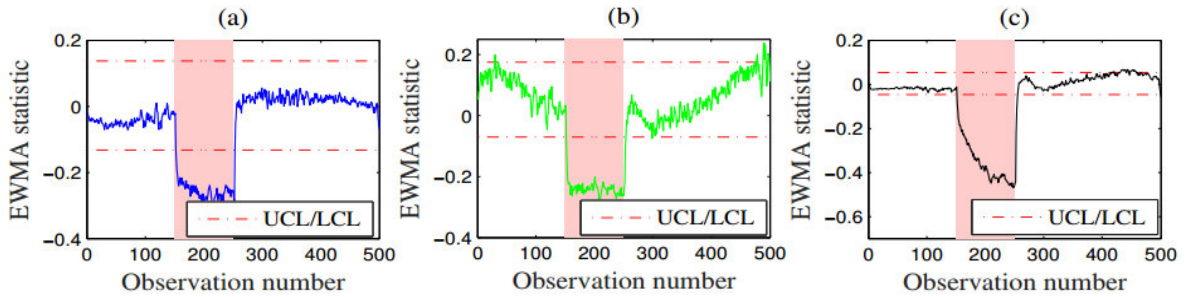
Figure 4.24: Monitoring results of EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of four PV modules temporarily shaded in the PV system.

b- Temporarily shading of one PV module:

In this example, the module #14 of the first string of the PV system was temporarily exposed to a fully shading (see Figure (4.22), fault #7), between samples 150 and 250. The shewhart chart fails to detect this fault, as shown in Figure (4.25(a-c)). Figure (4.26(c)) shows that the EWMA chart is able to detect the fault, but it cannot identify its type (see Figure 4.26(b-c)).
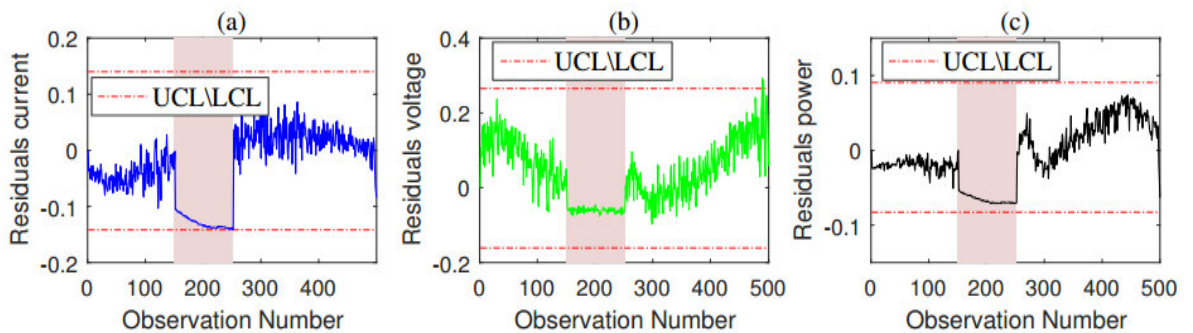
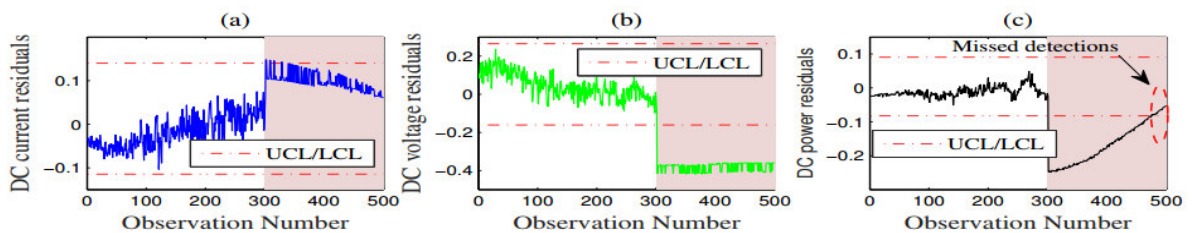

Figure 4.25: Monitoring results of Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of one PV module fully shaded in the PV system.
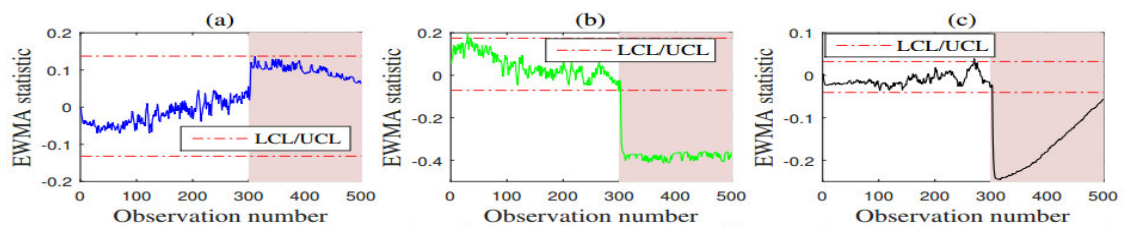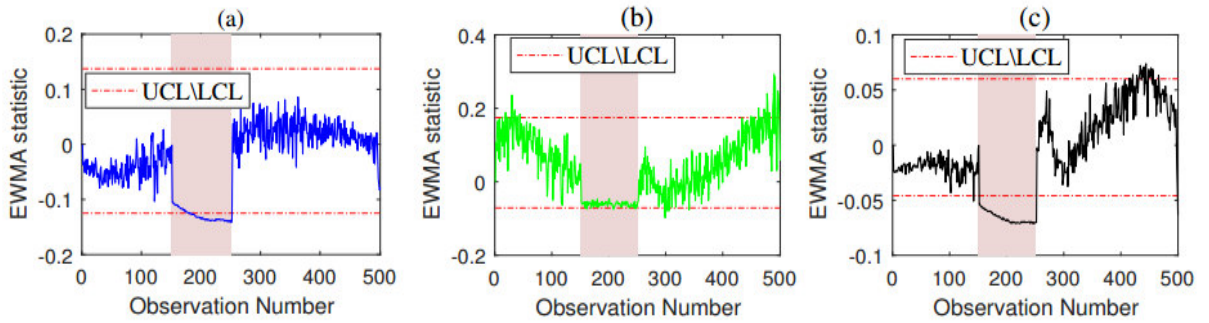
Figure 4.26: Monitoring results of EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of one PV module fully shaded in the PV system.

4.3.2.5. <u>Case of multiple faults:</u>

To assess the capacity of the proposed method to detect multiple faults, four modules in the PV system were exposed to partial shading (Figure (4.22), fault #6), between samples 150 and 250, and then five modules in the first string were short-circuited (Figure (4.22), fault #8). Monitoring results of the Shewhart and EWMA charts are illustrated in Figures (4.27-4.28), respectively. In this case, both charts can accurately detect and identify these multiple faults.
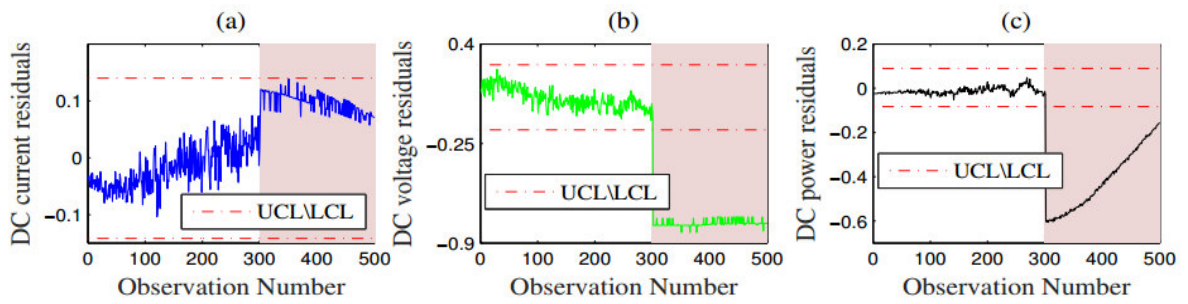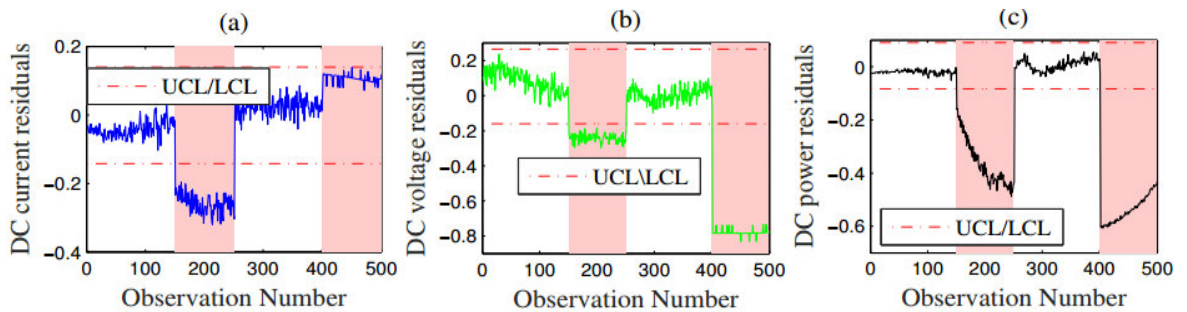

Figure 4.27: Monitoring results of a Shewhart chart for DC current (a), DC voltage (b) and DC power (c) in the presence of four PV modules that are partially shaded and five short-circuited modules in the PV system.
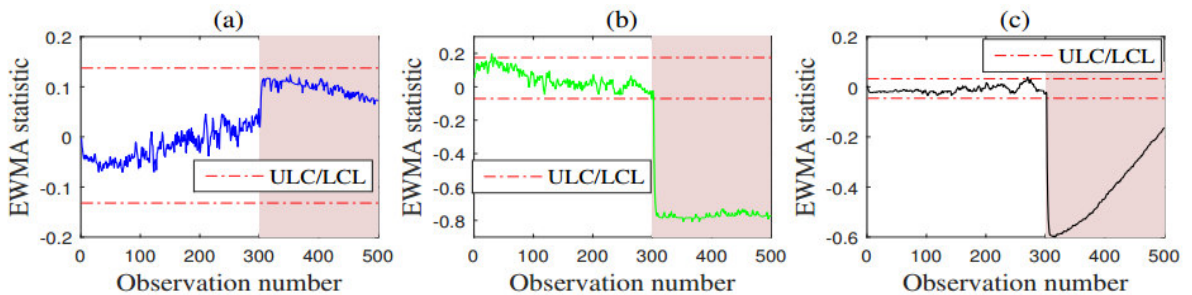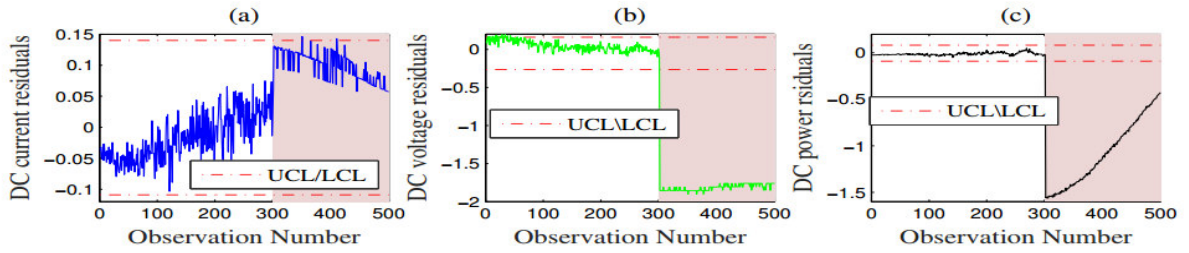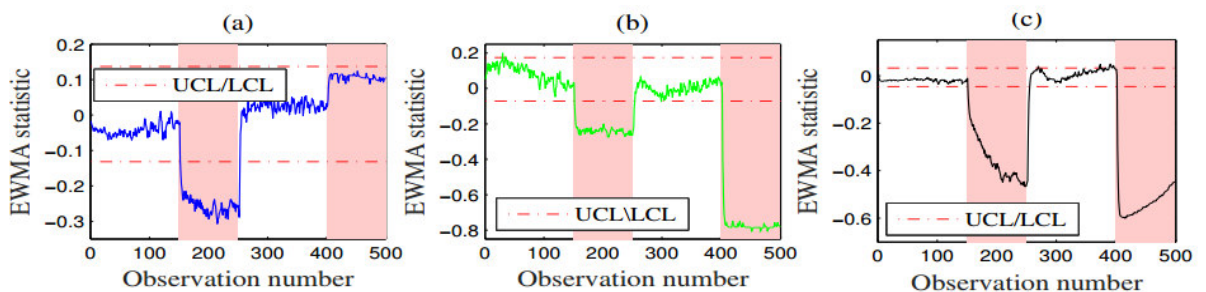

Figure 4.28: Monitoring results of a EWMA chart for DC current (a), DC voltage (b) and DC power (c) in the presence of four PV modules that are partially shaded and five short-circuited modules in the PV system.

4.4. <u>Conclusion:</u>

In this chapter, two statistical strategies of faults detection and diagnosis have been proposed. The first one consists on applying an improved ratio-based method to detect short-circuits and open-circuits faults in PV systems. The second strategy employs two statistical control charts, EWMA and Shewhart charts, to detect and diagnose faults in PV systems. This strategy provides satisfactory results in detecting and identifying short-circuit faults, open-circuit faults, and shading occurrence. In addition, to detect small changes (e.g., one short-circuited module in a string), the EWMA chart is more effective.

CONCLUSION AND PERSPECTIVES

The developed strategies carried out in this research deal with faults detection and diagnosis in GCPV systems. In fact, the main goal of this research was to improve the PV system efficiency and protect it from potential faults, by developing efficient strategies of fault detection and diagnosis. More particularly, the use of the artificial intelligence tools (such neuronal classifiers and statistical methods) to deal with this concern was the thesis principal goal.

However, using the artificial intelligence tools to deal with this problem, require the availability of a high-quality database that can, in one side, express the relationship between faults and PV system parameters, and in the other side, describes very well the system behavior for both healthy and faulty operations.

In practical point of view, obtaining such a database cannot often be guaranteed. Indeed, operating a PV system under some types of anomalies can lead to dangerous situations and even catastrophic damages. Therefore, the best way to deal with this concern was to develop an accurate simulation model that well mimics the PV system behavior for both healthy and faulty operations. In this thesis, a PSIM^{TM}/Matlab^{TM} co-simulation strategy has been developed to elaborate this model of simulation

In addition, the developed simulation model requires the use of the ODM five electrical parameters. For this reason, an efficient strategy, based on the ABC and the best-so-far ABC algorithms, has been developed to extract the ODM parameters. These algorithms have been utilized due to their efficiency in solving optimization problems, their convergence speeds and there simplicity in terms of real time implementation. The efficiency of the ODM parameters extraction stage has been experimentally validated using several PV modules of different technologies.

After that, the extracted ODM parameters have been used to develop an efficient strategy of MPP estimation. This strategy has been experimentally validated using real measurements collected from Algerian and Spanish GCPV systems. The efficiency test has been carried out for clear sky and cloudy sky conditions.

Based on the elaborated database, the next step was the development of a neuronal based strategy of fault detection and diagnosis. In fact, two PNN classifiers have been developed to deal with this concern. The first one was dedicated to the fault detection, while the second was responsible of the fault diagnosis. In addition, to test the efficiency of this strategy, the PNN classifiers have been compared, under real operating conditions, with the feed-forward back-propagation ANN classifiers method. This strategy has been validated using experimental measurements and simulation data. The obtained results have shown the high effectiveness of this strategy for noiseless and noisy data cases.

Finally, two statistical strategies for faults detection and diagnosis have been developed. The first strategy is the improved-ratio based method and it has been used to detect and diagnose open-circuit and short-circuit faults. The second strategy is the control-charts based method for fault detection and diagnosis. This strategy uses two control charts, EWMA and Shewhart charts, to detect and diagnose the faults. The efficiency of the statistical strategies has been experimentally validated.

Despite the promising results for fault detection and diagnosis, the strategies carried out in this thesis raise a number of questions and provides some directions for future works. In particular, the following points merit a serious consideration:

- The developed control-charts based method can be extended to be able to detect the number of open-circuited and short-circuited PV modules.
- In this thesis, the faults diagnosis is carried out using the climatic conditions and the MPP coordinates of current and voltage. However, using only these data does not allow the discrimination of large number of faults. To bypass this shortcoming, more input data, such as open circuit voltage (V_{oc}), short circuit current (I_{sc}) and fill factor (FF), can be added to the developed strategies to detect more categories of faults.
- The developed strategies can also be used to detect and diagnose AC side faults.

APPENDIX 'A'

LIST OF ABBREVIATIONS

ABC	Artificial Bee Colony
AC	Alternative Current
ANFIS	Adaptive Neuro-Fuzzy Inference System
ANN	Artificial Neural Network
CUSUM	Cumulative Sum
DC	Direct Current
DE	Differential Evolution
DUT	Device Under Test
ECM	Earth Capacitance Measurement
EWMA	Exponentially Weighted Moving Average
FET	Field Effect Transistor
FN	False Negative
FP	False Positive
GBSSL	Graph-Based-Semi-Supervised Learning
GCPV	Grid Connected PV system
GISTEL	Solar Radiation by Teledetection (Gisement solaire par télédétection)
IGBT	Insulated Gate Bipolar Transistor
LCL	Lower Control Limit
MAE	Mean Absolute Error
MLD	Multi-Level Decomposition
MLP	Multi-Layer Perceptron
MPPT	Maximum Power Point Tracking
NOCT	Normal Operating Cell Temperature
OCPD	Over Current Protection Device
ODM	One Diode Model
PCS	Power Conditioning system
PDF	Probability Density Function
PID	Potential Inducing Degradation
PNN	Probabilistic Neural Network
PS	Partial Shading

PSO	Particle Swarm Optimization
PV	Photovoltaic
RBF	Radial Basis Function
RMSE	Root Mean Square Error
ROC	Relative Absolute Change
STC	Standard Test Conditions
STD	Standard Deviation
TDR	Time Domain Reflectometry
TN	True Negative
TP	True Positive
UCL	Upper Control Limit

APPENDIX 'B'
LIST OF SYMBOLS

λ	The forgetting parameter [ ]
$\hat{I}_t$	The instantaneous predicted current [A]
$\hat{P}_t$	The instantaneous predicted power [W]
$\hat{V}_t$	The instantaneous predicted voltage [V]
C_d	The earth capacitance value for the whole transmission line [nF]
CR	The current ratio [ ]
C_T	The temperature coefficient of power [%/°C]
C_x	The earth capacitance value from the start point to the fault point [nF]
D	The length of the whole transmission line [m]
E_{dc}	The array produced energy [kWh]
E_i	The current error [A]
Ei	The residual error of DC current [A]
E_{i_ref}	The current error of the healthy system [A]
ELc	The error between the measured and simulated capture losses for the healthy system [Wh/Wp]
Ev	The residual error of DC voltage [V]
E_v	The voltage error [V]
E_{v_ref}	The voltage error of the healthy system [V]
FF	The file factor [ ]
ft	The faulty status [ ]
G	The irradiance level [W/m²]
γ	Gamma [m²]
G_0	The nominal irradiance (1000 [W/m²])
G_{meas}	The measured irradiance [W/m²]
I_0	The diode saturation current [A]
I_{ac}	The inverter output current [A]
id	The ideal status [ ]
I_{dc}	The DC side output current [A]
I_{mpp}	The current at the maximum power point [A]

I_{norm}	The normalized current [A]
I_{ph}	The light-generated current [A]
I_{sc}	The short circuit current [A]
I_{sc_ref}	The reference short-circuit current [A]
I_{str}	The PV string current [A]
I_t	The instantaneous measured current [A]
$I_{theorique}$	The simulated current [A]
IV_f	The (I-V) curve inflexion factor [ ]
K_B	The Boltzmann's constant (1.38×10^{-23} [J/K])
L,m	The Bishop adjustment coefficients[ ]
L_{array}	The array losses [W]
L_c	The total capture losses [Wh/Wp]
L_{cm}	The miscellaneous capture losses [Wh/Wp]
L_{ct}	The thermal capture losses [Wh/Wp]
Limit	The number of unsuccessful trials to produce better food source [ ]
LN	The number of onlooker bees [ ]
MCN	The maximum cycle number [ ]
MPP_f	The maximum power point factor [A^{-1}]
n	The diode ideality factor [ ]
η	The sensor efficiency [%]
Nf	The faulty norm [ ]
nm	The normal status [ ]
N_{mod}	The number of PV modules [ ]
Nn	The normal norm [ ]
N_{ss}	The number of PV modules in each string [ ]
N_{str}	The number of PV strings [ ]
Num_mod	The module number to the disconnection point [ ]
P_0	The maximum operating power, measured at STC [W]
P_{meas}	The measured power [W]
P_{mpp}	The peak power [W]
PR	The power ratio [ ]
P_{ref}	The array output maximum power [W]
P_{sim}	The simulated power [W]
P_t	The instantaneous measured power [W]

$P_{theorique}$	The simulated power [W]
q	The charge of an electron (1.602×10^{-19} [C])
R	The load [Ω]
R_s	The series resistance [Ω]
R_{Se}	The equivalent series resistance [Ω]
R_{sh}	The parallel resistance (shunt resistance) [Ω]
R_x	The residual signal of the attribute X [ ]
Si	The threshold value of current [A]
SN	The number of employed bees [ ]
ß	The voltage coefficient [ ]
ß_{Voc}	The temperature coefficient of open circuit voltage [%/°C]
Sv	The threshold value of voltage [V]
T	The PV cell temperature (in [Kelvins] or [Degrees Celsius])
T_{amb}	The ambient temperature [°C]
T_{ref}	The reference temperature (25 [°C])
V_0	The maximum operating voltage, measured at STC [V]
V_{ac}	The inverter output voltage [V]
V_b	The cell breakdown voltage of the bishop model [V]
V_{dc}	The DC side output voltage [V]
V_{mpp}	The voltage at the maximum power point [V]
V_{norm}	The normalized voltage [V]
V_{oc}	The open circuit voltage [V]
V_{oc_ref}	The reference open circuit voltage [V]
V_{oc_si}	The string's open circuit voltage [V]
VR	The voltage Ratio [ ]
V_t	The thermal voltage [V]
$V_{theorique}$	The simulated voltage [V]
W_i	The network weights [ ]
W_{max}, W_{min}	The Maximum and minimum percentage of the scout bees position [ ]
X_t	The actual observation [ ]
Y_a	The array yield [Hours]
Y_r	The reference yield [Hours]
Z_t	The control chart output characteristic [ ]

α	The current coefficient [ ]
α_{Isc}	The temperature coefficient of short-circuit current [%/°C]
δ	The smoothing parameter [ ]
ΔP	The difference of power [W]
$\Phi i,j$	Random number between [-1, +1]

PUBLICATIONS AND CONFERENCES

1. **Garoudja, E., Harrou, F., Sun, Y., Kara, K., Chouder, A., & Silvestre, S. (2017). Statistical fault detection in photovoltaic systems. Solar Energy, 150, 485-499.**
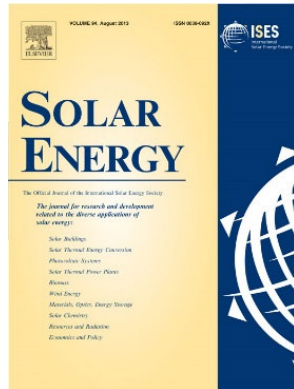
CiteScore: 4.52

Impact Factor: 4.018 (*2017 Journal Citation Reports*)

5-Year Impact Factor: 4.739 (*2017 Journal Citation Reports*)

Source Normalized Impact per Paper (SNIP): 1.957

SCImago Journal Rank (SJR): 2.063

2. **Garoudja, E., Chouder, A., Kara, K., & Silvestre Bergés, S. (2017). An enhanced machine learning based approach for failures detection and diagnosis of PV systems. Energy Conversion and Management (ECM_4383), 151, 496-513.**
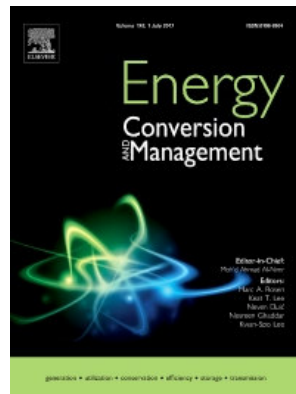
CiteScore: 6.04

Impact Factor: 5.589 (*2017 Journal Citation Reports*)

5-Year Impact Factor: 5.472 (*2017 Journal Citation Reports*)

Source Normalized Impact per Paper (SNIP): 1.948

SCImago Journal Rank (SJR): 2.156

3. **Garoudja, E., Kara, K., Chouder, A., & Silvestre, S. (2015, May). Parameters extraction of photovoltaic module for long-term prediction using artifical bee colony optimization. In Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on (pp. 1-6). IEEE.**

4. **Garoudja, E., Kara, K., Chouder, A., Silvestre, S., & Kichou, S. (2016, November). Efficient fault detection and diagnosis procedure for photovoltaic systems. In Modeling, Identification and Control (ICMIC), 2016 8th International Conference on (pp. 851-856). IEEE.**

5. **Garoudja, E., Harrou, F., Sun, Y., Kara, K., Chouder, A., & Silvestre, S. (2017, May). A statistical-based approach for fault detection and diagnosis in a photovoltaic system. In Systems and Control (ICSC), 2017 6th International Conference on (pp. 75-80). IEEE.**

REFERENCES:

[1]   Elrayyah, A., Sozer, Y.et Elbuluk, M., "Control of microgrid-connected PV-sources," in Industry Applications Society Annual Meeting (IAS), 2012 IEEE, (2012), 1-8.

[2]   Kanchev, H., Lu, D., Colas, F., Lazarov, V.et Francois, B., "Energy management and operational planning of a microgrid with a PV-based active generator for smart grid applications", IEEE transactions on industrial electronics, V. 58, n°10, (2011), 4583-4592.

[3]   Dhimish, M.et Holmes, V., "Fault detection algorithm for grid-connected photovoltaic plants", Solar Energy, V. 137, (2016), 236-245.

[4]   Alam, M. K., Khan, F., Johnson, J.et Flicker, J., "A comprehensive review of catastrophic faults in PV arrays: types, detection, and mitigation techniques", IEEE Journal of Photovoltaics, V. 5, n°3, (2015), 982-997.

[5]   Mohammedi, A., Mezzai, N., Rekioua, D.et Rekioua, T., "Impact of shadow on the performances of a domestic photovoltaic pumping system incorporating an MPPT control: A case study in Bejaia, North Algeria", Energy Conversion and Management, V. 84, (2014), 20-29.

[6]   Hu, Y., Gao, B., Song, X., Tian, G. Y., Li, K.et He, X., "Photovoltaic fault detection using a parameter based model", Solar Energy, V. 96, (2013), 96-102.

[7]   Gokmen, N., Karatepe, E., Celik, B.et Silvestre, S., "Simple diagnostic approach for determining of faulted PV modules in string based PV arrays", Solar Energy, V. 86, n°11, (2012), 3364-3377.

[8]   Nordmann, T., Clavadetscher, L., van Sark, W.et Green, M., "Analysis of long-term performance of PV systems," ed: International Energy Agency, 2015.

[9]   Specht, D. F., "Probabilistic neural networks and the polynomial adaline as complementary techniques for classification", IEEE Transactions on Neural Networks, V. 1, n°1, (1990), 111-121.

[10]   Specht, D. F., "Probabilistic neural networks for classification, mapping, or associative memory," in IEEE international conference on neural networks, (1988), 525-532.

[11]   Montgomery, D. C., "Introduction to statistical quality control", John Wiley & Sons (New York), (2009),

[12]   Lucas, J. M.et Saccucci, M. S., "Exponentially weighted moving average control schemes: properties and enhancements", Technometrics, V. 32, n°1, (1990), 1-12.

[13]   Zhao, Y., Lehman, B., Ball, R., Mosesian, J.et de Palma, J.-F., "Outlier detection rules for fault detection in solar photovoltaic arrays," in Applied Power Electronics Conference and Exposition (APEC), 2013 Twenty-Eighth Annual IEEE, (2013), 2913-2920.

[14]   Garoudja, E., Harrou, F., Sun, Y., Kara, K., Chouder, A.et Silvestre, S., "Statistical fault detection in photovoltaic systems", Solar Energy, V. 150, (2017), 485-499.

[15] Hu, Y., Cao, W., Ma, J., Finney, S. J.et Li, D., "Identifying PV module mismatch faults by a thermography-based temperature distribution analysis", IEEE Transactions on Device and Materials Reliability, V. 14, n°4, (2014), 951-960.

[16] Munoz, M., Alonso-García, M. C., Vela, N.et Chenlo, F., "Early degradation of silicon PV modules and guaranty conditions", Solar Energy, V. 85, n°9, (2011), 2264-2274.

[17] Garoudja, E., Kara, K., Chouder, A., Silvestre, S.et Kichou, S., "Efficient fault detection and diagnosis procedure for photovoltaic systems," in Modelling, Identification and Control (ICMIC), 2016 8th International Conference on, (2016), 851-856.

[18] Karaboga, D., "An idea based on honey bee swarm for numerical optimization," Technical report-tr06, Erciyes university, engineering faculty, computer engineering department2005.

[19] Karaboga, D.et Basturk, B., "On the performance of artificial bee colony (ABC) algorithm", Applied soft computing, V. 8, n°1, (2008), 687-697.

[20] Banharnsakun, A., Achalakul, T.et Sirinaovakul, B., "The best-so-far selection in artificial bee colony algorithm", Applied soft computing, V. 11, n°2, (2011), 2888-2901.

[21] Banharnsakun, A.et Tanathong, S., "Object detection based on template matching through use of best-so-far ABC", Computational intelligence and neuroscience, V. 2014, (2014), 7.

[22] Razykov, T. M., Ferekides, C. S., Morel, D., Stefanakos, E., Ullal, H. S.et Upadhyaya, H. M., "Solar photovoltaic electricity: Current status and future prospects", Solar Energy, V. 85, n°8, (2011), 1580-1608.

[23] Pappas, C., Karakosta, C., Marinakis, V.et Psarras, J., "A comparison of electricity production technologies in terms of sustainable development", Energy Conversion and Management, V. 64, (2012), 626-632.

[24] Tadj, M., Benmouiza, K., Cheknane, A.et   Silvestre, S., "Improving the performance of PV systems by faults detection using GISTEL approach", Energy Conversion and Management, V. 80, (2014), 298-304.

[25] Bun, L., "Détection et Localisation de Défauts pour un Système PV," Université Grenoble Alpes, 2011.

[26] Picault, D., "Reduction of mismatch losses in grid-connected photovoltaic systems using alternative topologies," Institut National Polytechnique de Grenoble-INPG, 2010.

[27] Verhoeven, B., "Utility aspects of grid connected photovoltaic power systems", International Energy Agency, (1998),

[28] Silvestre, S.et Chouder, A., "Effects of shadowing on photovoltaic module performance", Progress in Photovoltaics: Research and applications, V. 16, n°2, (2008), 141-149.

[29] Pau1l, M. M. R., Mahalakshmi, R., Karuppasamypandiyan, M., Bhuvanesh, A.et Ganesh, R. J., "Classification and Detection of Faults in Grid Connected Photovoltaic System",

[30] Kichou, S., Abaslioglu, E., Silvestre, S., Nofuentes, G., Torres-Ramírez, M.et Chouder, A., "Study of degradation and evaluation of model parameters of micromorph silicon photovoltaic modules under outdoor long term exposure in Jaén, Spain", Energy Conversion and Management, V. 120, (2016), 109-119.

[31] Mekki, H., Mellit, A.et Salhi, H., "Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules", Simulation Modelling Practice and Theory, V. 67, (2016), 1-13.

[32] Bonsignore, L., Davarifar, M., Rabhi, A., Tina, G. M.et Elhajjaji, A., "Neuro-Fuzzy fault detection method for photovoltaic systems", Energy Procedia, V. 62, (2014), 431-441.

[33] Zhao, Y., Ball, R., Mosesian, J., de Palma, J.-F.et Lehman, B., "Graph-based semi-supervised learning for fault detection and classification in solar photovoltaic arrays", IEEE Transactions on Power Electronics, V. 30, n°5, (2015), 2848-2858.

[34] Chine, W., Mellit, A., Lughi, V., Malek, A., Sulligoi, G.et Pavan, A. M., "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks", Renewable Energy, V. 90, (2016), 501-512.

[35] Dhimish, M., Holmes, V., Mehrdadi, B., Dales, M.et Mather, P., "Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system", Energy, V. 140, (2017), 276-290.

[36] Jiang, L. L.et Maskell, D. L., "Automatic fault detection and diagnosis for photovoltaic systems using combined artificial neural network and analytical based methods," in 2015 International Joint Conference on Neural Networks (IJCNN), (2015), 1-8.

[37] Andrianajaina, T., Sambatra, E. J. R., Andrianirina, C. B., Razafimahefa, T. D.et Heraud, N., "PV fault detection using the least squares method," in Electrical and Power Engineering (EPE), 2016 International Conference and Exposition on, (2016), 846-851.

[38] Spataru, S., Sera, D., Kerekes, T.et Teodorescu, R., "Diagnostic method for photovoltaic systems based on light I–V measurements", Solar Energy, V. 119, (2015), 29-44.

[39] Salem, F.et Awadallah, M. A., "Detection and assessment of partial shading in photovoltaic arrays", Journal of Electrical Systems and Information Technology, V. 3, n°1, (2016), 23-32.

[40] Vergura, S., Acciani, G., Amoruso, V., Patrono, G. E.et Vacca, F., "Descriptive and inferential statistics for supervising and monitoring the operation of PV plants", IEEE transactions on industrial electronics, V. 56, n°11, (2009), 4456-4464.

[41] Chouder, A.et Silvestre, S., "Automatic supervision and fault detection of PV systems based on power losses analysis", Energy Conversion and Management, V. 51, n°10, (2010), 1929-1937.

[42] Silvestre, S., Chouder, A.et Karatepe, E., "Automatic fault detection in grid connected PV systems", Solar Energy, V. 94, (2013), 119-127.

[43] Drews, A., De Keizer, A., Beyer, H. G., Lorenz, E., Betcke, J., Van Sark, W., et al., "Monitoring and remote failure detection of grid-connected PV systems based on satellite observations", Solar Energy, V. 81, n°4, (2007), 548-564.

[44] Hariharan, R., Chakkarapani, M., Ilango, G. S.et Nagamani, C., "A method to detect photovoltaic array faults and partial shading in PV systems", IEEE Journal of Photovoltaics, V. 6, n°5, (2016), 1278-1285.

[45] Kim, I.-S., "On-line fault detection algorithm of a photovoltaic system using wavelet transform", Solar Energy, V. 126, (2016), 137-145.

[46] Schirone, L., Califano, F.et Pastena, M., "Fault detection in a photovoltaic plant by time domain reflectometry", Progress in Photovoltaics: Research and applications, V. 2, n°1, (1994), 35-44.

[47] Takashima, T., Yamaguchi, J., Otani, K., Oozeki, T., Kato, K.et Ishida, M., "Experimental studies of fault location in PV module strings", Solar Energy Materials and Solar Cells, V. 93, n°6-7, (2009), 1079-1082.

[48] Garoudja, E., Kara, K., Chouder, A.et Silvestre, S., "Parameters extraction of photovoltaic module for long-term prediction using artifical bee colony optimization," in Control, Engineering & Information Technology (CEIT), 2015 3rd International Conference on, (2015), 1-6.

[49] Garoudja, E., Chouder, A., Kara, K.et Silvestre, S., "An enhanced machine learning based approach for failures detection and diagnosis of PV systems", Energy Conversion and Management, V. 151, (2017), 496-513.

[50] Garoudja, E., Harrou, F., Sun, Y., Kara, K., Chouder, A.et Silvestre, S., "A statistical-based approach for fault detection and diagnosis in a photovoltaic system," in Systems and Control (ICSC), 2017 6th International Conference on, (2017), 75-80.

[51] Karaboga, D., Gorkemli, B., Ozturk, C.et Karaboga, N., "A comprehensive survey: artificial bee colony (ABC) algorithm and applications", Artificial Intelligence Review, V. 42, n°1, (2014), 21-57.

[52] Hamid, N. F. A., Rahim, N. A.et Selvaraj, J., "Solar cell parameters extraction using particle swarm optimization algorithm," in Clean Energy and Technology (CEAT), 2013 IEEE Conference on, (2013), 461-465.

[53] Ishaque, K., Salam, Z., Taheri, H.et Shamsudin, A., "Parameter extraction of photovoltaic cell using differential evolution method," in Applied Power Electronics Colloquium (IAPEC), 2011 IEEE, (2011), 10-15.

[54] Wagner, A., "Peak-power and internal series resistance measurement under natural ambient conditions," in Proceedings EuroSun, (2000).

[55] Banharnsakun, A., Sirinaovakul, B.et Achalakul, T., "Job shop scheduling with the best-so-far ABC", Engineering Applications of Artificial Intelligence, V. 25, n°3, (2012), 583-593.

[56] Nantapat, T., Kaewkamnerdpong, B., Achalakul, T.et Sirinaovakul, B., "Best-so-far ABC based nanorobot swarm," in Intelligent Human-Machine Systems and Cybernetics (IHMSC), 2011 International Conference on, (2011), 226-229.

[57] Dilmac, S.et Korurek, M., "ECG heart beat classification method based on modified ABC algorithm", Applied soft computing, V. 36, (2015), 641-655.

[58] Hunter, J. S., "The exponentially weighted moving average", J. Quality Technol., V. 18, n°4, (1986), 203-210.

[59]   Winkel, P.et Zhang, N. F., "Shewhart Control Charts", Statistical Development of Quality in Medicine, (2007), 37-77.

[60]   Kang, C. W.et Kvam, P. H., "Shewhart Control Charts", Basic Statistical Tools for Improving Quality, (2011), 97-124.

[61]   Koutras, M., Bersimis, S.et Maravelakis, P., "Statistical process control using Shewhart control charts with supplementary runs rules", Methodology and Computing in Applied Probability, V. 9, n°2, (2007), 207-224.

[62]   Montgomery, D. C., "Introduction to statistical quality control", John Wiley & Sons, (2007),

[63]   Cook, D., Coory, M.et Webster, R., "Exponentially weighted moving average charts to compare observed and expected values for monitoring risk-adjusted hospital indicators", BMJ quality & safety, V. 20, n°6, (2011), 469-474.

[64]   Hawkins, D. M.et Olwell, D. H., "Cumulative sum charts and charting for quality improvement", Springer Science & Business Media, (2012),

[65]   Abbas, N., Riaz, M.et Does, R. J., "Mixed exponentially weighted moving average–cumulative sum charts for process monitoring", Quality and Reliability Engineering International, V. 29, n°3, (2013), 345-356.

[66]   Beiles, C. B.et Morton, A. P., "Cumulative sum control charts for assessing performance in arterial surgery", ANZ journal of surgery, V. 74, n°3, (2004), 146-151.

[67]   Holt, C. C., "Forecasting seasonals and trends by exponentially weighted moving averages", International journal of forecasting, V. 20, n°1, (2004), 5-10.

[68]   Ross, G. J., Adams, N. M., Tasoulis, D. K.et Hand, D. J., "Exponentially weighted moving average charts for detecting concept drift", Pattern recognition letters, V. 33, n°2, (2012), 191-198.