الجمهوريـــــــــة الجزائريـــــــــة الـديمقراطـيـــــــة الشعبـيـــــــــة
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليـــــــــم العـــــــــالي و البحـــــــــث العلمـــــــــي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة يحيى فارس-المدية
Université Dr. Yahia Farès de Médéa
كلية التكنولوجيا
Faculté de la Technologie
قسم الهندسة الكهربائية
Département de Génie Electrique

THESE

Pour l'obtention du Diplôme de

Doctorat LMD

Spécialité : Automatique des Systèmes

Contribution au Développement d'un Algorithme de Détection et Diagnostic des Défauts Dans le Système Photovoltaïque fonctionnant en mode connecté au Réseau et Autonome

Présentée par :

Rabah BENKERCHA

Soutenue le 07/04/2021 devant la commission d'examen composé de :

Mr. BOUCHHIDA Ouahid	Professeur de l'université de Médea	Président
Mr. MAIDI Ahmed	Professeur de l'université de Tizi-Ouzou	Examinateur
Mr. MEROUANI Khoudir	MCA de l'Ecole Militaire Polytechnique Alger	Examinateur
Mr. KEBECHE Nadir	MCA de l'université de Médea	Examinateur
Mr. MOULAHOUM Samir	Professeur de l'université de Médea	Directeur de thèse
Mr. CHOUDER Aissa	Professeur de l'université de Msila	Co-Directeur de thèse

REMERCIEMENTS

Je remercie **Dieu** la compassion et le miséricordieux de m'avoir aidé et donné patience et courage pour accomplir ce modeste travail.

Le travail présenté dans cette thèse a été réalisé dans le **Laboratoire de Recherche en Génie Electrique et Automatique (LREA)** de l'université de Médéa.

A l'issue de ce travail je tiens à adresser ma reconnaissance et mes remerciements à toutes les personnes qui ont contribué, chacune à leur manière, à l'accomplissement de cette thèse.

Je tiens également à remercier très chaleureusement :

Mon directeur et co-directeur de thèse, monsieur **Samir Moulahoum** Professeur à l'Université de Médéa, et monsieur **Aissa Chouder** Professeur à l'Université de Msila, pour ses encadrements, ses suivi permanents et ses précieux conseils.

Le directeur de laboratoire monsieur **Nadir KABACHE** et aussi monsieur **Hamza HOUASSINE** pour leurs soutient et encouragement. Tous les membres du **Laboratoire de Recherche en Génie Electrique et Automatique (LREA)** de l'université de Médéa ;

J'exprime toute ma gratitude au professeur **Ouahid BOUCHHIDA** pour avoir accepté de juger ce travail et d'en être le président de jury.

Je tiens à remercier le professeur **Ahmed MAIDI** et le docteur **Khoudir MEROUANI** d'avoir accepté de faire partie du jury.

Je tiens aussi a remercié mes collègue les doctorants **Bilel TAGHEZOUIT, Hamza OUAZENE, Abderrahmane BENYAMINA et Oualid AISSA** pour ses aides et pour l'atmosphère de fraternité et le confort que j'ai senti durant ces années de recherche.

J'exprime mes vifs remerciements à monsieur **Noureddine Ayda et Youcef Bouchahma** pour ses aides et pour ses encouragements.

Je remercie mes parents, ainsi que toute ma famille, pour leur amour, soutient et encouragement. Ils m'ont toujours donné confiance en soi et m'ont motivé pour toujours marcher en avant. A toutes les personnes, qui ont contribué de près ou de loin, directement ou indirectement à l'aboutissement à ce travail, on les remercie tous.

Enfin, on ne pourrait terminer ces remerciements sans une pensée à l'ensemble de nos enseignants qui sont à l'origine de tout notre savoir.

RESUME

En raison de la forte demande d'électricité dans le monde, les systèmes d'installations photovoltaïques (PVs) se multiplient. Par conséquent, les défauts dans les systèmes augmentent, ce qui entraîne une perte d'énergie. Afin de minimiser les pertes d'énergie dans les systèmes PV installés, le diagnostic et la surveillance du système PV jouent un rôle très important. Objectivement, notre préoccupation porte sur le système PV connecté au réseau (SPVCR) où le but de cette thèse est de construire un modèle capable de distinguer les défauts prédéfinis qui surviennent dans le système. Dans un premier temps, la modélisation du système dans son état sain est étudiée, basée sur l'algorithme de pollinisation des fleurs pour extraire les paramètres des modèles à seule et double diodes. Ensuite, les paramètres identifiés sont utilisés pour estimer le courant, la tension et la puissance au point de puissance maximale. Plus tard, une comparaison entre les données mesurées et simulées est effectuée pour évaluer les résultats du processus d'extraction. Les résultats obtenus montrent que les données simulées correspondent bien à celles mesurées. Deuxièmement, trois types de défauts sont créés artificiellement dans le générateur PV afin d'obtenir un ensemble de données. Ces données sont collectées à partir de SPVCR par le système d'acquisition dans plusieurs conditions météorologiques. Ensuite, les données sont utilisées pour construire deux modèles, l'un pour la détection des défauts et l'autre pour le diagnostic des défauts. En d'autres termes, l'un capable de reconnaître l'état du système et le second est autorisé à classer le type de défaut. A cet effet, une nouvelle approche basée sur l'algorithme d'arbre de décision C4.5 est proposée. Trois attributs numériques et deux cibles sont choisis pour être respectivement les entrées et les sorties des modèles. Les entrées choisies sont la température ambiante, l'irradiation et le rapport de puissance, ce dernier est calculé à partir de la puissance mesurée et estimée. La puissance estimée provient du modèle SANDIA qui représente le SPVCR dans son état sain. Les sorties du modèle de détection sont des états sains ou défectueux. D'autre part, le modèle de diagnostic a quatre classes à savoir: Sans défaut, défaut de chaîne, défaut de court-circuit et défaut de ligne. Les données sont divisées en deux parties, où 66% pour le processus d'apprentissage et le reste pour le test. Les résultats des tests indiquent que les modèles ont une performance de prédiction élevée dans la détection avec une grande précision tandis que le modèle de diagnostic a une précision égale à 99%.

ABSTRACT

Due the high demand of the electricity in the world, the photovoltaic installations systems are increasing. Consequently, the faults in systems are rising which leading to the energy loses. In order to minimize energy losses in the installed PV systems, the diagnosis and monitoring of PV system plays a very important role. Objectively, our concern is on the grid connected photovoltaic system (GCPVS) where the aim of this thesis is to construct a model able to distinguish between the predefined faults that occurs in the system. Firstly, the modeling of the system in its healthy state is studied, based on the flower pollination algorithm to extract the parameters of the single and double diode models. After, the identified parameters are used to estimate the current, the voltage and the power at maximum power point. Later, a comparison between the measured and simulated data is performed to evaluate the results of extraction process. The obtained results show that the simulated data fit well the measured one. Secondly, three type of faults are created artificially in PV array in order to get a data set. These data are collected from GCPVS by the acquisition system under several weather conditions. After, the data is used to build two models, one for the fault detection and the other for fault diagnostic. In other words, one capable to recognize the system state and the second is allowed to classify the type of fault. For this purpose, a new approach based on C4.5 decision tree algorithm is proposed. Three numerical attributes and two targets are chosen to be the inputs and the outputs of the models respectively. The chosen inputs are temperature ambient, irradiation and power ratio, this later is calculated from measured and estimated power. The estimated power is got from SANDIA model that represent the GCPVS in its healthy state. The outputs of the detection model are healthy or faulty states. On other hand, the diagnostic model have four classes namely: free fault, string fault, short circuit fault and line-line fault. The data is divided into two parts, where 66% for the learning process and the remained for the test. Testing result indicate that the models have a high prediction performance in the detection with high accuracy while the diagnosis model have accuracy equal to 99%.

ملخص

نظرًا لارتفاع الطلب على الكهرباء في العالم ، تتزايد أنظمة التركيبات الكهروضوئية. وبالتالي ، فإن الأعطال في الأنظمة آخذة في الارتفاع مما يؤدي إلى فقدان الطاقة. من أجل تقليل فقد الطاقة في الأنظمة الكهروضوئية المركبة ، يلعب تشخيص ومراقبة النظام الكهروضوئي دورًا مهمًا للغاية. من الناحية الموضوعية ، ينصب اهتمامنا على النظام الكهروضوئي المتصل بالشبكة حيث تهدف هذه الأطروحة إلى بناء نموذج قادر على التمييز بين الأعطاب المحددة مسبقًا التي تحدث في النظام. أولاً ، تمت دراسة نمذجة النظام في حالته السليمة ، بناءً على خوارزمية التلقيح بالزهور لاستخراج معاملات نموذجي الصمام المفرد والمزدوج. بعد ذلك ، يتم استخدام المعلمات المحددة لتقدير التيار والجهد والطاقة عند أقصى نقطة طاقة. في وقت لاحق ، يتم إجراء مقارنة بين البيانات المقاسة و البيانات المحاكاة لتقييم نتائج عملية الاستخراج. أظهرت النتائج المتحصل عليها أن المحاكاة تتناسب مع البيانات المقاسة. ثانيًا ، يتم إنشاء ثلاثة أنواع من الأعطاب بشكل مصطنع في مصفوفة الكهروضوئية من أجل الحصول على مجموعة بيانات. يتم جمع هذه البيانات من على النظام الكهروضوئي المتصل بالشبكة بواسطة نظام الاستحواذ في ظل العديد من الظروف الجوية. بعد ذلك ، يتم استخدام البيانات لبناء نموذجين ، أحدهما لاكتشاف الأعطاب والآخر لتشخيص الأعطاب. بمعنى آخر ، يمكن للنموذج الأول أن يتعرف على حالة النظام والنموذج الثاني يسمح بتصنيف هذا العطب. لهذا الغرض ، تم اقتراح نهج جديد يعتمد على خوارزمية شجرة القرارات C4.5 . يتم اختيار ثلاث سمات عددية وهدفين ليكونا على التوالي مدخلات ومخرجات النماذج. المدخلات المختارة هي درجة الحرارة المحيطة والإشعاع ونسبة الطاقة ، ويتم حساب هذا الأخير إعتمادا على الطاقة المقاسة و الطاقة المقدرة. يتم الحصول على الطاقة المقدرة من نموذج SANDIA الذي يمثل على النظام الكهروضوئي المتصل بالشبكة في حالته السليمة. مخارج نموذج الاكتشاف تكون إما حالة صحية أو حالة معيبة. من ناحية أخرى ، يحتوي النموذج التشخيصي على أربع فئات وهي: وسليم، وعطب السلسلة ، وعطب الدائرة القصيرة ، وعطب الخط. وتقسم البيانات إلى قسمين حيث أن 66٪ من هته البيانات تستعمل لعملية التعلم و أما الباقي فيستعمل للاختبار النموذج المحصل عليه. تشير نتائج الاختبار إلى أن كلتا النموذجين لديهما أداء تنبئي عالي في الكشف و التشخيص الأعطاب بدقة عالية حيث تفوق هذه النسبة 99٪.

CONTENU

REMERCIEMENT ……. 1

RESUME …….. 2

TABLEAU DE MATIERES …….. 5

LISTE DES TABLEAUX ……... 9

LISTE DES FIGURES ……... 11

INTRODUCTION GENERALE

CHAPITRE 1

L'ETAT DE L'ART SUR LES SYSTEMES PVS : GENERALITES, FONCTIONNEMENT ET ANOMALIES, ET TECHNIQUES DE DETECTIONS ET DIAGNOSTICS

1.1Introduction……..…… 24

1.2 Notions fondamentales …….. 25

1.2.1 La Cellule PV…….. 25

1.2.2 L'effet photovoltaïque……. 26

1.2.3 Caractéristiques courant-tension (I-V) d'une cellule PV …….…… 27

1.2.4 L'effet de la température et l'éclairement sur la cellule PV…….…… 28

1.2.4.1 L'effet de la température…… 28

1.2.4.2 L'effet de l'ensoleillement …… 29

1.2.5 Regroupement des cellules PVs…… 29

1.2.5.1 Regroupement en série ……. 29

1.2.5.1 Regroupement en parallèle…… 30

1.3 Les types des Systèmes PVs et leurs composants…….. 31

1.3.1 Systèmes autonome……..... 31

1.3.2 Systèmes autonome hybride……..... 31

1.3.3 Système photovoltaïque connecté au réseau ……..... 32

1.3.4 Les Composants de système PV connecté au réseau …… 33

1.3.4.1 Le module PV……. 33
1.3.4.2 String PV…….....……... 35
1.3.4.3 Champ PV…….…….. 35
1.3.4.4 Dispositif de protection ……... 36
1.3.4.5 Les convertisseurs PV……..... 37
1.4 Les défauts et les anomalies dans le champ PV ……. 39
1.4.1 Les défauts dans le champ PV……. 41
1.4.2 Méthodes de détection et de diagnostic des défauts dans la littérature…… 43
1.5 Conclusion……..…… 66

CHAPITRE 2

DESCRIPTION DU SYSTEME PV CONNECTE AU RESEAU

2.1 Introduction…….…….. 68
2.2 Description et fonctionnement de la SPVCR……...…….. 68
2.2.1 Les composants d'un SPVCR monophasé …… 68
2.2.2 Description du système PV connecté au réseau installé …….. 69
2.3 Description du système d'acquisition de données ……. 70
2.3.1 Les capteurs …….. 71
2.3.1.1 Mesure de courant …….…….. 71
2.3.1.2 Mesure de tension …….…….. 72
2.3.1.3 Mesure de l'ensoleillement sur le plan incliné…….…… 73
2.3.1.4 Mesure de la température ambiante ……..….. 74
2.3.2 Circuit alimentation ……..…… 74
2.3.3 Appareil d'acquisition de données ……..... 75
2.4 Récupération de données et calibration sous LabVIEW…… 76
2.5 Résultats d'acquisition de données……..... 78
2.6 Conclusion…….……. 80

CHAPITRE 3

MODELISATION DE SYSTEME PV CONNECTE AU RESEAU

3.1 Introduction …… 83

3.2 Modélisation de module PV …… 83

3.2.1 Modèle à une seule diode …….. 83

3.2.2 Modèle à double diode …… 84

3.2.3 Fonction de coût …….. 85

3.3 Algorithme de pollinisation des fleurs (FPA)…….…… 86

3.3.1 FPA dans le contexte de l'optimisation : source inspirée de la nature……. 86

3.3.2 Les lois fondamentales de l'FPA……..……. 88

3.3.2.1 Recherche globale de FPA (Biotique) …….…… 88

3.3.2.2 Recherche locale de FPA (Abiotique) ……..... 90

3.3.2.3 Commutateur probabilité en FPA…….…… 90

3.4. Mesure de caractéristique I-V…… 92

3.5 Simulation et résultat expérimental…… 94

3.6 L'estimation de MPP……..…… 101

3.7 Conclusion……...……... 106

CHAPITRE 4

DETECTION ET DIAGNOSTIC DES DEFAUTS PAR L'ARBRE DE DECISION

4.1 Introduction…….…….. 108

4.2 Modèle de SANDIA ……..…. 108

4.2.1 Identification des paramètres de modèle de SANDIA…….. 110

4.3 Les défauts considérés……. 115

4.4 La base de données de l'entrainement…… 116

4.4.1 Les attributs électriques…….…… 117

4.4.2 Les attributs sélectionnés…….. 118

4.5 L'arbre de décision (AD) ……..…….. 120

4.5.1 Notions de base ……..……... 120

4.5.2 Types de données ……..……... 121

4.5.3 Exemple introductif à l'arbre de décision ……..…….. 122
4.5.4 Construction d'un arbre de décision…….. 124
4.5.5 Choix de l'algorithme …….. 130
4.5.6 L'algorithme C4.5…….…… 130
4.5.7 Critères d'arrêt de construction d'un AD ……...……... 131
4.5.8 Élagage……...…….. 132
4.5.8.1 Pré-élagage……...…… 133
4.5.8.2 Post-élagage …….…….. 133
4.5.9 Évaluation du modèle …….. 136
4.5.9.1 Matrice de Confusion ……..…….. 136
4.6 Résultats expérimentaux et discussion……..……. 136
4.6.1. Résultats de l'entrainement ……..…….. 137
4.6.2 Résultats de validation…….. 143
4.6.2.1 Jour sans défaut……..……...…… 143
4.6.2.2 Jour de défaut chaîne ouvert …….……. 148
4.6.2.3 Jour de défaut court-circuit …… 153
4.6.2.4 Jour de défaut de ligne…… 158
4.6.2.5 Jour des défauts…….…… 162
4.7. Conclusion……..…….. 168
Conclusion générale…… 171
Référence ……..……...…… 177
Annexe…….…… 189

LISTE DES TABLEAUX

Tableau 1.1 Les différents types des cellules avec leur rendement…….…..…… 26
Tableau 1.2 Principaux défauts et anomalies rencontrés dans un générateur PV……..…… 44
Tableau 1.3 Les défauts éventuellement survenus qui pourraient être détectés correctement par la stratégie de Chine et al.……..……. 52

Tableau 2.1 Caractéristiques techniques de panneau solaire DIMEL……..… 69
Tableau 2.2 Caractéristiques techniques de l'onduleur Solis mini 700 W……..…… 70
Tableau 2.3 La description des capteurs avec leur valeur d calibration…… 77

Tableau 3.1 Pollinisation et ses composants d'optimisation…… 87
Tableau 3.2 Paramètres de l'algorithme FPA…….. 91
Tableau 3.3 Les paramètres électriques aux conditions d'essai normalisées (STC) pour les modules PV utilisés……..…… 93
Tableau 3.4 L'espace de recherche des paramètres…….……. 94
Tableau 3.5 Les résultats obtenus par les trois algorithmes appliqués sur le MSD……. 95
Tableau 3.6 Les résultats obtenus par les trois algorithmes appliqués sur le MDD…..……. 99

Tableau 4.1 Les valeurs des paramètres identifiés…… 110
Tableau 4.2. Données météorologiques pour jouer le tennis …… 127
Tableau 4.3 : La matrice de Confusion…….. 136
Tableau 4.4 Matrice de confusion du modèle de détection……..……. 141
Tableau 4.5 Résultat de test du modèle de détection……..……... 141
Tableau 4.6 Matrice de confusion du modèle de diagnostic…….……. 142
Tableau 4.7 Résultat de test du modèle de diagnostic…….……. 142
Tableau 4.8 Matrice de confusion du modèle de détection …….……. 144
Tableau 4.9 Résultat de test du modèle de détection (Jour sans défaut)……..…… 144
Tableau 4.10 Matrice de confusion du modèle de diagnostic …….……. 144
Tableau 4.11 Résultat de test du modèle de diagnostic ……..…….……. 144
Tableau 4.12 Matrice de confusion du modèle de détection (Jour de défaut chaine ouvert).. 152

Tableau 4.13 Résultat de test du modèle de détection ……. 152
Tableau 4.14 Matrice de confusion du modèle de diagnostic …….. 153
Tableau 4.15 Résultat de test du modèle de diagnostic …..…… 153
Tableau 4.16 Matrice de confusion du modèle de détection (Jour de défaut court-circuit).. 154
Tableau 4.17 Résultat de test du modèle de détection …… 154
Tableau 4.18 Matrice de confusion du modèle de diagnostic. …… 157
Tableau 4.19 Résultat de test du modèle de diagnostic ……..…….. 158
Tableau 4.20 Matrice de confusion du modèle de détection (Jour de défaut de ligne) …… 158
Tableau 4.21 Résultat de test du modèle de détection ……. 158
Tableau 4.22 Matrice de confusion du modèle de diagnostic ……..…… 159
Tableau 4.23 Résultat de test du modèle de diagnostic …… 159
Tableau 4.24 Matrice de confusion du modèle de détection (Jour des défauts)…….. 167
Tableau 4.25 Résultat de test du modèle de détection …… 167
Tableau 4.26 Matrice de confusion du modèle de diagnostic …….……. 168
Tableau 4.27 Résultat de test du modèle de diagnostic…… 168

LISTE DES FIGURES

Figure 1.1 Les différents types de cellules. …….. 26
Figure 1.2 Construction et illustration de l'effet photovoltaïque dans cellule PV cristalline. 26
Figure 1.3 Les Caractéristiques de la cellule PV **(a)** de courant-tension (I-V) et **(b)** de puissance-tension (P-V)…… ……. 27
Figure 1.4 Influence de la température sur les caractéristiques I-V.……..…….. 29
Figure 1.5 Influence de l'ensoleillement sur les caractéristiques I-V.……..……. 29
Figure 1.6 Caractéristiques courant tension de Ns cellules en série……..…….. 30
Figure 1.7 Caractéristiques courant tension de Np cellules en parallèle…… 30
Figure 1.8 Système PV connecté au réseau…….. 32
Figure 1.9 Caractéristiques d'un assemblage série/parallèle de Ns et Np cellules PV identique…….. 33
Figure 1.10 (a) Image réelle du module PV DIMEL (50) P, **(b)** Boîte de connexion du module PV, **(c)** Connexions de la sous-chaîne du module PV aux diodes de dérivation ….. 34
Figure 1.11 Schéma synoptique d'un champ PV…… 35
Figure 1.12 Points de fonctionnement résultant de la combinaison entre un module PV avec une charge résistive variable. ……..……..……..……..…….. 37
Figure 1.13 Différents types de configuration des onduleurs, (a) onduleur central, (b) onduleur de string et (c) onduleur multi-string. ……...……. 38
Figure 1.14 Les types de défauts se produisaient généralement dans les systèmes PVs…… 40
Figure 1.15 Diagramme schématique de différents types de défauts dans le champ PV**.....** 45
Figure 1.16 Les fonctions d'appartenance de la logique floue……..……..…….. 47
Figure 1.17 Les différentes étapes du modèle GISTEL ……...……. 48
Figure 1.18 Organigramme de la procédure de détection des défauts…….…….. 49
Figure 1.19 Modèle de six points le long de la courbe I-V…….…….. 50
Figure 1.20 Classification de l'état du système PV par Norm-tests…….. 51
Figure 1.21 Les différentes étapes de la méthode ANN……..…….. 52
Figure 1.22 (a) Le réseau de neurone MLP avec la matrice de confusion, **(b)** Le réseau de neurone RBF avec la matrice de confusion……..……..…… 53
Figure 1.23 Toutes les couches de l'algorithme de détection de défaut dans SPVCR…….. 55

Figure 1.24 Comparaison des performances d'un système PV. ……. 57
Figure 1.25 Un aperçu général sur PVSAT2……...…….. 58
Figure 1.26 Méthode de calcul du rapport de performance dans un SPVCR…… 59
Figure 1.27 Valeurs de PR quotidiennes, hebdomadaires et mensuelles pour un système PV pour 2001……...…….……...…….…… 60
Figure 1.28 Organigramme proposé du système de détection……...…… 62
Figure 1.29 Organigramme de la procédure de diagnostic des défauts …… 63
Figure 1.30 Schéma de principe pour la configuration TDR dans le domaine PV…… 64
Figure 1.31 Entrée échelon et les formes d'onde réfléchies avec le câble……. 64
Figure 1.32 (a) Un mismatch causé par une cellule ombrée avec son image thermique correspondante, (b) l'effet d'une polarisation inverse sur une cellule PV en silicium monocristallin avec l'image thermique de ce dernier……. 65

Figure 2.1 Schéma block d'un système photovoltaïque connectées au réseau…….. 68
Figure 2.2 Schéma générale de la SPVCR…….…… 70
Figure 2.3 Aperçu global du SPVCR …….…..……. 71
Figure 2.4 Capteur ACS712: -a- la conception du capteur. -b- le brochage du capteur -c-Schéma de brochagede du capteur…….……..…… 72
Figure 2.5 Capteur LEM LV 25-P: -a- la conception du capteur. -b- le brochage du capteur -c-Schéma de brochagede du capteur……..……. 72
Figure 2.6 Capteurs de rayonnement SPEKTRON 200…….…… 73
Figure 2.7 Circuit de l'amplificateur différentiel…….……. 73
Figure 2.8 Capteur de température LM35……..……..... 74
Figure 2.9 Conception du circuit d'alimentation…….……... 75
Figure 2.10 (1) Cartes de mesure, et (2) carte d'alimentation…….……... 75
Figure 2.11 Multiplexeur d'acquisition de données NI-USB 6008…….…….. 76
Figure 2.12 Assistant fenêtre DAQ…….…….. 76
Figure 2.13 (**a**) Face avant du programme d'acquisition LabVIEW (**b**) Diagramme LabVIEW VI pour l'acquisition …….. 77
Figure 2.14 L'irradiation mesurée…….. 78

Figure 2.15 La température ambiante mesurée. …… 79
Figure 2.16 Le courant DC mesuré…… 79
Figure 2.17 La tension DC mesurée…….. 80
Figure 2.18 La puissance DC mesurée…….. 81

Figure 3.1 Circuit électrique équivalent d'un modèle à une seule diode pour une cellule PV 83
Figure 3.2 Circuit électrique équivalent du modèle à double diode pour une cellule PV..... 84
Figure 3.3 L'organigramme de l'algorithme FPA……..……..…….. 89
Figure 3.4 Schéma synoptique pour l'acquisition de données des caractéristiques I-V…..... 92
Figure 3.5 Le schéma synoptique du processus d'optimisation……..…….. 93
Figure 3.6 Les courbes de la variation des erreurs obtenues pour les trois algorithmes FPA, SA et FA utilisant le MSD…….……..…….. 96
Figure 3.7 Données mesurées vs simulées avec MSD pour cellule de type Monocristallin **(a)** ISOFOTON 106 **(b)** SANYO……..……..…….. 97
Figure 3.8 Données mesurées vs simulées avec MSD pour cellule de type Polycristallin **(a)** CONDOR **(b)** KYOCERA……..……..……..……. 98
Figure 3.9 Les courbes de la variation des erreurs obtenues pour les trois algorithmes FPA, FA et SA utilisant le MDD …….……..……. 100
Figure 3.10 Données mesurées vs simulées avec MDD pour cellule de type Monocristallin **(a)** ISOFOTON 106 **(b)** SANYO…….…….. 101
Figure 3.11 Données mesurées vs simulées avec MMD pour cellule de type Polycristallin **(a)** CONDOR **(b)** KYOCERA……...……..…….. 102
Figure 3.12 Comparaison entre les valeurs réelles et estimées au MPP du ciel clair pour (a) courant, (b) tension et (c) puissance……...……..…… 105
Figure 3.13 Comparaison entre les valeurs réelles et estimées au MPP du ciel nuageux pour (a) courant, (b) tension et (c) puissance…….……. 106

Figure 4.1 La courbe de la variation d'erreur durant le processus d'identification des paramètres. …… 111
Figure 4.2 Représentation de dispersion de la puissance mesurée et la puissance estimée... 111

Figure 4.3 Profil du jour du ciel clair (a) température ambiante; b) l'irradiation; c) Température de cellule estimée; d) Puissance estimé vs mesurée du SPVGC (20 juillet 2017)…… …….……. 113
Figure 4.4 Profil du jour à ciel nuageux (a) température ambiante; b) température estimée de la cellule; c) irradiation; d) puissance estimé VS la puissance mesurée du SPVCR (4 novembre 2017)……. ……..…… 115
Figure 4.5 Représentation de dispersion de la puissance mesurée et la puissance estimée du jour à ciel nuageux……..……..…….. 115
Figure 4.6 Les Différents types de défauts considérés dans le système PV…….……. 116
Figure 4.7 Histogramme de l'ensemble de données…….…….. 117
Figure 4.8 P-V caractéristique au MPP dans diverses conditions météorologiques. …… 118
Figure 4.9 Tracé 3D du modèle d'entrées DT proposé…….. 119
Figure 4.10 L'induction par l'arbre de décision……..…….. 121
Figure 4.11 Caractéristiques I-V en fonctionnement Normal et Défaut…….…….. 123
Figure 4.12 Modèle d'arbre de décision pour détection de défaut……. 123
Figure 4.13 Structure de l'arbre de décision pour l'ensemble D…….…….. 125
Figure 4.14 Les sous-ensembles associés aux valeurs des quatre attributs…….…….. 128
Figure 4.15 Les sous-ensembles correspondant à la valeur ensoleillée de l'attribut Temps.. 129
Figure 4.16. L'arbre de décision final pour la base météo……..…….. 130
Figure 4.17 Evolution de taux d'erreur en apprentissage et en test…….. …….. 132
Figure 3.18 Arbres de décision avec et sans élagage ……..…….. 133
Figure 4.19 Elagage d'un sous-arbre ……...……. 135
Figure 4.20 Une partie de l'arbre du modèle pour la détection de défauts……..…….. 137
Figure 4.21 Partie des règles du modèle AD pour la détection de défauts…….. 138
Figure 4.22 Une partie de l'arbre du modèle AD pour diagnostic de défaut……. 140
Figure 4.23 Une partie des règles du modèle AD pour le diagnostic de défaut……. 141
Figure 4.24 Courbes de (a) la température ambiante, (b) la température estimée de la cellule (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance. …….……..……..…… 148

Figure 4.25 Courbes de (a) la température ambiante, (b) la température estimée de la cellule (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance. ……..……..…….. 152
Figure 4.26 Courbes de (a) la température ambiante, (b) la température estimée de la cellule (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance. ……..……..…… 157
Figure 4.27 Courbes de (a) la température ambiante, (b) la température estimée de la cellule (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance. ……..……..…….. 162
Figure 4.28 Courbes de (a) la température ambiante, (b) la température estimée de la cellule (c) l'irradiation, (d) la puissance estimée et mesurée et (e) le rapport de puissance……..……... 166
Figure 4.29 Résultat de la classification de (a) modèle de détection et (b) modèle de diagnostic…….…….. 167

LISTES DES ACRONYMES ET ABREVIATIONS

PV : Photovoltaïque

CDER	:	Centre de Développement des Énergies Renouvelables
SPVCR	:	Système Photovoltaïque Connecté au Réseau
$MPPT$	:	Suivi de Point de Puissance Maximale (Maximum Power Point Traking)
MPP	:	Point de Puissance Maximale
AD	:	Arbre de Décision
N	:	Jonction N de la diode
P	:	Jonction P de la diode
GPV	:	Générateur Photovoltaïque
N_S	:	Nombre de module en série du champ PV
V_{ocN_s}	:	la tension de circuit ouvert de (Ns) cellules en série [V]
I_{scN_s}	:	le courant de cout circuit de (Ns) cellule en série [A]
N_P	:	Nombre de module en parallèle du champ PV
V_{ocNp}	:	la tension de circuit ouvert de (N_p) cellules en parallèle [V]
I_{scN_p}	:	le courant de cout circuit de (N_p) cellule en parallèle [A]
I	:	Courant fournit par la cellule [A]
I_{ph}	:	Courant généré par l'éclairement [A]
I_{Rsh}	:	Courant circulant dans la résistance shunt [A]
I_d	:	Curant circulant dans la diode [A]
I_{0d}	:	Courant de saturation de la diode [A]
E	:	L'ensoleillement / l'éclairement / rayonnement absorbé par la cellule
E_0	:	L'ensoleillement / l'éclairement / rayonnement de référence (1000) [W/m²]
R_s	:	Résistance série [Ω]
R_{sh}	:	Résistance shunt [Ω]
K	:	Constante de Boltzmann (K=1,38.10^{-23}) [J/°K]
q	:	Charge de l'électron (q=1,602.10^{-19}) [C]
I_{sc}	:	Courant de shunt [A]
V_{oc}	:	Tension de circuit-ouvert [V]
T	:	Température ambiante
T_c	:	Température de cellule
I_{dc}	:	Courant continue en MPP mesuré [A]

V_{dc}	:	Tension continue en MPP mesuré [V]
P_{dc}	:	Puissance continue en MPP mesuré [VA]
I_{mpp}	:	Courant en MPP [A]
V_{mpp}	:	Tension en MPP [V]
CC	:	Court-Circuit
P_{mpp}	:	Puissance en MPP [W]
P_T	:	Puissance maximale théorique [W]
DC	:	Courant continue [A]
AC	:	Courant Alternative [A]
X	:	L'ensemble des attributs
n	:	Le nombre d'attributs
C	:	L'ensemble de classes
k	:	Le nombre de classes
$\lvert C_j \rvert$	:	Le nombre des exemples qui appartiennent à la classe C_j
D	:	L'ensemble des exemples
$\lvert D \rvert$	:	Le nombre total des exemples
Info	:	Entropie / information [Bits]
Gain	:	Gain d'information [Bits]
Rapport_Gain	:	Le rapport de gain d'information normalisé
S_j	:	Le seuil de test sur l'attribut
E_M	:	Le nombre d'exemples mal classés sur une feuille
Z	:	Le Seuil de confiance de la distribution Normale centré réduite
N_T	:	Le nombre Totale d'exemples couvres sur une feuille
f	:	Le taux d'erreur réel
α	:	Le niveau de confiance
e	:	Le taux d'erreur prédit

INTRODUCTION GENERALE

Introduction générale

Au cours de ces dernières années, l'électricité souffre d'une contrainte irréfutable: à tout moment, les systèmes électriques doivent assurer un équilibre entre la production et la consommation tout en conservant une tension satisfaisante. Historiquement, la fiabilité du réseau était principalement assurée par l'excès de capacité dans le système avec le flux unidirectionnel vers les consommateurs dispersés à partir de centrales de grande puissance.

Aujourd'hui, les énergies renouvelables sont considérées comme une solution potentielle pour la réduction des émissions de gaz à effet de serre et la sécurité énergétique. Alimentée par des facteurs économiques, environnementaux et sociaux, l'utilisation des générateurs photovoltaïques augmente. Grâce à son fonctionnement sans bruit, il peut être facilement installé à l'extérieur et sur les toits. Cependant, le développement de la production photovoltaïque connectée au réseau est limité par la production d'électricité intermittente et le décalage entre la production électrique photovoltaïque, et la consommation réelle. Un déploiement massif de systèmes photovoltaïques complique l'équilibre entre production et consommation, ce qui peut provoquer des pannes de courant s'il est perturbé.

Un système photovoltaïque peut être soumis, au cours de son fonctionnement, des défauts différentes et des anomalies conduisant à une baisse des performances du système voir son indisponibilité totale [**1,2**]. Réellement, plusieurs facteurs peuvent affecter d'une manière significative les performances du système PV [**3,4**]. En effet, ces installations sont fréquemment sujettes à différentes sources de défauts et d'anomalies affectant la puissance générée par les générateurs PV et conduisent à une diminution de leur efficacité [**5,6**]. En plus, ces défauts pourraient réduire considérablement le taux de la productivité et la durée de vie du champ PV. Toutes ces conséquences défavorables vont évidemment réduire le rendement de l'installation, sans compter le coût de maintenance nécessaire pour remettre le système en état normal et le risque d'incendie suivant d'un défaut persistant.

Le système PV, en particulier le type DC, peut avoir à des défauts et des anomalies entraînant une baisse de performances globales du système [**8**]. Par conséquent, la procédure de détection et de diagnostic des pannes est nécessaire et importante en même

temps, non seulement pour réduire les coûts de maintenance, mais également pour éviter toute perte d'énergie, tout dégât matériel et tout risque (incendie, etc…).

La détection et le diagnostic précoces des défauts dans une installation PV réduisent considérablement les coûts de maintenance, et augmentent la productivité tout en augmentant le temps de disponibilité des installations PVs et optimisent le rendement de ces dernières. Parmi les approches les plus utilisées, dans la littérature, pour la détection et le diagnostic de pannes dans les systèmes PV, on distingue celles basées sur les techniques de modélisations statistiques et mathématiques et celles basées sur les techniques d'intelligences artificielles basées d'apprentissage automatique.

Les techniques de modélisations mathématiques et statistiques utilise un processus de comparaison entre deux grandeurs l'une est mesurée et l'autre est estimée (obtenues par des modèles mathématiques ou analytiques), et lorsqu'un écart important apparait cela indique qu'il y a une panne probable [**7,8**]. Ce genre de technique nécessite l'utilisation des modèles très précis pour le fonctionnement sain et défectueux. Néanmoins, ces approches ont montré leurs limites dans l'identification du type de défaut avec précision.

D'autre part, les techniques d'intelligences artificielles à base d'apprentissage automatique nécessitent généralement trois phases pour obtenir un modèle fiable qui permet de détecter et diagnostiquer les défauts. La première phase est la collection et le traitement des données mesurées pour être utilisées dans la phase suivante. La phase est l'entrainement de modèle par un algorithme basé sur l'intelligence artificielle comme le réseau de neurones, la logique flou, les séparateurs à vaste marge, les K plus proches voisins ou les arbres de décision [**9-14**]. La en troisième phase est celle processus d'évaluation. Ce processus est nécessaire avant d'utiliser en temps réel les modèles résultants. La fiabilité des modèles obtenus est fortement liée à la qualité et à la quantité de données utilisées dans la phase d'entrainement [**15**]. De plus, quelques modèles sont sensibles aux données bruitées ou erronées [**16**]. Par conséquent, ce type de données pourrait évidemment affecter la fiabilité, et même l'utilisation en temps réel de ces modèles.

L'objectif de cette thèse consiste à proposer une approche de détection et de diagnostic des défauts dans les systèmes photovoltaïques basée sur une technique de l'intelligence artificielle qui est l'arbre de décision.

Dans cette étude, on s'intéresse aux différents défauts et anomalies qui peuvent survenir dans partie DC de SPVCR. L'objectif consiste essentiellement en la modélisation de GPV, la détection et le diagnostic des défauts tel que le défaut de chaine coupé (ouverte), le défaut de court-circuit et défaut de ligne. Le contenu de la thèse est scindé en quatre chapitres.

Le premier chapitre est divisé en deux parties, la première partie est dédiée à des généralités sur les systèmes photovoltaïques (effet photovoltaïque, modèle et caractéristique d'une cellule PV), ainsi que les définitions classiques des différents composants qui constituent le système PV. Nous énumérons également les différentes applications des systèmes PVs et leurs architectures. La deuxième partie présente dans un premier temps les différents défauts associés au système PV connecté au réseau et les causes de ces défauts dans les parties DC et AC. A la fin, nous présentons l'état de l'art sur les différentes méthodes de diagnostic utilisées pour détecter les défauts dans un système photovoltaïque.

Les générateurs PVs nécessitent un suivi par un monitoring photovoltaïque en utilisant plusieurs stratégies [**17**]. Ce monitoring est réalisé à partir des données climatiques et électriques du système PV. Alors, il est nécessaire de procéder à l'acquisition de ces données en temps réel via un système d'acquisition de données. Ainsi, le chapitre a pour objectif la présentation des différentes étapes de réalisation d'un système d'acquisition et d'enregistrement de données à utiliser pour l'étude du comportement dynamique d'un système PV connecté au réseau.

Le troisième chapitre est consacré à la modélisation des modules PVs afin d'estimer le courant et la tension au point de puissance maximale (MPP). Deux modèles équivalents de la cellule PV sont présentés, en l'occurrence le modèle à une seule diode (MSD) et le modèle à double diode (MDD) [**18**]. Chaque modèle possède des paramètres inconnus qui doivent être identifiés. L'algorithme de pollinisation des fleurs (Flower Pollination algorithm FPA) est un algorithme d'optimisation inspiré de la nature [**19**]. Cet algorithme est utilisé pour identifier les valeurs optimales des paramètres des deux modèles. Le principe de cet algorithme d'optimisation consiste à imiter les voies de transfert du pollen pour aider à produire des plantes dans la nature, en d'autres termes, ils existent de nombreuses façons dont le pollen peut être transporté pour reproduire les plantes. Ces méthodes peuvent être exploitées pour développer des algorithmes

d'optimisation efficaces. Afin d'évaluer les performances de l'algorithme plusieurs données expérimentales ont été utilisées. Ces dernières sont acquises sous forme de courbes courant-tension appelées aussi courbes I-V. Ces courbes I-V sont obtenues pour deux types de technologies de cellules PVs à savoir Monocristalline et Polycristalline. Les résultats de la simulation sont comparés aux données expérimentales. Le modèle MSD identifiés est utilisé pour estimer le courant, la tension et la puissance au MPP qui seront par la suite comparés aux grandeurs réelles du système PV connecté au réseau (SPVCR).

Le quatrième chapitre est consacré aux différentes phases de la construction des modèles de détection et de classification de défauts qui surviennent dans un système PV connecté au réseau. Pour cela, une nouvelle approche basée sur l'algorithme C4.5 d'arbre de décision est proposée [**20,21**]. Ainsi, un modèle non paramétrique a été proposé par le biais de cet algorithme et utilisé pour prédire l'état du SPVCR. Un ensemble de données a été recueilli par le système d'acquisition dans plusieurs conditions météorologiques. Trois attributs numériques et deux cibles sont choisis pour former les données finales utilisées. Les attributs sont la température ambiante, l'irradiation et le rapport de puissance. Ce dernier est calculé à partir de la puissance mesurée et estimée du système. La première cible est l'état du système qui possède deux cas, soit sain ou défectueux (c'est pour le modèle de détection); la deuxième cible contient quatre étiquettes de classes nommées sain, défaut de chaîne, défaut de court-circuit ou défaut de ligne (c'est pour le modèle de diagnostic). Le modèle SANDIA est utilisé pour estimer la puissance générée par le SPVCR fonctionnant à l'état sain. L'ensemble de données a été divisé en deux parties, où 66 % ont été utilisées pour l'entrainement (construction de modèle) et le reste pour le test. Par la suite, de nouvelles données obtenues pendant cinq jours sont utilisées afin d'évaluer la robustesse et l'efficacité des deux modèles. Les résultats des tests réalisés démontrent les performances de prédiction de ces modèles dont la précision est supérieure à 99%. La thèse se termine par une conclusion générale.

CHAPITRE 1

L'ETAT DE L'ART SUR LES SYSTEMES PVS : GENERALITES, FONCTIONNEMENT ET ANOMALIES, ET TECHNIQUES DE DETECTIONS ET DIAGNOSTICS

Chapitre 1

L'Etat de l'Art sur les Systèmes PVs : Généralités, Fonctionnement et Anomalies, et Techniques de Détections et Diagnostics

1.1 Introduction

L'énergie renouvelable est l'énergie provenant de sources qui se réapprovisionnent naturellement, mais qui sont limitées au débit; la durée des ressources renouvelables est pratiquement inépuisable, mais limitée dans la quantité d'énergie disponible par unité de temps. Les principaux types de sources d'énergie renouvelables sont : biomasse, hydroélectricité, géothermie, éolienne et solaire (photovoltaïque).

L'énergie photovoltaïque est la plus jeune des énergies renouvelables, elle a l'avantage d'être non polluante, souple et fiable **[22]**. Les systèmes photovoltaïques sont utilisés depuis 40 ans. Les applications ont commencé avec le programme spatial pour la transmission radio des satellites. Elles se sont poursuivies ensuite avec les balises en mer et l'équipement des sites isolés. La production d'électricité photovoltaïque dépend de l'ensoleillement du lieu et de la température, donc de sa localisation géographique, de la saison et de l'heure de la journée : la production est ainsi maximale à midi (heure solaire) par ciel dégagé.

Les systèmes PVs peuvent être généralement considérés comme fiables, mais ils ne peuvent pas être sans défauts et leur production peut être affectée de façon considérable par la présence des défauts. Un système de détection de défaut permet de réduire les pertes afin d'augmenter la production annuelle, il est défini comme un système qui est capable de distinguer entre les différentes conditions de fonctionnement de l'installation PV. Si la présence d'un défaut est détectée, le système doit être aussi en mesure d'indiquer les caractéristiques de ce défaut.

Dans ce chapitre, des notions fondamentales sur les systèmes PVs seront présentées. De plus, les différents éléments du système PV connecté au réseau (SPVCR) sont abordés. Ensuite, on passe aux définitions des défauts et des anomalies qui peuvent resurgir dans le SPVCR, et à la fin, on présent les solutions et les approches proposées dans la littérature pour mentionnées.

1.2 Notions fondamentales

Cette partie contient un regard de près sur l'élément de base du système PV avec ses caractéristiques électriques telles que le courant, la tension et puissance. De plus, pour savoir comment les cellules PVs produisent de l'énergie électrique, nous décrirons attentivement le fonctionnement de cet élément.

1.2.1 La cellule PV

Une cellule PV est une technologie de collecte d'énergie, qui convertit l'énergie solaire en électricité grâce à un processus appelé l'effet photovoltaïque. Il existe plusieurs types de cellules PVs qui utilisent tous des semi-conducteurs pour interagir avec les photons entrants du soleil afin de générer un courant électrique. Généralement, la tension générée par la cellule PV commerciale est d'environ 0,6 V, tandis que son courant produit dépend principalement de l'intensité du soleil (éclairement ou ensoleillement) et de la surface de la cellule solaire **[22,23]**.

À l'heure actuelle, la plupart des cellules photovoltaïques commerciales sont fabriquées à partir de silicium, le même matériau à partir duquel le sable est fait **[24]**. Il existe 3 grandes familles de la cellule PV qui dépend sur la matière basique de fabrication **[24]** :

- La cellule mono cristalline ;
- La cellule poly cristalline ;
- La cellule amorphe.

La figure 1.1 montre les différents types de cellules PV.



Silicium Monocristallin



Silicium poly-cristallin



Silicium amorphe

Figure 1.1. Les différents types de cellules.

Le tableau suivant présente les différents types des cellules avec leur rendement.

Tableau 1.1 Les différents types des cellules avec leur rendement [25].

Technologie de cellules	**Rendement en laboratoire**	**Rendement de production**
Silicium Amorphe (A-Si)	13%	7 à 9%
Silicium Poly-cristallin (P-Si)	19,8%	14 à 16%
Silicium Monocristallin (M-Si)	24,7%	16 à 17%

Le type de cellule qu'on a utilisé dans notre système est le silicium monocristallin (M-Si).

1.2.2 L'effet photovoltaïque

L'effet photovoltaïque est un processus qui génère de la tension ou du courant électrique dans une cellule PV lorsqu'elle est exposée aux rayons du soleil. Ces cellules solaires sont composées de deux types différents de semi-conducteurs, un type P et un type N, qui sont réunis pour créer une jonction p-n comme montré dans la figure 1.2 **[26,27]**.

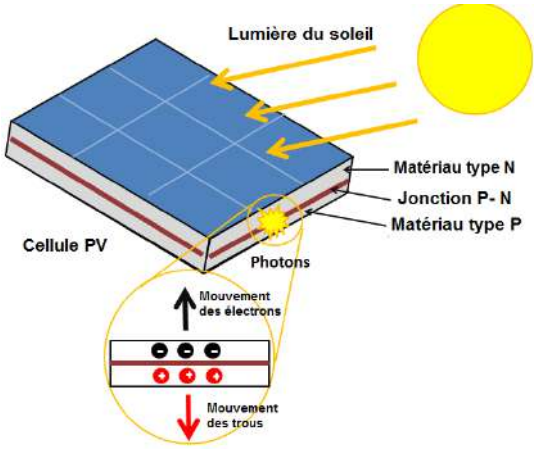


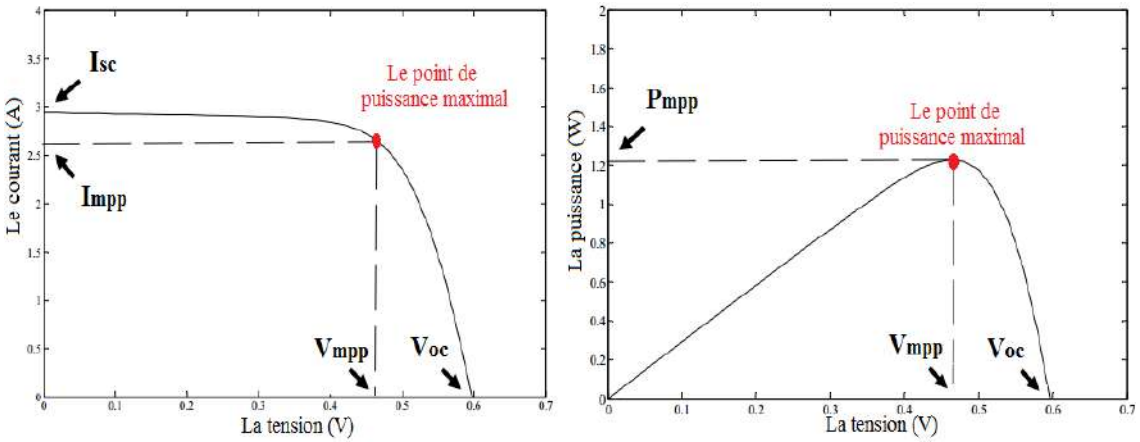
Figure 1.2 Effet photovoltaïque dans cellule PV cristalline [28].

En joignant ces deux types de semi-conducteurs, un champ électrique se forme dans la région de la jonction que les électrons se déplacent vers le côté P positif et les trous se déplacent vers le côté négatif N. Ce champ provoque des particules chargées négativement de se déplacer dans une direction et des particules chargées positivement dans l'autre direction **[28,29]**. Le rayon solaire est composé de photons, qui sont simplement de petits faisceaux de rayonnement électromagnétique ou d'énergie. Lorsque le rayon d'une longueur d'onde appropriée est incident sur ces cellules, l'énergie du photon est transférée à un électron du matériau semi-conducteur, ce qui le fait basculer dans un état d'énergie plus élevé connu sous le nom de bande de conduction. Dans leur état excité dans la bande de conduction, ces électrons sont libres de se déplacer à travers le matériau, et c'est ce mouvement de l'électron qui crée un courant électrique dans la cellule.

1.2.3 Caractéristiques courant-tension (I-V) d'une cellule PV

Les principales propriétés électriques de la cellule PV sont résumées dans la relation entre le courant et la tension résultante sur une courbe de caractéristiques I-V typique de la cellule solaire. L'intensité du rayonnement solaire (irradiation) qui descendit sur la cellule contrôle le courant (I), tandis que les augmentations de température de la cellule solaire réduisent sa tension (V).

Les figures 1.3.a et 1.3.b représentent respectivement la courbe I = f (V) et la courbe P = f(V) d'une cellule PV typique dans des conditions constantes d'irradiation et de température :



(a) (b)

Figure 1.3 Caractéristiques de la cellule PV **(a)** de courant-tension (I-V) et **(b)** de puissance tension (P-V) **[30].**

Une cellule PV produira son courant maximal lorsqu'il n'y a essentiellement pas de résistance dans le circuit. Ce serait un court-circuit entre ses bornes positives et négatives. Ce courant maximum est appelé le courant de court-circuit (Isc). Lorsque la cellule est court-circuitée, la tension dans le circuit est nulle. Inversement, la tension maximale est produite quand il y a une coupure (résistance infinie ou circuit ouvert) dans le circuit. C'est ce qu'on appelle la tension de circuit ouvert (Voc). Dans cette condition, la résistance est infiniment élevée et il n'y a pas de courant puisque le circuit est incomplet. Ces deux extrêmes dans la résistance de la charge et l'ensemble des conditions entre eux sont représentés sur un graphe appelé une courbe I-V (courant-tension) ou un graphe appelé une courbe P-V (puissance tension).

Les points de genou, dans les deux courbes I-V et P-V, représentent le point de puissance maximale (MPP), où ce point est la puissance de sortie optimale produite par la cellule PV qui peut être déterminée en choisissant la valeur idéale de la charge connectée. Le point de la puissance maximale est défini par la multiplication de la tension et du courant au MPP, où V_{mpp} et I_{mpp} sont les valeurs de tension et de courant au MPP qui donne la sortie de fonctionnement maximale de puissance.

1.2.4 L’effet de la température et l'éclairement sur la cellule PV

La température et l’ensoleillement sont deux grandeurs qui influencent sur la cellule PV. C'est pourquoi il est préférable de voir comment chacun d'entre eux affecte les caractéristiques I-V de la cellule PV.

1.2.4.1 L’effet de la température

La température est un paramètre très important dans le comportement des cellules PV. D’après la figure 1.4, on voit que le courant augmente avec la température; par contre la tension de circuit ouvert diminue. Ceci entraîne une diminution de la puissance maximale disponible. C’est-à-dire, la température et la tension générée par la cellule PV a une corrélation inverse.

1.2.4.2 L’effet de l’ensoleillement

La figure 1.5 représente la variation des caractéristiques I-V et P-V pour valeurs de l'ensoleillement. D’après cette figure, la valeur de l’ensoleillement augmente, et le courant généré par la cellule augmente aussi. C'est-à-dire, l'ensoleillement et le courant généré par la cellule PV ont une corrélation directe.

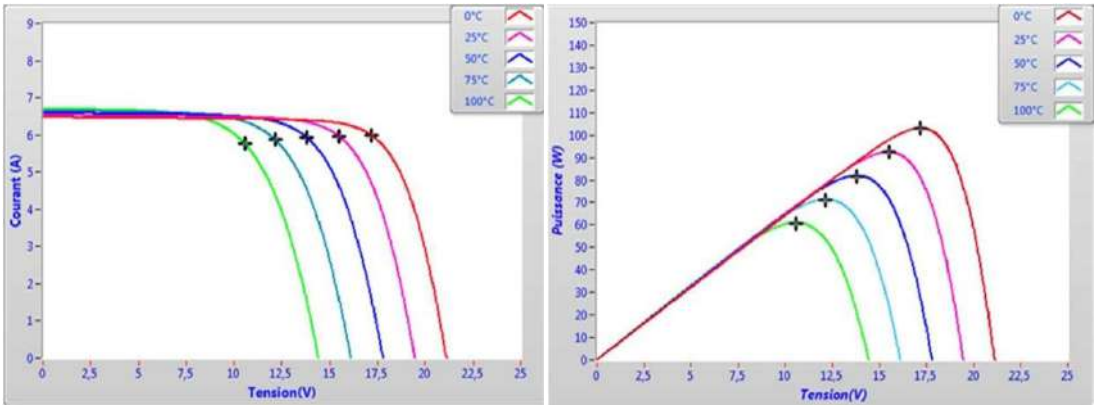


Figure 1.4 Influence de la température sur les caractéristiques I-V [**21**].

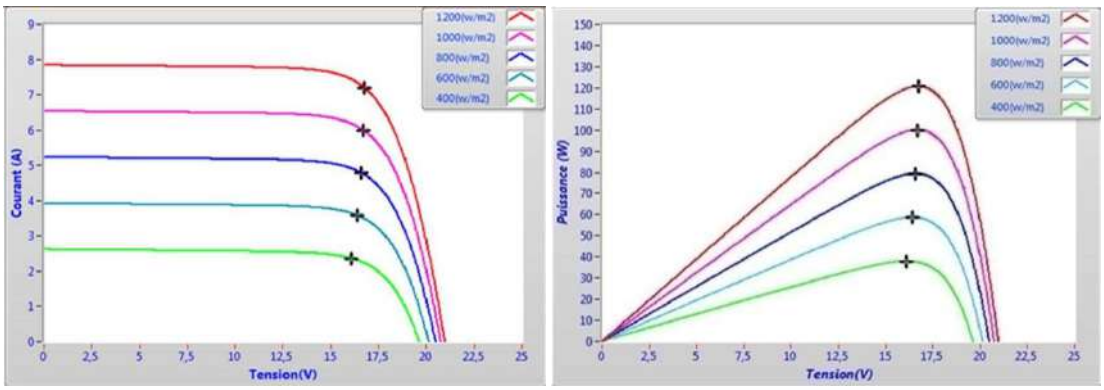


Figure 1.5 Influence de l’ensoleillement sur les caractéristiques I-V [**21**].

1.2.5 Regroupement des cellules PVs

1.2.5.1 Regroupement en série

Une association de Ns cellule en série comme illustré dans la figure 1.6 permet d'augmenter la tension générée. Les cellules sont alors traversées par le même courant et les caractéristiques résultants du groupement série sont obtenues par l’addition des tensions

élémentaires de chaque cellule. Les équations suivantes résument les caractéristiques électriques d'une association série de (Ns) cellules **[31-33].**

$$V_{ocN_s} = N_s \times V_{oc} \quad (1.1)$$

$$I_{scN_s} = I_{sc} \quad (1.2)$$

V_{ocN_s} : La tension en circuit ouvert de Ns cellules en série.

I_{scN_s} : Courant de court-circuit de Ns cellules en série.

1.2.5.1 Regroupement en parallèle

Une association parallèle de N_p cellules figure 1.7 est possible et permet d'accroître le courant de sortie du générateur ainsi créé.

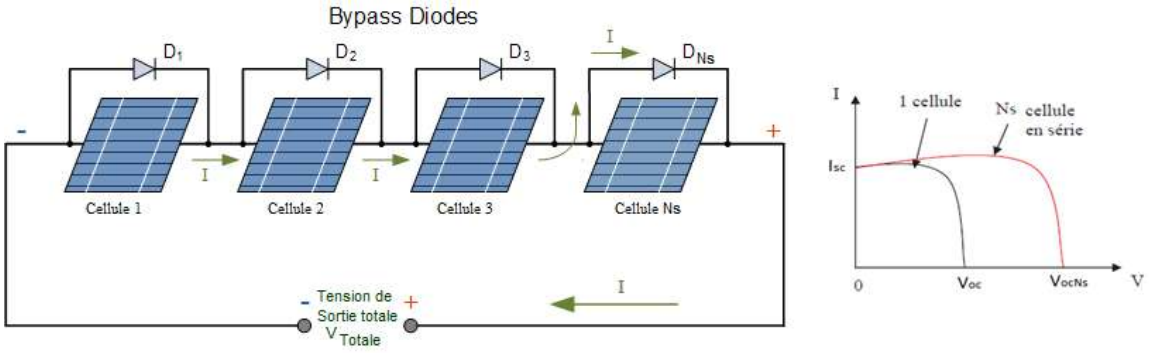


Figure 1.6 Caractéristiques courant-tension de Ns cellules en série.

Dans un groupement de cellules identiques connectées en parallèle, les cellules sont soumises à la même tension et les caractéristiques résultantes du groupement sont obtenues par l'addition des courants **[31-33]**.

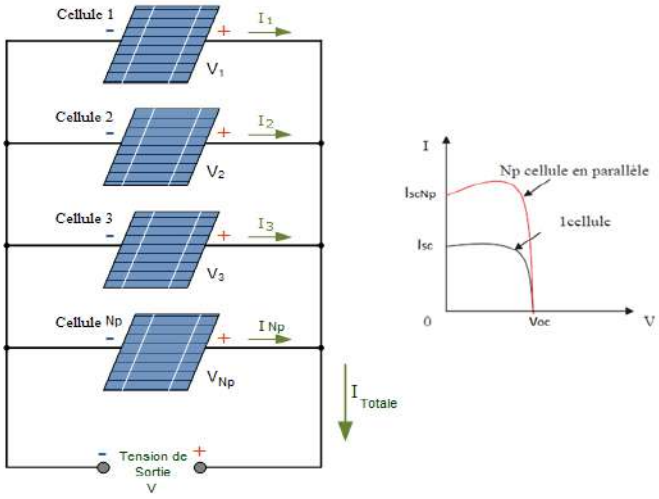


Figure 1.7 Caractéristiques courant-tension de Np cellules en parallèle.

$$I_{scN_p} = N_p \times I_{sc} \quad (1.3)$$

$$V_{ocN_p} = V_{oc} \quad (1.4)$$

I_{scN_p} : Le courant de cout circuit de N_p cellules en parallèle.

V_{ocNp} : La tension de circuit ouvert de N_p cellules en parallèle.

1.3 Les types des Systèmes PVs et leurs composants

Un module PV seul ne suffit pas pour alimenter régulièrement une application. L'énergie fournit par le générateur PV est en courant continu (DC), elle est variable en fonction du niveau d'ensoleillement reçu : il faut souvent la stocker, adapter ou parfois la transformer [**32**]. Nous appelons un système PV l'ensemble des composants nécessaires à l'alimentation d'une application en toute fiabilité. Les différentes configurations des systèmes photovoltaïques dépendront des besoins, de l'emplacement et du budget alloué.

1.3.1 Système autonome

C'est la configuration la plus courante des systèmes PV. Ce système est le plus souvent en DC, ce qui est préférable vu sa simplicité. Mais dans le cas de présence d'une charge AC usuelle, le système nécessite l'ajout un convertisseur DC/AC. Les appareils alimentés en DC sont donc câblés sur la batterie à travers d'un régulateur de charge. Lorsque la batterie est pleine, le régulateur de charge coupe le générateur pour éviter qu'elle ne souffre de surcharge [**33**]. Quand une charge doit fonctionner en courant alternatif AC, a recourt à une conversion DC/AC. Il existe plusieurs applications pour les systèmes PVs autonomes citant par exemple : l'alimentation maison, télécoms, éclairage, météorologie, sécurité maritime, pompage et gestion d'eau.

1.3.2 Systèmes autonome hybride

Une des limites d'un système autonome purement photovoltaïque, comme on vient de le décrire, est qu'il fournit une puissance donnée, variable selon la saison, mais que l'on ne peut pas dépasser, au risque de détruire la batterie par décharge profonde.

Avoir un système hybride, c'est disposer d'une autre source d'électricité autonome qui vient de compléter l'apport photovoltaïque [**34**]. Cette autre source peut être un groupe électrogène ou un aérogénérateur. On retiendra l'aérogénérateur si le site est bien venté, de préférence au cours des saisons ou l'ensoleillement est plus bas. Mais quand

l'approvisionnement en diesel est possible, le groupe électrogène est plus confortable, car on y a recours à volonté. La gestion des charges en alimentation directe sur le groupe ou sur la batterie dépend du type de charge et du mode de fonctionnement souhaité.

1.3.3 Système photovoltaïque connecté au réseau

Les systèmes PVs sont généralement classés en fonction de leurs besoins fonctionnels et opérationnels, de leurs configurations de composants et de la façon dont l'équipement est connecté à d'autres sources d'énergie et charges électriques. Les deux principales classifications sont les systèmes connectés au réseau ou les systèmes interactifs par les services publics et les systèmes autonomes. Les systèmes PVs peuvent être conçus pour fournir un service d'alimentation électrique DC et/ou AC, et peuvent fonctionner interconnectés avec ou indépendamment du réseau de service public, et peuvent être connectés à d'autres sources d'énergie ou des systèmes de stockage d'énergie [34,35].

Le système connecté au réseau électrique (ou relié à une centrale de distribution), ce qui représente une économie importante en investissement et en fonctionnement. Il utilise le réseau comme un stock, et transforme le surplus de l'électricité produite vers le réseau sous forme d'un courant alternatif, tant en termes d'énergie que de coûts. Ces installations se composent de modules PV interconnectés, d'un (ou plusieurs) onduleur(s) raccordé(s) au réseau électrique. L'onduleur convertit le courant continu généré par le champ PV en un courant alternatif conforme au réseau électrique (220V/380V, 50Hz) [36]. Il gère également des fonctions de commande et de réglage afin d'optimiser l'apport énergétique.

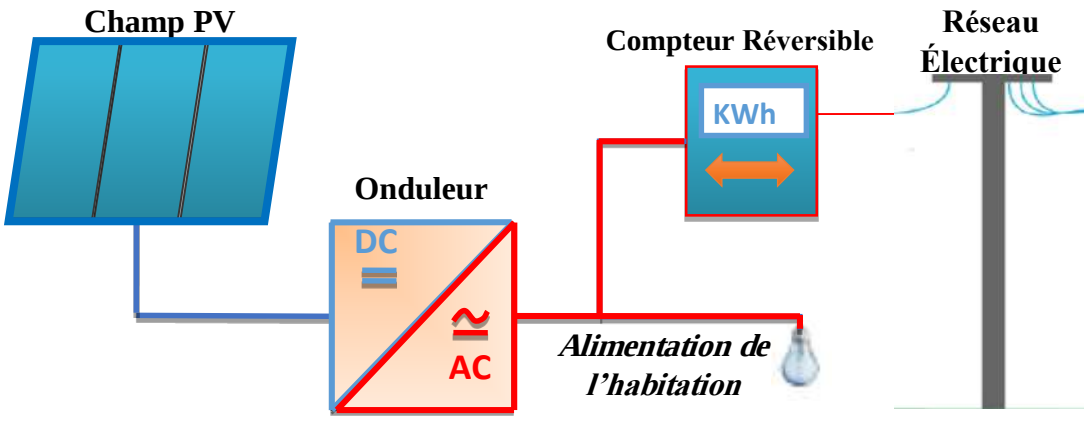


Figure 1.8 Système PV connecté au réseau.

Le système connecté au réseau est constitué de quatre éléments essentiels [35,36] :

• Le champ PV qui produit l'énergie électrique;

• L'onduleur est composé de deux convertisseurs. Le premier est un convertisseur DC/DC de type élévateur et le deuxième est un convertisseur DC/AC qui assure la synchronisation et l'injection de courant généré à partir de champ PV dans le réseau public. C'est-à-dire, l'onduleur joue le rôle d'une interface entre le champ PV et la compagnie d'électricité;

• Un compteur d'énergie.

• Le réseau électrique est un élément indispensable dans les SPVCRs.

On distingue deux types d'installations connectées au réseau :

- Installation centralisée : génération de la puissance PV à grande-échelle centralisée dans les stations de puissance PV;
- Installation décentralisée : forme alternative de génération de puissance distribuée dans les unités situées dans la région du consommateur.

1.3.4 Les Composants de système PV connecté au réseau

1.3.4.1 Le module PV

La quantité de courant et de tension générés, c'est-à-dire la puissance produite, d'une cellule solaire, n'est pas suffisante pour des applications réelles **[37]**. Par conséquent, les cellules sont connectées en série/parallèle pour augmenter la tension et le courant produit respectivement, comme est montré ci-dessous (figure 1.9).

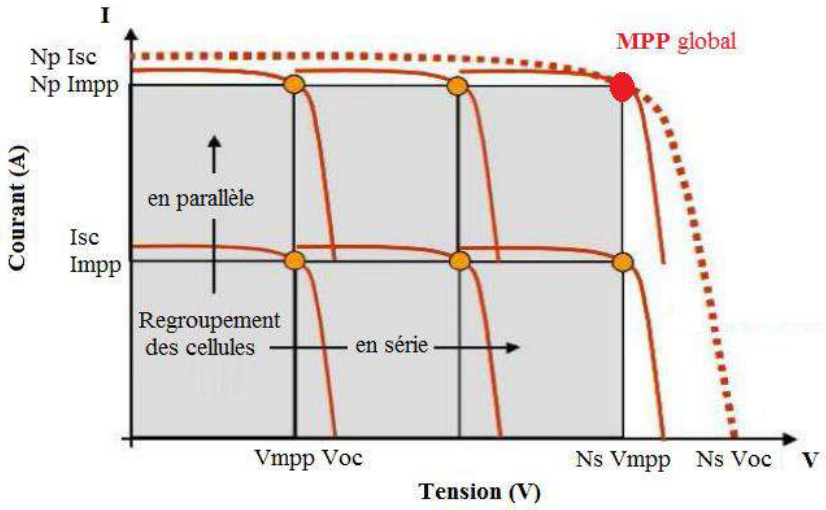


Figure 1.9 Caractéristiques d'un assemblage série/parallèle de Ns et Np cellules PV identiques **[33]**.

En fait, la série et l'interconnexion parallèle de plusieurs cellules solaires forment un module PV. De plus, ces cellules sont disposées en forme de grille à la surface des panneaux solaires. Dans le marché les modules PV sont construits par 36, 40, 54, 60, 72, ou même 108 cellules, afin d'obtenir un courant ou une tension souhaitable **[65]**. Chacun d'eux est relié en parallèle avec une diode de dérivation. L'objectif principal de cette diode est de protéger le groupe de cellules de la polarisation inversée.

a. Diode by-pass

Pour éviter les effets indésirables du phénomène "point chaud" (Hot-Spot), chaque panneau contient des diodes dites bypass dont le principe est de court-circuiter les cellules ombragées **[38]**. Les constructeurs de modules implantent généralement entre 2 à 5 diodes bypass dans les modules (dans le boitier de jonction du module).

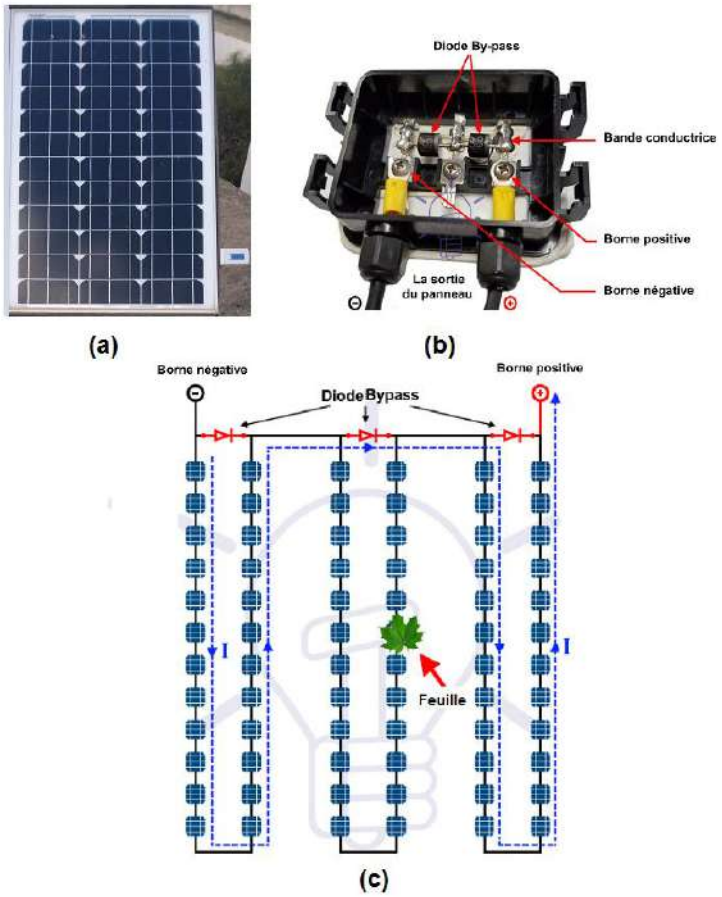


Figure 1.10 (a) Image réelle du module PV DIMEL (50 W), **(b)** Boîte de jonction du module PV, **(c)** Connexions de la sous-chaîne du module PV aux diodes by-pass.

Chacune des diodes bypass est associée à un sous-réseau de cellules du module. Lorsqu'une des cellules du sous-réseau est ombragée, la diode bypass devient passante, c'est-à-dire que le courant circule dans la diode en isolant ainsi du circuit électrique le sous-réseau de cellule associé.

1.3.4.2 Chaine PV (String PV)

Un string PV, appelé également chaine PV, est formé par un ensemble de modules PVs. Ces modules sont liés en série afin d'augmenter la tension générée pour obtenir une tension de sortie spécifiée.

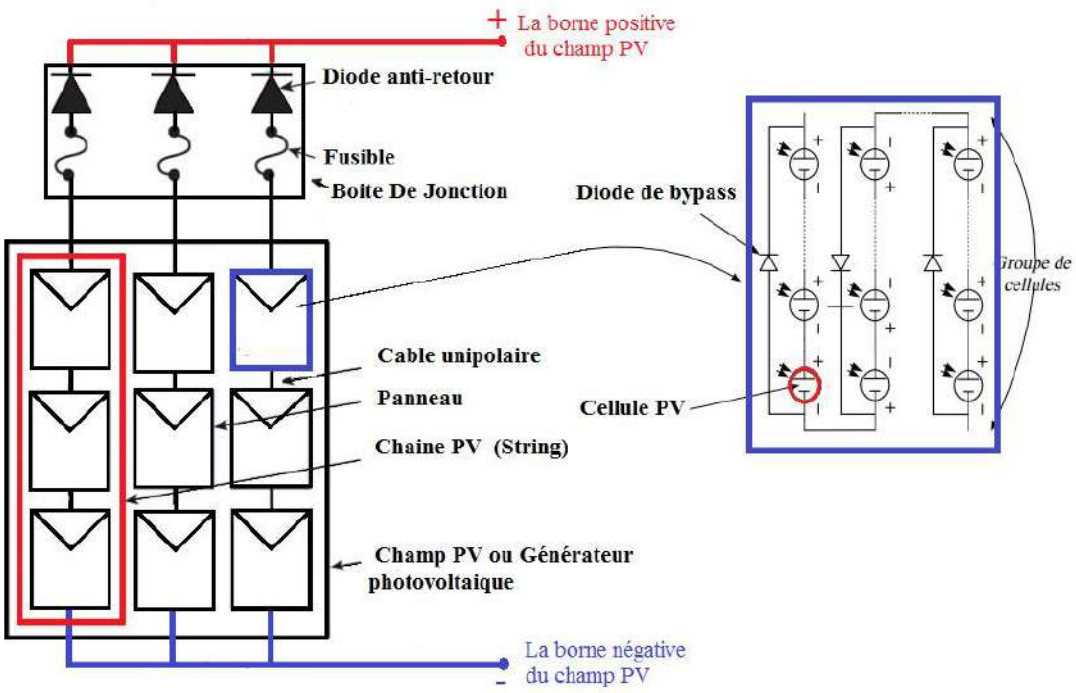


Figure 1.11 Schéma synoptique d'un champ PV [39].

1.3.4.3 Champ PV

Le champ photovoltaïque est un grand composant dans les systèmes PVs. Cette unité est capable de produire des puissances de quelques kW avec une tension convenable. Il est nécessaire d'arranger plusieurs strings PV en parallèle. Ce groupement forme un champ PV. Il existe plusieurs configurations possibles pour interconnecter les modules dans un champ PV : connexion série parallèle simple, connexion Total Cross Tied ou connexion Bridge Linked **[39]**. La figure 1.11 c'est une illustration détaillée d'un champ PV.

1.3.4.4 Dispositif de protection

Un champ PV doit posséder des dispositifs de protection afin de garantir certaines mesures de sécurité (pour des biens et des personnes) comme la diode anti-retour, les fusibles ou les disjoncteurs DC [**40**].

a) Diode anti-retour

Dans un champ PV, chaque string peut produire différentes valeurs de tension. Le string avec la tension la plus faible peut absorber un courant inverse provenant des autres strings [**41**]. Le déséquilibre des strings entre eux conduit donc à une baisse de production et peut se développer pour endommager les panneaux de ce string. Ceci est dû au courant inverse traversant ce dernier. Pour cela une diode anti-retour est nécessaire et doit être placée au bout de chaque string (voir figure 1.11).

Parmi les effets secondaires de la diode anti-retour, c'est la conduite à une perte de puissance du fait de la chute de tension causée par cette diode pendant le fonctionnement normal du champ PV. De plus, ces diodes peuvent se mettre en défaut et demandent par conséquent un contrôle régulier.

b) Les fusibles

Le courant de court-circuit disponible dans les systèmes PVs est limité et les dispositifs de protection contre les surintensités doivent fonctionner efficacement avec de faibles courants de fuite [**41**]. C'est pourquoi les chercheurs ont mené des travaux de recherche et des travaux de développement approfondis portant sur des fusibles spécifiquement conçus et testés pour protéger en toute sécurité les systèmes PVs avec des tensions de courant continu (VDC) élevées et de faibles courants de fuite. Parfois, le fusible remplace la diode anti-retour mais dans ce cas le string ne peut pas protéger contre les courants inverses.

c) Disjoncteur DC

Si le nombre de chaînes PVs est supérieur à deux, il faut par défaut monter un disjoncteur ou un porte-fusible en série avec chaque string [**41**]. En effet, dans le cas d'un court-circuit d'une chaîne, les autres chaînes débiteront dans ce court-circuit. Là encore ces composants doivent avoir la capacité à couper des courants de court-circuit, sous des

tensions VDC élevées. De plus, le disjoncteur DC sert à la protection contre les surintensités et aussi garantir la fonction anti-incendie [41].

1.3.4.5 Les convertisseurs PV

Les SPVCRs contiennent généralement deux types de convertisseurs. Le premier type est le convertisseur DC/DC qui sert à chercher et à suivre le point de puissance maximale du champ PV, tandis que l'autre est un convertisseur DC/AC qui a pour but de convertir et injecter l'énergie produite dans le réseau électrique.

A. Les convertisseurs DC/DC

Les convertisseurs DC/DC sont largement utilisés dans les systèmes PV comme une interface entre le module PV et une charge permettant le suivi du point de puissance maximale (MPP). Pour extraire la puissance maximale, il faut ajuster la charge pour correspondre au courant et à la tension du module PV. Le convertisseur doit être conçu pour être connecté directement au panneau PV et effectuer le fonctionnement pour rechercher le point de puissance maximale (MPPT). Les convertisseurs DC/DC ainsi que les systèmes de suivi des MPPs sont utilisés pour éviter ces pertes [42].

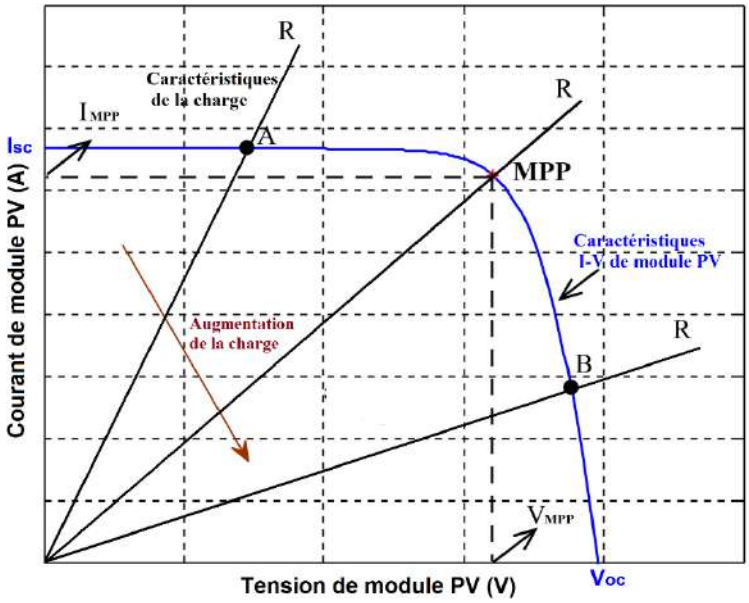


Figure 1.12 Points de fonctionnement résultant de la combinaison entre un module PV avec une charge résistive variable [42].

B. Les convertisseurs DC/AC

Le convertisseur DC/AC ou parfois appelé l'onduleur solaire est l'un des éléments les plus importants du système d'énergie solaire électrique. La fonction de base d'un onduleur est de convertir l'énergie à courant continu (DC) que les champs PVs ont générée en une énergie de courant alternative (AC) utilisable dans les maisons, les entreprises ou pour alimenter le réseau public [39].

C. Topologies des onduleurs PV connectés au réseau

Les onduleurs connectés au réseau peuvent être classés en trois types : les onduleurs centraux, les onduleurs strings et les onduleurs multi-string [39]. Les trois types sont illustrés par la figure 1.13.

i. Onduleur central

Les modules connectés en série forment une chaine et chacune de ces chaînes génère une tension suffisamment élevée pour éviter une amplification supplémentaire comme dans l'onduleur centralisé de la figure 1.13.a. Le niveau de puissance est atteint en connectant ces chaines en parallèle à travers la diode anti-retour et l'arrangement forme un tableau [39].

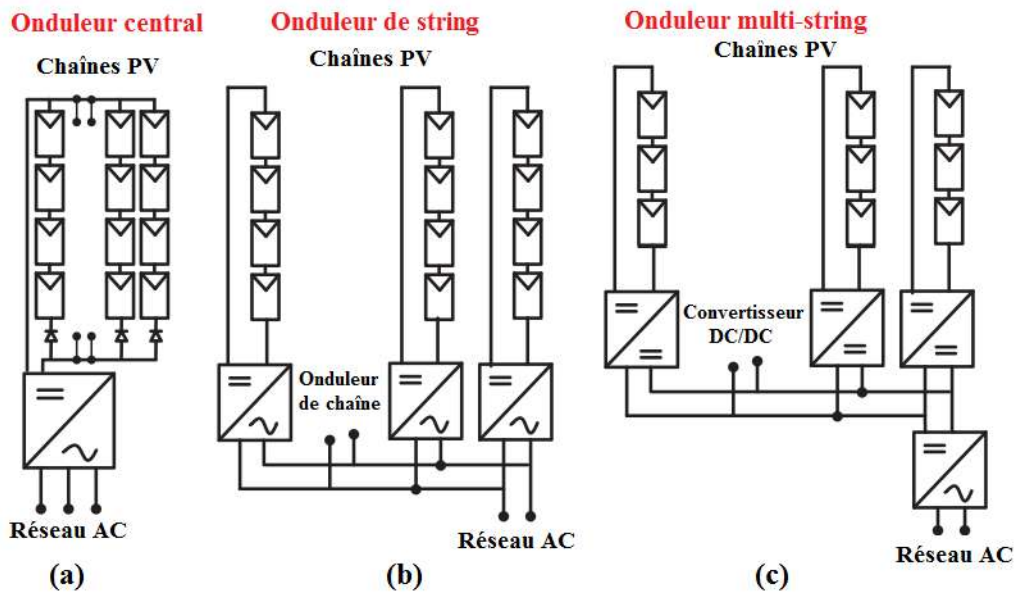


Figure 1.13 Différents types de configuration des onduleurs, (a) onduleur central, (b) onduleur de string et (c) onduleur multi-string [39].

Le tableau est interfacé avec le réseau AC via un onduleur centralisé. Les limites d'une telle configuration sont : l'interconnexion entre les modules PVs et de l'onduleur nécessitent des câbles DC à haute tension. L'utilisation d'un seul convertisseur DC/DC, pour l'ensemble du champ PV ne permet pas d'extraire avec précision le point de puissance maximale optimale dans des conditions d'ombrage partiel et de plus, le système peut être complètement arrêté dans ces conditions.

ii. Onduleur de string

Dans cette structure, chaque string a son propre convertisseur DC/DC pour extraire la puissance maximale plus un autre convertisseur DC/AC pour être connecté au réseau, donc le contrôle MPPT est séparé pour chaque chaine comme indiqué par la figure 1.13.b. Cette configuration n'a pas la perte de diode anti-retour et aussi garantit la disponibilité énergétique en cas de panne d'un onduleur par exemple [**39**].

iii. Onduleur multi-string

L'onduleur à plusieurs canaux comme démontré dans la figure 1.13.c a une commande

MPPT pour chaque string mais interfacé avec le réseau par un onduleur centralisé. Le principal avantage de cette configuration est son faible coût comparé avec l'onduleur pour chaque string. Cependant, il ne peut garantir la disponibilité de l'électricité en cas de panne de l'onduleur [**39**].

1.4 Les défauts et les anomalies dans le champ PV

Les propriétaires des systèmes PVs ont tendance à penser que le système PV est un équipement sans entretien en raison de sa configuration simple **[43]**. En fait, l'absence d'une pièce mobile réduit considérablement la possibilité d'une défaillance. Cependant, des enquêtes et des inspections sur le terrain, comme l'étude des systèmes PVs résidentiels au Japon **[44]** et le 1000 toits programmés en Allemagne **[45]**, montre qu'un grand pourcentage de défauts dans les systèmes PVs, nécessitent un entretien et le remplacement des pièces, et chaque événement qui amène à une production de puissance plus faible que celui attendu est considéré comme un défaut. Il est possible de classer les défauts en

cherchant différents aspects, en particulier : le composant qui présente le défaut, les causes, les effets et la réponse à ce défaut.

Afin de classer les défauts, il est intéressant de diviser le système PV en trois parties : la partie AC, la partie DC et entre eux l'onduleur. Si une défaillance est enregistrée dans la partie AC, les sources possibles sont : le déséquilibre du réseau, ou un problème dans la liaison avec l'onduleur ou avec la charge. La probabilité d'une défaillance due au réseau est fortement liée à l'état de ce dernier, et aussi aux points de raccordement. Les défauts dus à des erreurs internes de l'onduleur sont les plus courants et ils augmentent avec le vieillissement des composants. Ces défauts peuvent être liés avec tous les composants internes de l'onduleur, comme les commutateurs, le ventilateur, le MPPT ou les varistances (en fait une cause fréquente de l'échec de l'onduleur est un condensateur électrolytique défectueux)

Bien que la partie DC présente la plus faible quantité d'échecs, il est le plus compliqué à analyser et détecter l'origine du défaut en raison du nombre de composants inclus et des caractéristiques de sortie non linéaires des panneaux PV. Une variété de défauts peut être difficile à distinguer, pour cela une division doit être faite entre les défauts qui amènent à un arrêt complet de tout un string ou bien l'installation complète et ceux qui déterminent seulement une diminution de la puissance produite.

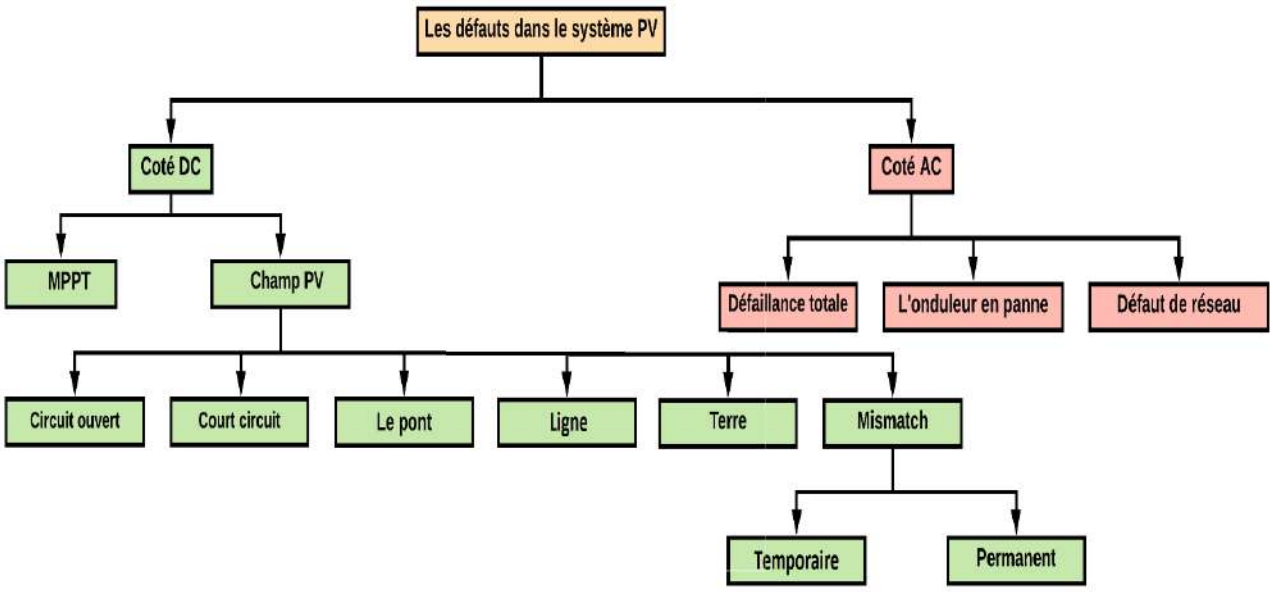


Figure 1.14 Les types de défauts se produisaient généralement dans les systèmes PVs [17].

Il est possible de classer les défauts dans le SPVCR comme a été montré dans figure 1.14. Ci-après on présente une description détaillée des défauts les plus courants dans le système PV.

1.4.1 Les défauts dans le champ PV

Au cours de son fonctionnement, le générateur photovoltaïque peut être sujet à plusieurs types de défauts [**46**]. On mentionne les plus courants : défauts de mismatch, défauts de circuit ouvert, défauts de terre, défauts de ligne (défaut de pont) et défauts de court-circuit.

a. Défaut de mismatch

Les défauts de mismatch sont causés par l'interconnexion des cellules PVs ou des modules qui n'ont pas des propriétés identiques ou qui connaissent des conditions de fonctionnements différentes les unes des autres. Ces défauts représentent un problème sérieux dans les modules ou les champs PVs dans certaines conditions parce que la sortie de l'ensemble des modules PVs dans les pires conditions est déterminée par la cellule ou le module PV avec la plus faible production.

De plus, ce type de défaut a été classé parmi les problèmes graves en raison de la complexité du processus de détection et de son impact significatif sur la réduction de la puissance. Les défauts de mismatch sont divisés en deux groupes principaux: les défauts de mismatch temporaires et les défauts de mismatch permanents [**46**].

i. Défaut de mismatch temporaire

Ces défauts reposent sur la distribution non uniforme de la température (couverture de neige) [**46**]. En fait, pendant la saison hivernale, la neige recouvre au sommet des modules PVs et entraîne une distribution de température non uniforme. Cet effet peut réduire considérablement la production énergétique des systèmes PVs.

ii. Défaut de mismatch permanent

Le défaut de mismatch permanent se produit principalement lorsque l'une de ces défaillances apparaît, telles que la dégradation de cellule ou module PV, le phénomène de point chaud (Hotspot) ou le défaut de soudage [**46**].

- **Défaut de dégradation**

La dégradation du champ PV est reliée directement à la durée de vie des modules c'est-à-dire son vieillissement au cours du temps de fonctionnement. Donc, cette dégradation résultera une augmentation de la résistance série entre les cellules ou les modules due à la diminution de l'adhérence des contacts ou à la corrosion causée par la vapeur d'eau [**46, 47**].

- **Défaut de soudure**

Des contacts à souder non fiables, des interconnexions de cellules peuvent provoquer une résistance élevée et par conséquent un échauffement dans le module PV. Dans les cas extrêmes, le risque d'incendie augmente. Ces mauvais contacts de soudure sont causés par une basse température de soudage où la soudure entre les cellules solaires et leurs rubans en contact est accidentellement déconnectée **[48]**.

- **Phénomène point chaud (Hotspot)**

Parallèlement à un certain nombre de cellules PVs, des diodes bypass sont intégrées dans le module PV. Ces diodes sont activées lorsqu'il y a un ombrage partiel. Par conséquent, certaines cellules reçoivent moins de courant que les autres. Dans ce cas, cette cellule devient polarisée en inverse et dissipe le courant au lieu de le générer, cela conduit à des points chauds qui peuvent entraîner un brunissement, des marques de brûlure ou, dans le pire des cas, un incendie [**49**].

b. Défaut de terre

Un défaut à la terre dans les champs PVs est un court-circuit électrique accidentel qui comprend la terre et un ou plusieurs conducteurs qui transportent normalement du courant [**50**]. Entre autre, les défauts à la terre dans les panneaux PVs soulèvent souvent des problèmes de sécurité car ils peuvent générer des arcs au point de défaut sur le chemin de défaut à la terre. En plus, d'un risque pour la sécurité des biens, ces défauts peuvent créer un risque d'incendie car le métal est chauffé par un courant court-circuité.

c. Défaut de circuit ouvert

Plusieurs situations conduisent à ce défaut. On trouve parmi elles la connexion en série entre le chemin d'acheminement du courant et la charge est accidentellement rompue, ouverte ou bien endommagée [**46,48**]. Une autre possibilité similaire peut se produire en raison de la mauvaise connexion entre les cellules PVs ou la rupture des câbles entre les modules PVs [**46,48**].

d. Défaut de ligne

Le défaut de ligne ou aussi appelé défaut de pont dans le champ PV pourrait être dû à l'une de ces raisons [**51,52**]:

- Défaillance d'isolation des câbles ;
- Court-circuit accidentel entre les conducteurs de courant ;
- Défauts de ligne dans la boîte de jonction DC, causés par des dommages mécaniques, des infiltrations d'eau ou de la corrosion ;
- Une connexion de faible résistance est établie entre deux points différents de potentiel dissemblable dans une même chaîne PV ou parfois dans différentes chaines.

e. Défaut de court-circuit

Un défaut de court-circuit est principalement dû à un mauvais câblage dans une chaîne PV ou entre les chaînes PVs. De plus, le vieillissement, les vibrations et l'abrasion des modules PV sont également parmi les principales sources de défauts de court-circuit [**53-55**]. Ce type de défaut peut être généré en fonction d'une connexion accidentelle entre deux points de tensions différentes dans une chaine ou un module PV.

Le reste de défauts et anomalies rencontrés dans un générateur PV sont résumés dans le tableau 1.2.

1.4.2 Méthodes de détection et de diagnostic des défauts dans la littérature

Afin de minimiser les pertes d'énergie dans les installations PVs, le diagnostic et la surveillance du système PV joue un rôle très important [**56**]. La puissance de sortie générée

par un système PV dépend de l'ensoleillement et de la température qui descendent sur la surface des cellules PVs. Si certains défauts existent dans un système PV qui entraînent une puissance de sortie inférieure, les utilisateurs peuvent penser que cette diminution est due à une faible irradiation solaire ou à d'autres conditions changeantes plutôt qu'à une condition de défaut. Même si les utilisateurs savent que le système PV a une anomalie, mais ils ne peuvent pas être en mesure d'identifier quel type de défaut, ni avoir une méthode pour déterminer quel défaut est présent. La détection et le diagnostic des défauts peuvent aider à découvrir les pannes dans les systèmes et à diagnostiquer leurs causes.

Tableau 1.2 Principaux défauts et anomalies rencontrés dans un générateur PV [**39**].

Eléments du générateur PV	**Origines de défauts et d'anomalies**
Générateur PV	-Feuilles d'arbres, déjections, pollution, sable, neige etc ; -Détérioration des cellules, fissure, échauffement des cellules ; -Pénétration de l'humidité, dégradation des interconnections, corrosion entre les cellules ; -Modules de performances différentes ; -Module arraché ou cassé ; -Modules court-circuités, modules inversés.
Boite de jonction	-Rupture de circuit électrique ; -Court-circuit du circuit électrique ; -Destruction de la liaison ; -Corrosion des connections.
Câblage et connecteur	-Circuit ouvert ; -Court-circuit ; -Mauvais câblage (module inversé) ; -Corrosion des contacts ; -Rupture du circuit électrique.
Diode de protection (diode de bypass et diode anti-retour)	-Destruction des diodes -Absence ou non fonctionnement de diodes -Inversion de la polarité des diodes au montage, diode mal connectée

En outre, les pronostics des défauts constituent un autre domaine de recherche pour prédire les performances de production sur la durée de vie restante et l'éventuelle défaillance du système à l'avenir.

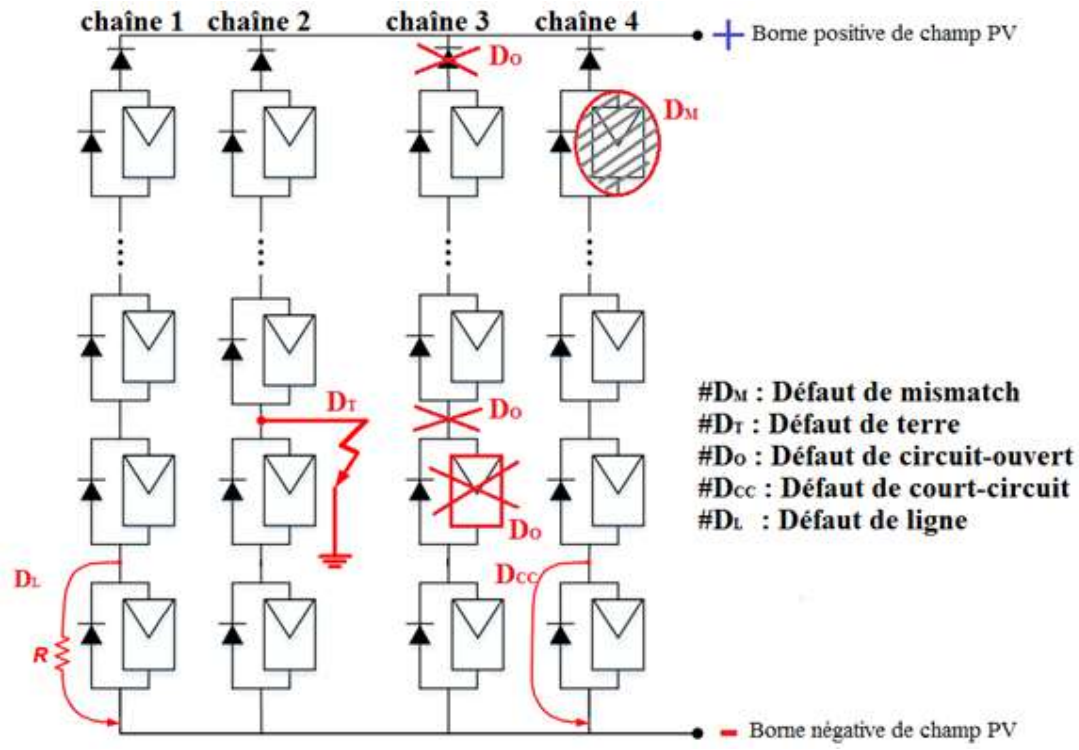


Figure 1.15 Diagramme schématique de différents types de défauts dans le champ PV [57].

Le contrôle de performances régulières du le fonctionnement des SPVCRs sont nécessaires pour assurer une production d'énergie fiable et optimale. L'objectif est la détection précoce des défauts dans les systèmes PVs et changer les conditions de fonctionnement pour empêcher les pertes d'énergie et financières subséquentes pour l'opérateur. Ainsi, un système simple et sûr à manipuler à long terme pour vérifier les performances et l'état du système PV; donc ce processus est nécessaire pour garantir la sécurité et réduire les pertes de ce dernier. Le but de systèmes de détections et diagnostics est non seulement la découverte des dysfonctionnements, mais aussi la mise à la disposition des informations sur la source de défaillance la plus probable. Les méthodes de détection, diagnostic et pronostics des défauts constituent les approches de pronostics et de gestion [58]. De plus, ces méthodes offrent de nombreux avantages, tels que:

- la prévision à l'avance du temps d'échec;
- Réduire au minimum l'entretien imprévu, les cycles d'entretien prolongés, l'efficacité par des mesures de réparation en temps opportun;
- réduire les temps d'arrêt;

- Améliorer la qualification, l'assistance à la conception et le soutien logistique pour les futurs systèmes.

Cette section montrera les récents algorithmes de détection de pannes qui sont implémentés pour détecter les modules PV défectueux dans les chaînes ou les champs PVs. Dans la littérature, il a été identifié que les algorithmes de détection des pannes PV peuvent être classés en deux catégories principales: ***les techniques basées sur la modélisation statistique et mathématique***, et ***les techniques basées sur l'intelligence artificielle ou l'apprentissage automatique*** telles que la logique flou et les réseaux de neurones artificiels [**22**].

Plusieurs critères ont été utilisés pour évaluer l'efficacité d'une telle méthode, certains d'entre eux sont énumérés ci-dessous :

- La capacité de détecter le défaut survenue à un stade précoce ;
- L'aptitude de définir correctement la nature de défauts détectés ;
- La robustesse de l'algorithme proposé contre les erreurs de modélisation et les données bruitées ;
- La qualité et le montant des données requises à utiliser lors de la détection et du diagnostic des pannes.

De nombreuses approches de diagnostic ont été proposées pour détecter les défauts dans un système PV. Nous résumons dans cette partie quelques différentes méthodes proposées.

- **Techniques d'intelligence artificielle (IA) dans la détection de défaut PV**

Cette section décrit l'utilisation des algorithmes basés sur l'intelligence artificielle (IA) dans le diagnostic des défauts. Dans ces algorithmes, les modèles empiriques implicites provenus de l'analyse des données disponibles, sont utilisés pour détecter et diagnostiquer les défauts du système PV. Par conséquent, ces approches reposent principalement sur les données enregistrées pour permettre d'établir des modèles grâce à l'apprentissage automatique ou les méthodes d'intelligence artificielle. Ci-après, un certain nombre de travaux les proposés dans la littérature.

En 2011, Karatepe & Hiyama ont proposé une méthode de diagnostic des pannes utilisant plusieurs réseaux neurones [**59**]. La couche d'entrée de la technique proposée contient le rayonnement solaire, la température de cellule, la tension et le courant au point

de puissance maximale (V_{mpp} et I_{mpp}). La technique proposée est capable d'identifier l'emplacement des modules PVs court-circuités dans une chaine. Cependant, la technique ne peut pas identifier les conditions d'ombrage partiel, et le défaut PV des chaines.

tadj et al a proposé une approche capable de détecter trois différents niveaux d'irradiation solaire dans les systèmes PVs [60]. L'algorithme utilise un système basé sur la logique flou pour prédire la couverture nuageuse dans le ciel qui peut être classée comme suit : ciel nuageux, ciel nuageux partiel et ciel clair.

Les fonctions d'appartenance sont illustrées à la figure 1.16, où K_0 est l'indice d'un ombrage partiel. Cette approche permet de détecter les éventuels défauts dans l'installation photovoltaïque sans avoir recours à un capteur de rayonnement solaire. Cependant, l'approche nécessite une image satellite visible du ciel afin de définir le paramètre d'indice de clarté. Une autre limite de l'approche proposée est de ne pas indiquer si l'ombrage partiel est dû à un nuage, à la poussière, à des hélicoptères ou à un brouillard affectant les modules PVs.

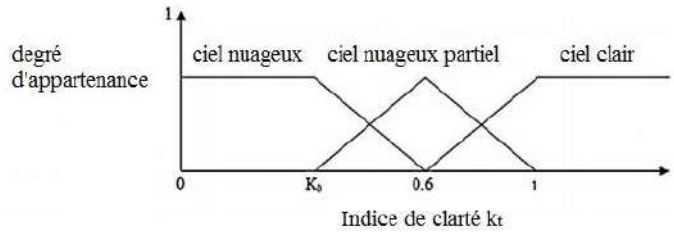


Figure 1.16 Les fonctions d'appartenance de la logique floue [60].

Gisement solaire par télédétection (Solar Radiation by Teledectection « GISTEL ») est une méthodologie satellitaire basée sur un simple modèle. Il est utilisé pour estimer l'irradiation solaire globale à partir des données Meteosat. La méthodologie adoptée a plusieurs étapes qui sont résumées par le schéma ci-dessous. Ainsi, dans ce travail, une méthode de simulation proposée basée sur l'environnement MATLAB/Simulink a été utilisée pour détecter et puis diagnostiquer les défauts. Cette méthode se compose de deux parties principales. Premièrement, le modèle GISTEL a été appliqué pour estimer les séries chronologiques horizontales mondiales du rayonnement solaire;

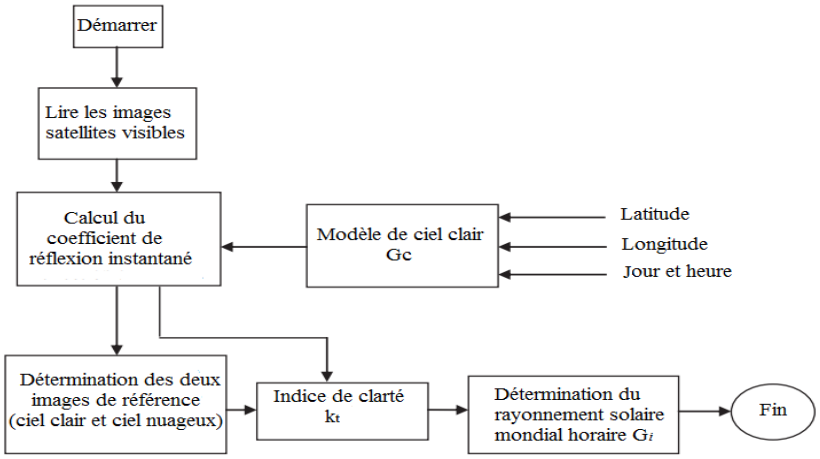


Figure 1.17 Les différentes étapes du modèle GISTEL [**60**].

Ensuite les résultats obtenus ont été comparés à ceux mesurés. À la deuxième étape, une comparaison de la puissance de sortie de GISTEL et des données mesurées a été utilisée. La méthode proposée est présentée par la figure 1.17.

Un autre algorithme de diagnostic de défaillances dans le système PV basé sur le système de classification Neuro-Flou a été proposé par Bonsignore, et al [**61**]. L'algorithme a été implémenté à l'aide du logiciel MATLAB/Simulink. Six paramètres ont été utilisés afin de mettre en œuvre l'algorithme développé qui sont : I_{mpp}, V_{mpp}, V_{oc}, Isc, S1 et S2, où S1 et S2 sont calculés à l'aide de l'équation (1.5) et de l'équation (1.6) respectivement :

$$S1 = \frac{I_{mpp} - I_{sc}}{V_{mpp} - V_{oc}} \tag{1.5}$$

$$S2 = \frac{0 - I_{mpp}}{V_{oc} - V_{mpp}} \tag{1.6}$$

Le système Neuro-Flou proposé est capable de détecter diverses failles dans le système PV, telles que la diode court-circuitée dans un module PV et les conditions de l'ombrage partial. L'objectif de cette approche est d'appliquer des techniques ANFIS pour modéliser le module PV à partir des données calculées au moyen d'un modèle hybride.

Pour cela, les étapes suivantes sont exécutées : **1)** entrée et sortie de modèle sont respectivement (T_{PV}, G_{PoA}) et (I_{mpp}, V_{mpp}, I_{sc}, V_{oc}, S1, S2) ; **2)** génération du système ANFIS ; **3)** formation et test du modèle ANFIS. De cette façon, au lieu de sauvegarder l'ensemble des courbes I-V, seulement les six variables sont considérées et utilisées comme cible de sortie du modèle général ANFIS dans des conditions normales et défectueuses.

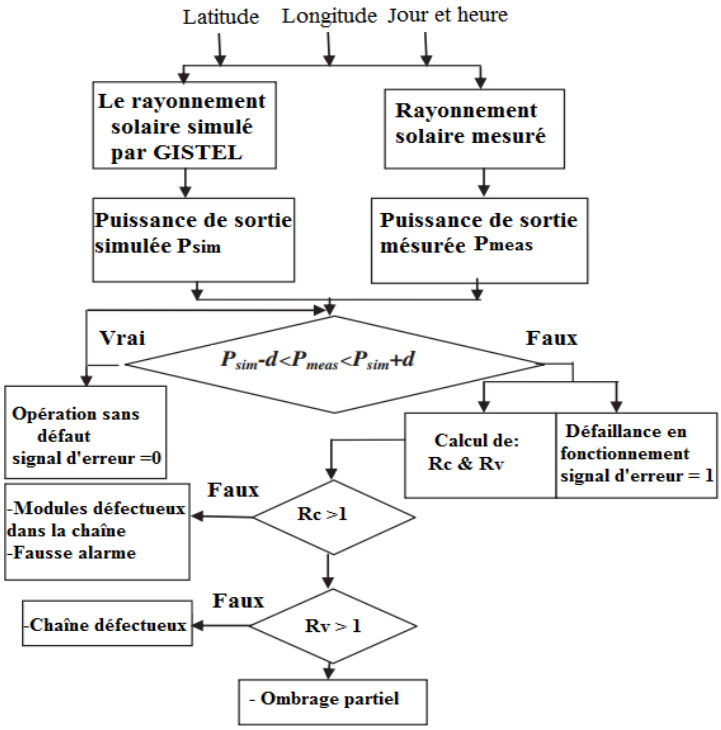


Figure 1.18 Organigramme de la procédure de détection des défauts [61].

La figure 1.18 montre de manière qualitative comment les défauts peuvent affecter la forme d'une courbe I-V dans des conditions de fonctionnement normales (ligne rouge) où trois zones distinctes peuvent être identifiées : ***i)*** une région légèrement inclinée au-dessus du point de court-circuit ; ***ii)*** un coude ou un «genou» dans la courbe dans la région du point de puissance maximale ; ***iii)*** une région en pente raide sous le point de circuit ouvert.

La figure 1.19 présente six caractéristiques (nommées Cx) (voir les lignes pointillées rouges) qui diffèrent de la courbe I-V en condition de fonctionnement normal pour certains symptômes dus à la présence de défauts. Chaque défaut génère un ensemble de symptômes identifiables à l'aide des variations d'un ou de plusieurs paramètres.

Les signaux de résidus normaux et défectueux, $RX_{id,nm}$ et $RX_{id,ft}$, sont calculés comme montrés dans les équations (1.7) et (1.8).

$$RX_{id,nm} = X_{id} - X_{nm} \tag{1.7}$$

$$RX_{id,ft} = X_{id} - X_{ft} \tag{1.8}$$

Où : X est l'attribut adéquat ; les index (id, nm, ft) indiquent respectivement les états idéal, normal et défaillant.

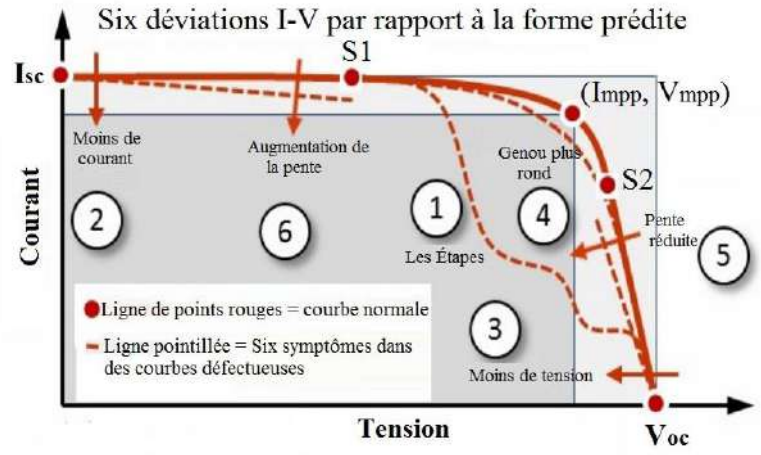


Figure 1.19 Modèle de six points le long de la courbe I-V [**61**].

Les normes normales et défectueuses, Nn et Nf, sont calculées comme suit [**61**] :

$$|Nn| = \sqrt{\sum_{i=1}^{6} \left(RX_{id,nm}\right)^2} \tag{1.9}$$

$$|Nf| = \sqrt{\sum_{i=1}^{6} \left(RX_{id,ft}\right)^2} \tag{1.10}$$

Pour terminer, l'évaluation des normes consistait à comparer les normes obtenues (Nn et Nf) avec des valeurs de seuil (S) définies expérimentalement. Cette comparaison pourrait identifier cinq cas de fonctionnement : cas de fonctionnement normal, cas de fonctionnement de court-circuit de diode, cas de panne de terre inférieure, cas de fonctionnement de défaut de terre supérieure et cas de fonctionnement avec ombrage partiel. La stratégie de classification de l'état du système basée sur le test de norme est résumée dans la figure 1.20 [**61**].

La principale limitation de l'algorithme est le nombre de variables utilisées dans la classification des défauts, puisque les unités MPP utilisées dans les installations PV ne peuvent pas identifier toutes les variables requises, en particulier les Voc et Isc.

Les auteurs Chine et coll., ont proposé une technique de diagnostic des défauts pour le système PV basée sur le réseau de neurone artificiel (ANN) [**62**]. La différence entre la puissance de sortie du générateur PV mesurée et simulée est d'abord calculée et comparée à la valeur de seuil (S) pour détecter l'occurrence des défauts.

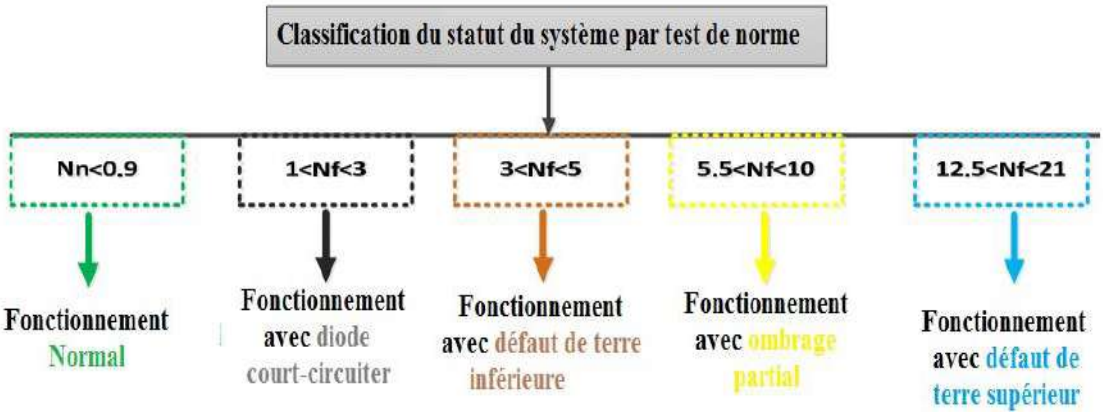


Figure 1.20 Classification de l'état du système PV par Norme-tests [**61**].

Ensuite, l'analyse des attributs principaux, dérivée par les caractéristiques I-V de chaque chaîne PV séparée, est élaborée pour diagnostiquer et localiser les défauts.

Un certain nombre d'attributs tels que le courant, la tension et le nombre de pics de la courbe I-V de la chaîne PV sont calculés. Par la suite, ces attributs sont collectés et stockés dans un ensemble donné pour certaines conditions de fonctionnement, telles que l'irradiation solaire et la température du module PV. Les réseaux de perceptron multicouches (MLP) à de fonction radiale de base (RBF) ont été utilisés pour détecter la défaillance.

Les différentes étapes de cette approche pour la détection, le diagnostic et la localisation sont résumées dans la figure 1.21 [**62**].

Cette méthode pourrait détecter, diagnostiquer et localiser huit types de défauts décrits dans le tableau 1.2 [**35**]. A partir de la figure 1.19, on voit que cette méthode utilise deux algorithmes différents, nommés ***algorithme 1*** et ***algorithme 2***. Le premier algorithme est une approche basée sur le seuil de signal qui permet la discrimination de six cas défectueux [**62**]. Les paramètres utilisés dans ce premier algorithme et leurs descriptions sont récapitulées dans le tableau 1.3 [**62**].

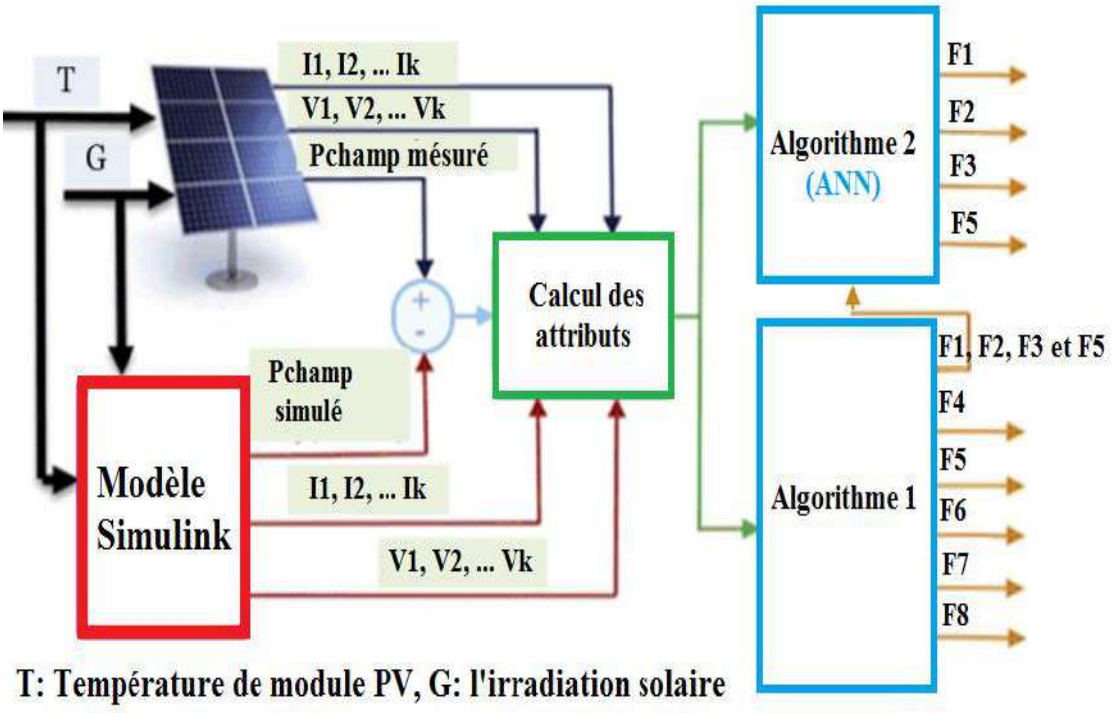
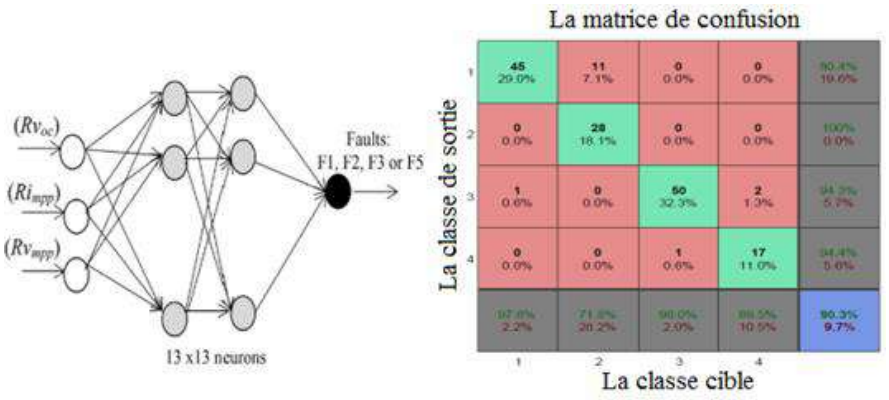


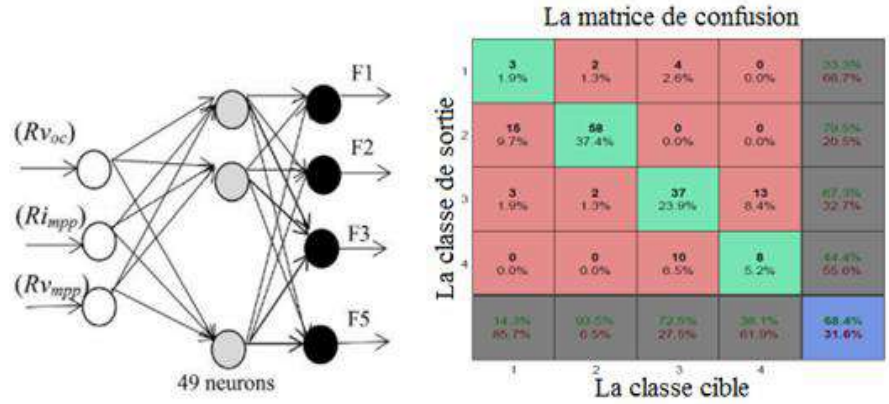
Figure 1.21 Les différentes étapes de la méthode ANN [62].

Table 1.3 Les défauts qui pourraient être détectés correctement par la stratégie de Chine et al [62].

Type de défaut	Désignation	Code
Module	- Défaut de court-circuit dans toute diode de dérivation ou (cellule ou module) - Défaut de diode de dérivation inversé ou (cellule ou module) - Défaut de diode de dérivation shunté de (cellule ou module) - Défaut de circuit ouvert dans n'importe quelle cellule (ou module)	F1 F2 F3 F4
Défaut de connections	- Résistance de connexion entre les modules PV	F5
Défaut d'ombrage partial	- L'effet de l'ombre dans les modules PV avec le fonctionnement normal des différents composants de la chaîne PV	F6
L'effet de l'ombre avec la diode de dérivation défectueuse	- L'effet de l'ombre dans un groupe de cellules avec leur diode de dérivation en panne.	F7
L'effet de l'ombre avec défaut de connexion	-L'effet de l'ombre dans un groupe de modules reliés par une résistance connectée.	F8



(a)



(b)

Figure 1.22 (a) Le réseau de neurone MLP avec la matrice de confusion, **(b)** Le réseau de neurone RBF avec la matrice de confusion [62].

Les matrices de confusions de classification des sorties pour les deux réseaux neurones sont montrées dans la figure 1.22. F4 est manquante parce qu'elle est associée aux défauts de circuit ouvert dans n'importe quel module PV qui peut être détecté en utilisant des rapports mathématiques présentés dans l'article de Chine et al [62].

De plus, Dhimish et al, ont proposé une méthode efficace pour diagnostiquer plusieurs défauts qui surviennent dans les installations PVs. La méthode est basée sur la modélisation et l'analyse de la courbe des caractéristiques I-V et sur un système de classification floue [63]. L'algorithme de diagnostic passe par six couches : la première couche de l'algorithme est la transmission des valeurs de l'irradiation mesurée et la température du module PV au logiciel VI LabVIEW afin de générer la courbe P-V théorique attendue (système sain). La couche qui suit est pour la détermination si une défaillance s'est produite dans le système SPVCR, pour cela deux rapports ont été identifiés. Le rapport de puissance théorique (PR)

et le rapport de tension théorique (VR) ont été utilisés pour classer la région des défauts. Les rapports de puissance et de tension sont donnés par les expressions suivantes [**46**] :

$$PR = \frac{P_{T,G}}{P_{T,G} - nP_0} \tag{1.11}$$

$$VR = \frac{V_{T,G}}{V_{T,G} - nV_0} \tag{1.12}$$

Où : $P_{T,G}$ et $V_{T,G}$ sont respectivement la puissance et la tension théorique, générées par le champ PV, dans des conditions météorologiques spécifiques de la température (T) et de l'irradiation (G). Q est le nombre de modules PV. P_0 et V_0 qui dénotent respectivement la puissance et la tension au fonctionnement maximale et sont mesurées dans des conditions d'essai normalisées (Standard Test Conditions (STC) : T = 25 °C, G = 1000 W/m ²).

Les capteurs de courant et de tension ne sont pas parfaits donc ils possèdent une précision d'acquisition inférieure à 100%. Cette tolérance a été prise en considération par l'introduction dans le calcul de rapport de puissance et le rapport de tension [**63**]. En conséquence, les deux rapports possèdent donc deux limites, la limite inférieure de PR et la limite inférieure de VR qui sont citées ci-dessous :

$$PR\ limite\ basse = \frac{P_{T,G}}{(P_{T,G} - nP_0)\eta_{capteur}} \tag{1.13}$$

$$VR\ limite\ basse = \frac{V_{T,G}}{(V_{T,G} - nV_0)\eta_{capteur1}} \tag{1.14}$$

Où $\eta_{capteur}$ et $\eta_{capteur1}$ sont respectivement l'efficacité des capteurs. Le premier pour les deux capteurs courant et tension, tandis que le deuxième est l'efficacité du capteur de tension, $\eta_{capteur}$ comme est exprimée dans l'équation (1.15).

$$\eta_{capteur} = \eta_{capteur1}(efficacité\ de\ capteur\ de\ tension) + \eta_{capteur2}(efficacité\ de\ capteur\ de\ courant) \tag{1.15}$$

D'autre part, les limites hautes PR et VR sont celles données par les équations (1.11) et (1.12), et les valeurs réelles mesurées sont calculées comme suit :

$$PR\ mésuré = \frac{P_{T,G}}{P_{mésuré}} \tag{1.16}$$

$$VR\ mésuré = \frac{V_{T,G}}{V_{mésuré}} \tag{1.17}$$

Les limites de détection hautes et basses de PR et VR sont évaluées pour examiner le système PVCR dans plusieurs niveaux d'irradiation.

Des courbes polynomiales de 3ème ordre a pour but de générer une fonction de régression qui décrit les performances des courbes créées par les points théoriques à l'aide du logiciel VI LabVIEW. L'algorithme global de détection des défauts dans SPVCR est expliqué dans la figure 1.23. La couche 5 montre les données mesurées par rapport aux courbes polynomiales de 3ème ordre générées par le logiciel VI LabVIEW. Le PR et la VR mesurés peuvent être évalués à l'aide de la formule suivante :

$$PR\ mésuré\ vs. VR\ mésuré = \frac{V_{T,G}}{V_{mésuré}}\ vs. \frac{P_{T,G}}{P_{mésuré}} \tag{1.18}$$

Dans le cas où le PR mesuré vs. VR mesuré est hors de portée :

$$F\ limite\ haute < PR\ mésuré\ vs. VR\ mésuré < F\ limite\ basse \tag{1.19}$$

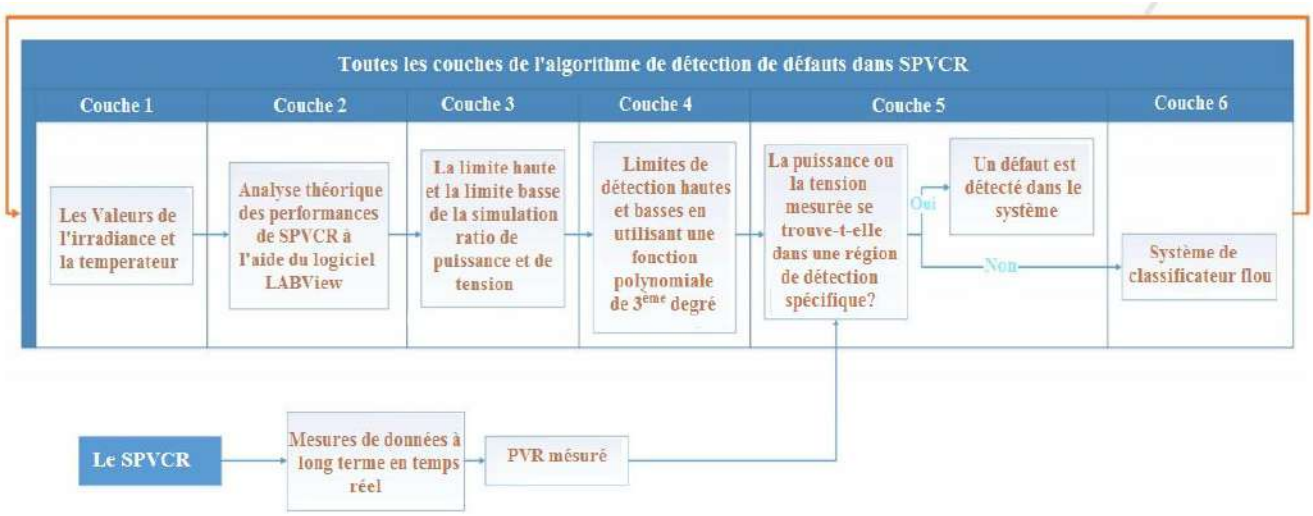


Figure 1.23 Toutes les couches de l'algorithme de détection de défaut dans SPVCR [**63**].

Par conséquent, l'algorithme de détection des pannes ne peut pas identifier le type de l'erreur qui s'est produite dans le SPVCR. Cependant, il peut prédire deux conditions erronées possibles qui pourraient avoir lieu dans le SPVCR. Donc, les échantillons hors région sont traités par un classificateur de logique floue comme illustré à la figure 1.23, qui est la couche 6.

Cette approche peut détecter avec précision les défauts suivants : la présence d'ombrage partiel (PS) dans le champ PV, Un module PV court-circuité et/ou partiellement ombré, deux modules PVs en court-circuit et/ou partiellement ombré. (Q-1) modules en court-

circuit ou en partiellement ombré, d'où Q dénote le nombre total des modules PVs. La complexité du processus de détection est l'un des désavantages de cette approche, qui dépend de plusieurs facteurs tels que les rapports de puissance et de tension.

La complexité du processus de détection est l'un des désavantages de cette approche, qui dépend de plusieurs facteurs tels que les rapports de puissance et de tension. Par conséquent, son efficacité dépend fortement de l'efficacité des composants d'instrumentation.

- **Avantages et inconvénients des méthodes d'intelligence artificielle**

Les approches et les contributions proposées dans cet axe (basées sur l'intelligence artificielle) ont prouvé leurs aptitudes et efficacités pour la détection et la classification des défauts et des anomalies qui surviennent aux systèmes PVs au cours du fonctionnement. En outre, ces approches utilisent les données collectées précédemment du système en état sain et défectueux, et ne nécessitent aucune connaissance sur la configuration ou les composants du système pour produire un modèle qui permet de distinguer entre les différents états de ce dernier.

Néanmoins, l'ensemble de données est une condition préalable pour pouvoir créer un modèle de classification. Par conséquence, les systèmes d'acquisition et d'enregistrement de données sont nécessaires. De plus, la qualité de ces données joue un rôle essentiel puisque elle est capable de donner une fausse information sur l'état réel du système ou mener à des résultats erronés. Même, certains algorithmes ne supportent pas les données bruitées. Ces faiblesses pourraient évidemment affecter la fiabilité, l'efficacité et aussi l'utilisation en temps réel de ces approches. Pour résoudre ces faiblesses une phase de prétraitement de données est préférable, mais elle conduit à une augmentation de la complexité de l'approche et affecte le temps d'exécution, l'implémentation et l'utilisation en temps réel du modèle.

- **Techniques basées sur la modélisation mathématique et statistique dans la détection de défauts PVs**

En plus des risques d'incendie et des problèmes de sécurité, les défauts des systèmes PVs peuvent entraîner une grande quantité de perte d'énergie. L'énergie perdue en raison de défauts PVs a été analysée et catégorisée dans les systèmes PVs domestiques britanniques et il a été estimé que la perte d'énergie annuelle atteignait 18,9% dans les

systèmes PV en raison des défauts [64]. Par conséquent, il est nécessaire de surveiller les performances du système PV, d'étudier le modèle de défaut et de développer les méthodes de détection de ces défauts.

Les méthodes qui sont basées sur la modélisation statistique et mathématique ont montré leur efficacité dans la détection et le diagnostic des défauts dans les systèmes PV. En fait, ils exigent un minimum de connaissance préalable de la configuration du système, et n'utilisent que des données du système sain et défectueux pour surveiller, détecter, diagnostiquer et localiser les défauts.

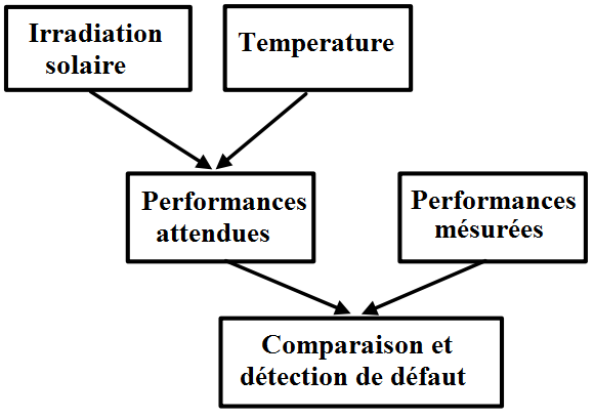


Figure 1.24 Comparaison des performances d'un système PV [65].

Les approches de comparaison des performances PV sont comparées entre celles qui sont mesurés de celles simulées en temps réel [66-68]. Récemment, plusieurs approches basées sur la comparaison des performances PVs ont été proposées pour la détection des défauts. En général, ils comparent la performance réelle avec la performance attendue. La méthode de détection des défauts est simple : une différence significative dans les performances de sortie produites (mesurées) et prédites (simulées) peut indiquer un défaut. Comme le montre la figure 1.24, l'évaluation des performances comporte généralement plusieurs composants, notamment des informations météorologiques telles que l'irradiation solaire et la température, les performances PV attendues comme référence, les performances PV réelles mesurées, la comparaison des performances et la détection des défauts.

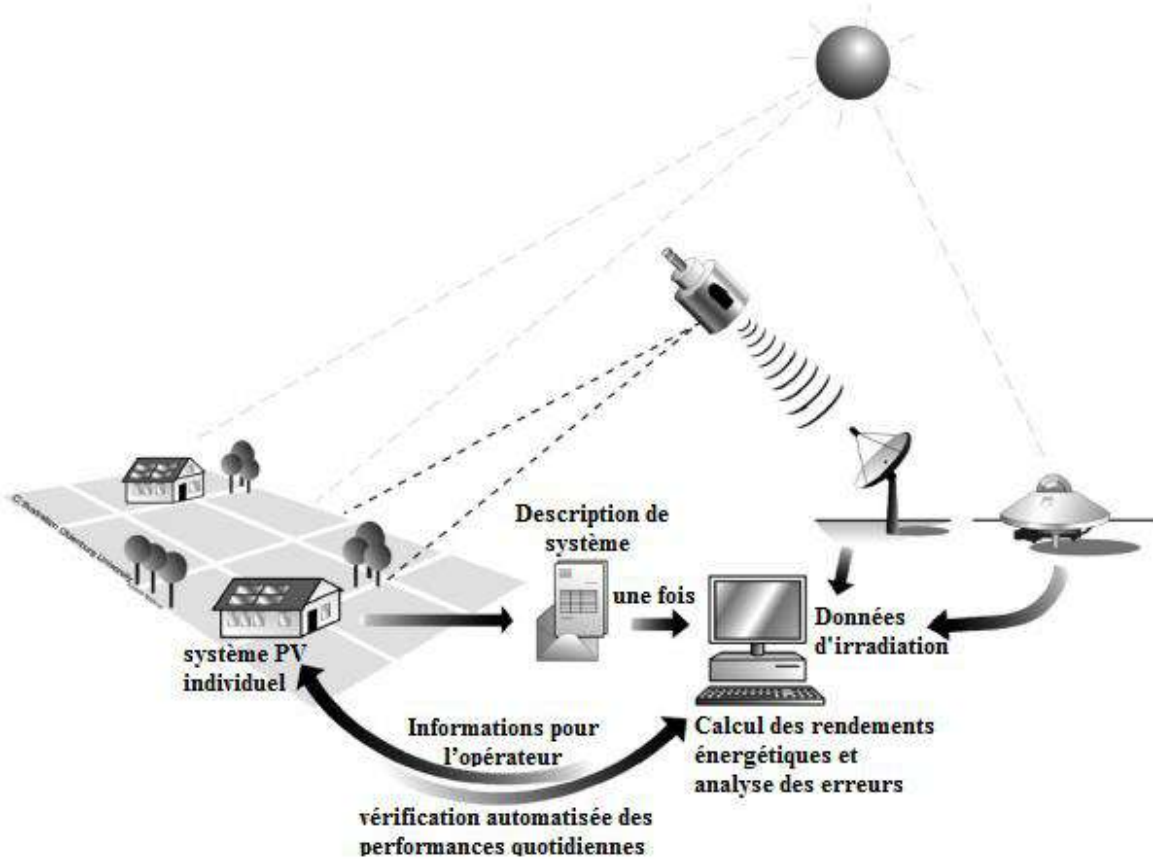


Figure 1.25 Un aperçu général sur PVSAT2 [**69**].

Par exemple, pour prévenir les pertes énergétiques et financières subséquentes dans les systèmes PVs, une comparaison automatique des performances a été proposée dans [**69**], qui surveillent la différence entre le rendement énergétique simulé et le rendement énergétique mesuré en temps réel (voir figure. 1.25). Le système de détection des défauts rassemble l'irradiation solaire délivrée par le satellite et la température ambiante qui sont introduites dans un modèle de simulation pour prédire la puissance de sortie AC du PV (P_{sim}). Pendant ce temps, il surveille la puissance de sortie réelle d'AC (P_{actual}) et la compare avec P_{sim}. Quatre catégories des défauts généraux sont les suivantes : perte d'énergie constante, perte d'énergie changeante, couverture de neige et défaillance totale.

La limitation existe dans cette approche : puisqu'il surveille la puissance PV sur une période de temps (équivalent au rendement énergétique), il a une réponse lente. Par exemple, cette méthode de détection des défauts peut prendre au moins une journée.

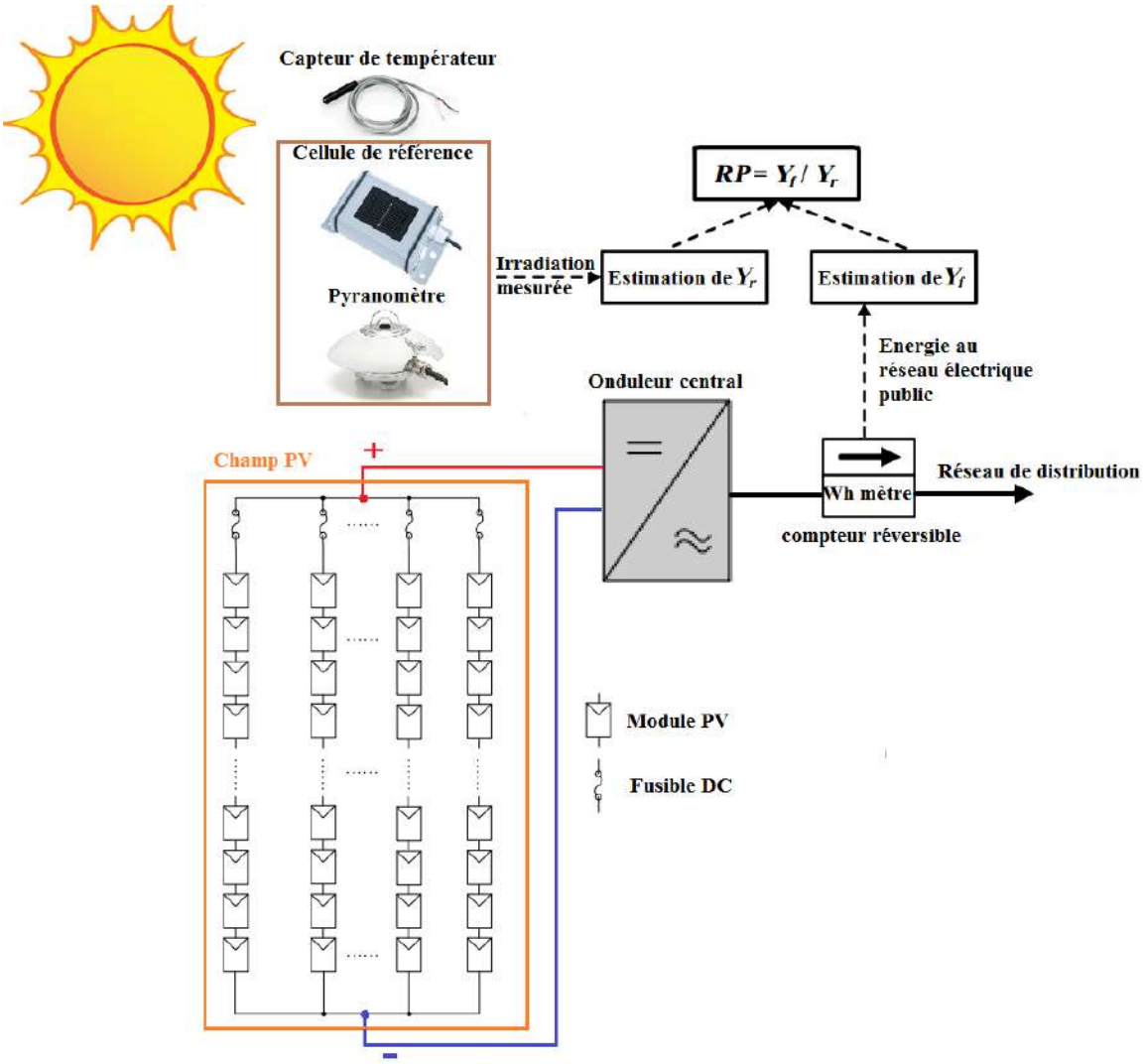


Figure 1.26 Méthode de calcul du rapport de performance dans un SPVCR [**69**].

Le rapport de performance (RP) est proposé dans [**70,71**] comme paramètre normalisé du rendement énergétique du système PV pour évaluer les performances du système. Indépendamment de l'orientation et de l'inclinaison du panneau, PR prend en compte les effets globaux des pertes du système afin qu'il puisse être utilisé pour la détection de défauts. Il est généralement donné par la formule 1.20 en utilisant le rendement final (Y_f) sur le rendement de référence (Y_r).

$$PR = \frac{Y_f}{Y_r}\ (sans\ dimension) \tag{1.20}$$

La figure 1.26 illustre comment obtenir Y_f et Yr dans un SPVCR. Plus précisément, Y_f représente la production d'énergie AC normalisée vers le réseau électrique public. Elle est

définie par l'équation 1.21, où E est la puissance AC de sortie nette et P0 est la puissance nominale PV.

$$Y_f = \frac{E}{P_0}\left(\frac{kWh}{kW}\right) ou\ (les\ heures) \tag{1.21}$$

De l'autre côté, Y_r représente les conditions d'irradiation solaire normalisées. Il est défini par expression 1.22, où H est l'irradiation solaire totale dans le plan par (kWh / m2) et G_{STC} est la valeur de l'irradiation dans la condition d'essai standard qui est égale à 1000 W/m²

$$Y_r = \frac{H}{G_{STC}}\ (les\ heures) \tag{1.22}$$

Les valeurs de rapport de performance des systèmes PVs sont communément rapportées sur une base mensuelle ou annuelle (à long terme). Pour une base à court terme, comme quotidienne ou hebdomadaire, le rapport de performance donne une meilleure résolution et il peut être utilisé pour la détection de défauts dans les systèmes PVs.

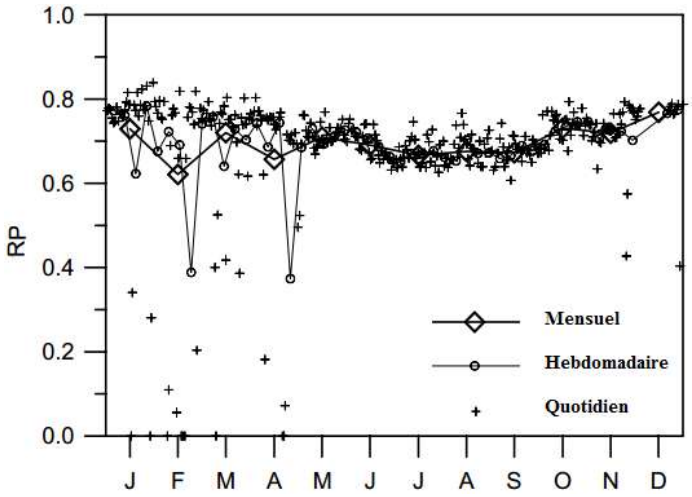


Figure 1.27 Valeurs de PR quotidiennes, hebdomadaires et mensuelles pour un système PV pour 2001 [72].

Par exemple, comme le montre la figure 1.27, la plupart des valeurs de rapport de performance se situent entre 0,65 et 0,8 lorsque le système PV fonctionne normalement. Cependant, le taux de performance est considérablement réduit quand les défauts ou les anomalies son présents dans le système PV, tels que des ombrages partiels, un dimensionnement inapproprié du système, des erreurs MPPT, des pannes d'onduleur et des défauts dans les panneaux PVs. En outre, la figure 1.27 montre que les valeurs de rapport de performance pour les intervalles plus petits (par exemple, quotidiennement) donnent

une meilleure résolution de détection des défauts et une réponse plus rapide que les données mensuelles ou hebdomadaires [**71**].

Pour surmonter la limitation du PR précédemment discutée (rapport de performance), le capteur des pertes du système (Lc) a été proposé dans [**73**] pour faire distinguer l'anomalie au niveau du système. Lc représente les pertes dues au fonctionnement du générateur PV. Cette grandeur est définie en équation (1.23), où Y_r est le rendement de référence et Y_A est la production d'énergie quotidienne du générateur PV installé par kW. De plus, L_c peut être divisé en deux types de pertes, à savoir la perte de capture thermique (L_{ct}) et la perte de capture diverse (L_{cm}) dans l'équation (1.24).

$$L_c = Y_r - Y_A \tag{1.23}$$

Dans cet axe, les auteurs Chouder et Silvestre ont proposé un système de surveillance automatique qui est par la suite utilisé pour la détection et le diagnostic des défauts du système PV [**74**]. L'approche présente dans ce travail est basée sur l'analyse des pertes de puissance. De plus, la méthode d'analyse est reposée sur deux indicateurs pour identifier la nature de défaut survenant dans le système PV. Dans la première phase, un algorithme de détection de défaut a été établi pour déterminer l'état du système PV (sain ou défaillant). À cette fin, l'équation 1.24 a été introduite.

$$L_{c_sim} = Y_r(G, T_c) - L_{a_sim}(G, T_c) \tag{1.24}$$

Où L_{c_sim} est les pertes capturées, T_R (G, T_C) est le rendement de référence mesuré et T_C est la température du module PV. Après avoir identifié les pertes capturées dans le système PV, une phase statistique a été utilisée comme représentée par l'équation suivante :

$$L_{c_sim} - 2\sigma < L_{c_mes} < L_{c_sim} + 2\sigma \tag{1.25}$$

Où L_{c_mes} est la perte mesurée du système PV examiné, σ est l'écart type qui est calculé à partir des pertes simulées pour des conditions de ciel clair.

Le processus d'analyse consiste à mesurer en continu les pertes de puissance du côté courant continu (pertes de capture) et à examiner ensuite si ces pertes se situent dans des limites théoriques (basant sur l'équation 1.25). Le système est considéré comme défectueux si les pertes de puissance mesurées dépassent ces limites. Enfin, afin de diagnostiquer le type de défaut, des indicateurs jouent un rôle très important afin d'accomplir le processus de détection. Ces indicateurs sont les rapports de courant et de tension indiqués dans les expressions suivantes :

$$R_C = \frac{I_{PV_sim}}{I_{PV_mes}} \tag{1.26}$$

$$R_V = \frac{V_{PV_sim}}{V_{PV_mes}} \tag{1.27}$$

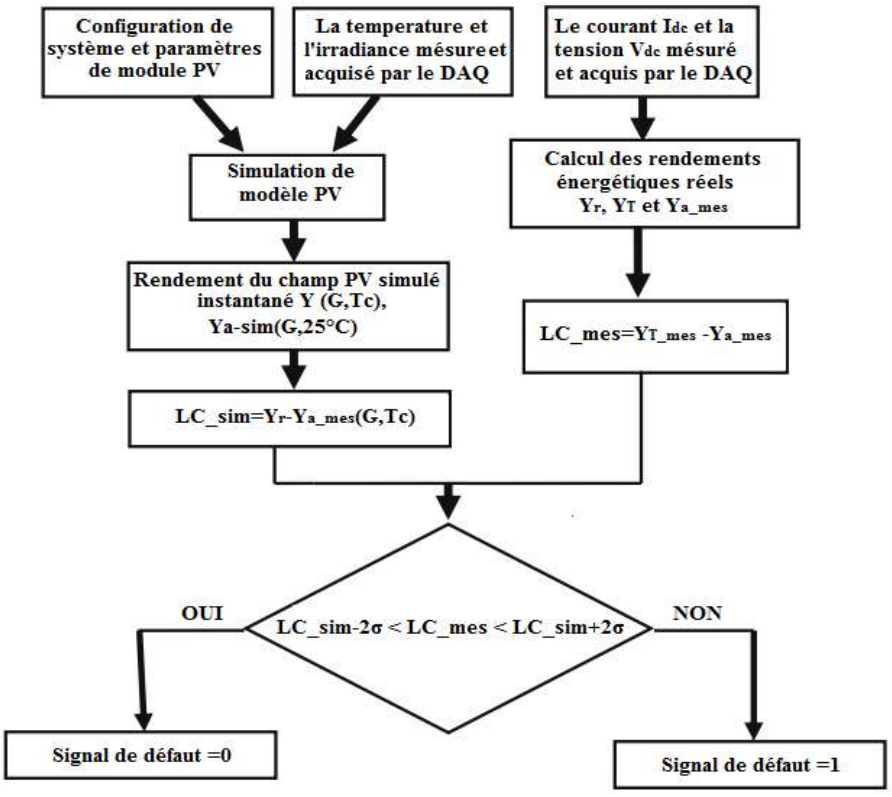


Figure 1.28 Organigramme proposé du système de détection [74].

L'organigramme donné à la figure 1.28 illustre une explication détaillée de la procédure proposée pour diagnostiquer les défauts dans une installation PV.

De plus, une procédure sophistiquée basé sur la même idée est proposée dans [75]. Le processus de comparaison entre les rendements simulés et mesurés est proposé, tandis que l'analyse de l'écart des erreurs dans le courant et la tension permet de déterminer l'état du système soit sain ou bien défectueux et spécifie également le type de défaut. Un système de surveillance est mis en œuvre dans LabVIEW [76]. Ce système est réalisé pour une modélisation sophistiquée et la supervision du SPVCR.

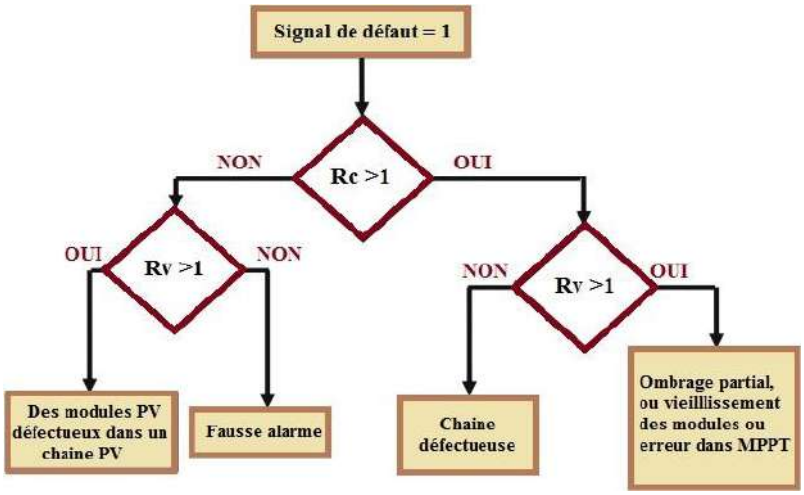


Figure 1.29 Organigramme de la procédure de diagnostic des défauts [74].

Finalement, l'objectif de ce nouveau système de surveillance est d'aider et d'intégrer l'approche de diagnostic. En outre, [77] décrit une procédure améliorée basée sur les indicateurs révisés tels que le courant et la tension avec la détermination et la définition des valeurs de seuil pour chaque indicateur respectivement. Ces seuils ont une forte corrélation avec la structure du système; le défaut est défini en fonction du seuil calculé et du seuil comparé.

- **Avantages et inconvénients des méthodes de modélisation mathématique et statistique**

Les approches présentées ci-dessus montrent leur capacité de distinguer les différents états, sain et défectueux, du système PV. En conséquence, elles sont capables d'aider à la détection et le diagnostic de défauts. D'autre part, parmi les inconvénients majors de ces méthodes, on peut citer la complexité de la méthode proposée et parfois certains modèles ne sont pas généralisés pour tous les systèmes, donc il ne peut pas être appliqué pour n'important quel système PV, même des fois, il faut faire une étude profonde sur le système pour développer un modèle. Aussi, l'implémentation dans des circuits programmables est difficile ou impossible.

- **Techniques basées sur les méthodes de traitement du signal**

La réflectométrie dans le domaine temporel (TDR) est une approche de mesure pour déterminer les caractéristiques du fil ou du câble électrique en injectant certaines formes d'ondes (généralement des signaux échelonnés ou impulsionnels) et en observant les

formes d'ondes réfléchies. TDR compare les réflexions de l'environnement de ligne inconnue à celles générées avec l'impédance donnée. Par conséquent, TDR peut être utilisé pour dépanner les fils ou câbles défectueux. En plus de la détection et de la classification des défauts, TDR est capable de localiser le défaut, ce qui est particulièrement utile dans un grand système électrique.

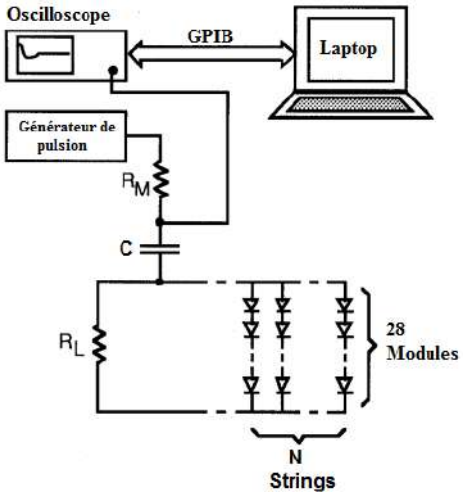


Figure 1.30 Schéma de principe pour la configuration TDR dans le domaine PV [78].

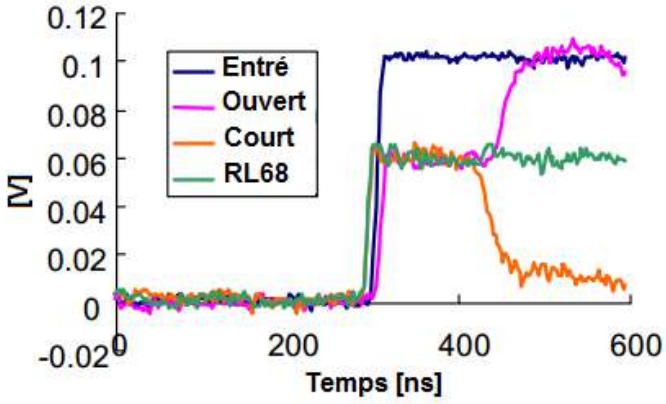


Figure 1.31 Entrée échelon et les formes d'ondes réfléchies avec le câble [79].

La localisation des défauts à l'aide du TDR a été proposée dans les champs PVs [79,80]. Le diagramme schématique est illustré à la figure 1.30. Par exemple, la figure 1.31 montre une entrée échelon dans une chaîne PV et ses formes d'onde réfléchies. Les formes d'onde diffèrent selon le type de conditions, telles qu'un défaut de circuit ouvert, un défaut de court-circuit ou une charge de résistance.

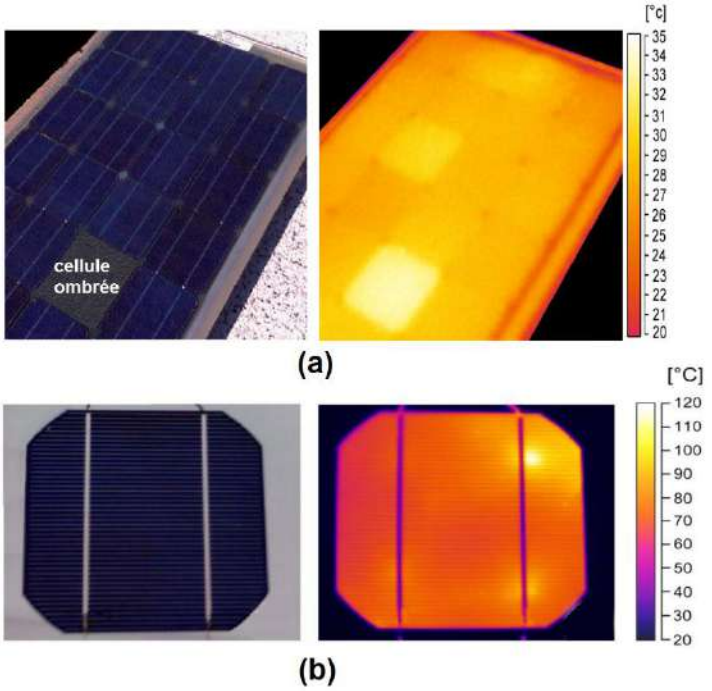


Figure 1.32 (a) Un mismatch causé par une cellule ombrée avec son image thermique correspondante, **(b)** l'effet d'une polarisation inverse sur une cellule PV en silicium monocristallin avec l'image thermique de ce dernier [**84**].

En tant que version améliorée du TDR, la réflectométrie dans le domaine temporel à spectre étalé (SSTDR) a été proposée dans le commerce pour détecter les défauts de câblage des avions [**81**]. Le SSTDR a de faibles niveaux de signal de test et une immunité au bruit élevée, ce qui est en fait une solution fiable pour détecter les défauts de fils sous tension (dans les systèmes sous tension). Pour la détection de défaut à la terre PV, les pics d'autocorrélation générés par la condition de défaut souterrain SSTDR sont nettement plus élevés que la normale. Selon cela, un algorithme de détection de défaut dans les systèmes PV a été proposé par Alam et al [**82**]. Cependant, cette approche n'intègre pas de classification des défauts.

La thermographie infrarouge est un processus qui utilise une caméra thermique pour détecter le rayonnement (température) provenant d'un objet, puis afficher une image de la distribution de la température de ce dernier. Les images de la distribution de température détectée sont appelées thermo grammes. Ces images permettent de voir les objets produisant la chaleur invisible à l'œil nu. Il est largement utilisé dans la maintenance prédictive et la surveillance de l'état du système.

Cette méthode a été utilisée pour identifier certains défauts de discordance, tels que les points chauds sur les modules PV, qui peuvent provoquer des dommages irréversibles et une perte de puissance sur les modules PVs (voir la figure 1.32) [**83, 84**]. Étant donné que l'inspection IR est périodique et répétitive, elle peut être coûteuse si elle nécessite du personnel de maintenance PV, des coûts de main-d'œuvre accrus au cours de la durée de vie des modules PVs.

- **Avantages et inconvénients des méthodes de traitement de signal**

Les méthodes et les approches basées sur le traitement de signal sont capables de détecter et dans certains cas de distinguer quelques défauts dans le module PV ou dans le champ PV. D'autre côté, le coût, la limitation et aussi la complicité de ces approches sont élevés. De plus, elles demandent des appareils très chers.

1.5 Conclusion

Dans ce chapitre, nous avons présenté les principales caractéristiques et les technologies des éléments constitutifs d'un système PV. Par la suite, on a expliqué brièvement comment augmenter le courant ou la tension d'un générateur PV, et nous avons brièvement décrit quelques types de défauts et anomalies qui peuvent survenir au champ PV au cours de fonctionnement.

Nous avons aussi présenté quelques méthodes de diagnostic de défauts proposées dans la littérature. Ces méthodes montrent une limitation de distinction entre les défauts. Ainsi, les méthodes basées sur l'exploitation des données sont identifiées comme la plus prometteuse [**85**].

CHAPITRE 2

DESCRIPTION DU SYSTEME PV CONNECTE AU RESEAU

Chapitre 2

Description du Système PV Connecté au Réseau

2.1 Introduction

Les systèmes PVs nécessitent un suivi tout en assurant une production d'énergie sans perte. Plusieurs stratégies de monitoring existent en fonction de la puissance de l'installation PV et la classe de monitoring choisie. Le monitoring est réalisé à partir des données climatiques et électriques du système PV; il est de ce fait nécessaire de procéder à l'acquisition de ces données en temps réel à travers un système d'acquisition adéquat.

Le présent chapitre a pour objectif la démonstration des différentes étapes de la réalisation d'un système d'acquisition et d'enregistrement de données qui permettent d'étudier le comportement dynamique d'un système PV connecté au réseau déjà en fonction.

2.2 Description du système PV connecté au réseau

2.2.1 Les composants d'un SPVCR monophasé

Un système photovoltaïque connecté au réseau (SPVCR) est composé généralement de 4 blocs fonctionnels différents schématisés dans la figure 2.1 :

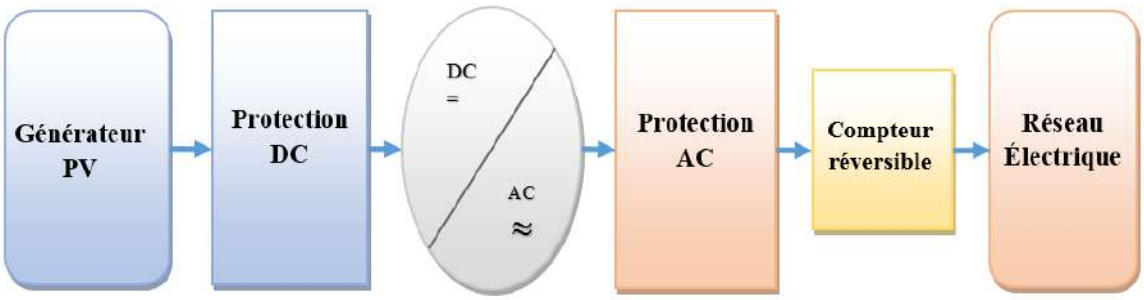


Figure 2.1 Schéma block d'un système photovoltaïque connecté au réseau.

Chacun des blocs du système possède un role qui sera défini dans ce qui suit.

a) Le générateur PV

Parfois appellé le champ PV est formé par des modules PVs pouvant être connectés en série et/ou en parallèle qui convertissent l'énergie du soleil en électricité continue. Les modules PVs sont montés et fixés sur un support métallique.

b) Le dispositif de conditionnement de puissance (DCP)

Le DCP est constitué d'un convertisseur DC/DC ayant pour fonction la recherche et la poursuite du point de puissance maximale (MPPT) et d'un convertisseur DC/AC qui permet d'adapter les caractéristiques de l'énergie produite par le générateur PV à celle du réseau électrique.

c) Les protections

Elles sont constituées de fusibles, d'un disjoncteur différentiel, d'un disjoncteur magnétothermique et de varistance. Elles assurent la protection de la SPVCR et du réseau électrique.

d) Compteur réversible

Il compte le flux d'energie dans les deux sens, l'un pour la consommation, et l'autre pour le surplus de la production.

e) Le réseau éléctrique

Le réseau de distribution électrique basse tension avec une tension AC monophasé 230V ou triphasé 400Vet une fréquence 50 Hz.

2.2.2 Description du système PV connecté au réseau installé

Notre système PV connecté au réseau est constitué d'un générateur PV de 10 modules PVs, du fournisseur DIMEL de puissance crête 50W et de tension nominale de 16.44 V. Le générateur PV est formé de 10 modules soit deux branches en parallele de 5 modules connectés en series. Les caractéristiques électriques de modules PV sont montrées dans le tebleau 2.1 ci-dessous.

Le générateur PV est connecté à un onduleur monophasé de 700 Wc. Les caractéristiques de ce dernier sont présentées dans le tableau 2.2 [**86**].

Tableau 2.1 Caractéristiques techniques de panneau solaire DIMEL 50W.

Paramètre	V_{oc}(V)	I_{sc}(A)	V_{mpp}(V)	I_{mpp}(A)	P_{mpp}(W)
Valeur	21.2	3.11	16.44	3.04	50

L'onduleur est branché sur une phase du réseau basse tension (BT) de Sonelgaz à travers toutes les protections nécessaires pour protéger le réseau et le SPVCR.

Tableau 2.2 Caractéristiques techniques de l'onduleur Solis mini 700W.

Les données d'entrée	
Puissance PV recommandée	500-700 Wp
Courant d'entrée max.	10 A
Tension d'entrée Max. à 1000W/m² et 10°C	450 V
La plage de tension MPP	50-400 V
Temps de connexion au réseau	50s
Les données de sortie	
Puissance nominale	700 W
Puissance Max	800 W
Rendement Max	> 96 %
Facteur de distorsion	< 3 %
Courant de sortie nominale	10.9 A
Consommation en fonctionnement	1 W
Tension nominale du réseau	230 V / 50Hz

Le schéma général de la SPVCR est représenté dans la figure 2.2.

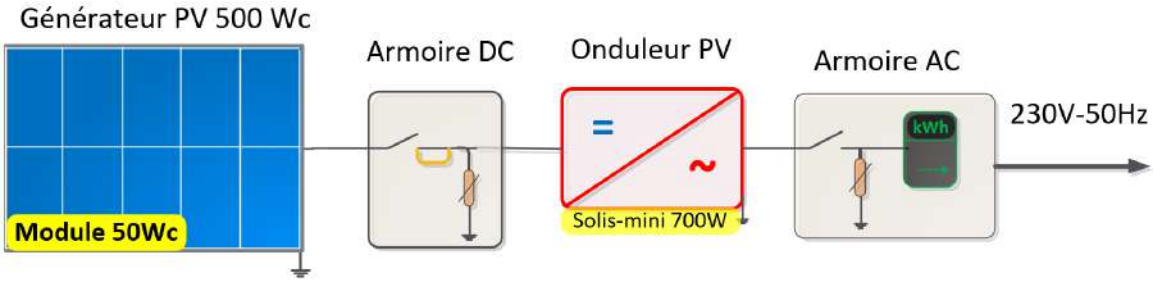


Figure 2.2 Schéma générale de la SPVCR.

2.3 Description du système d'acquisition de données

Le monitoring de l'installation se base sur la surveillance des principales grandeurs. Pour cela, il faut réaliser un système qui permet l'acquisition et le transfert des grandeurs influant le système tel que l'irradiation et la température, et les grandeurs électriques générées tel que le courant et la tension.

Les différents éléments constituant notre système d'acquisition sont représentés dans la figure 2.3.

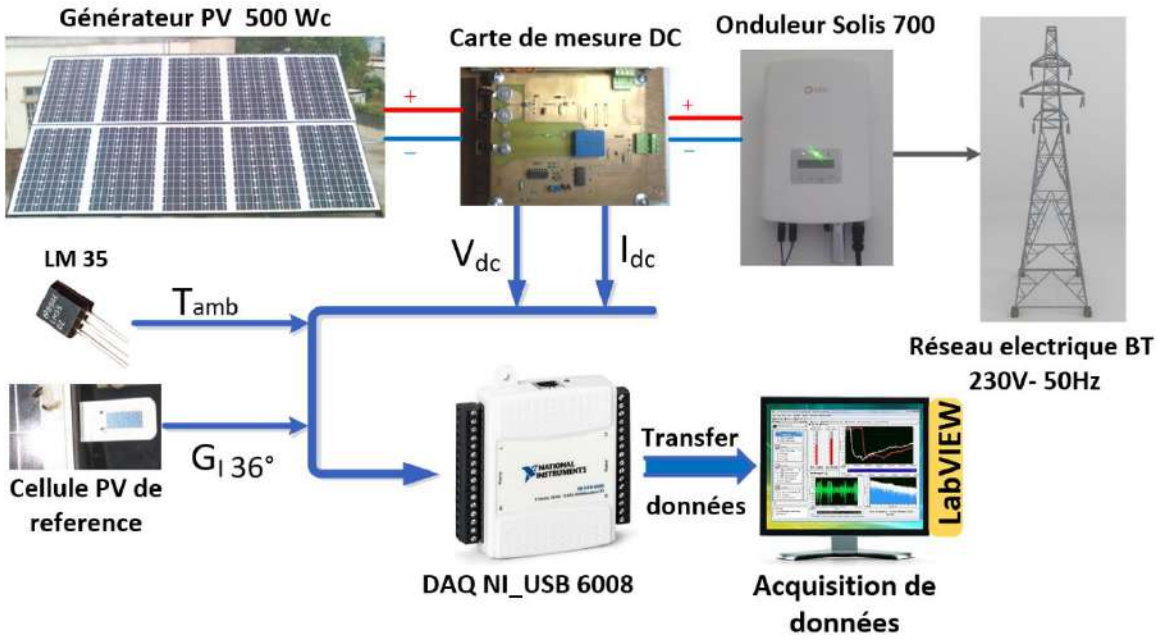


Figure 2.3 Aperçu global du SPVCR.

2.3.1 Les Capteurs

Les capteurs servent à acquérir les différents paramètres nécessaires (courant, tension, éclairement, température). Chacune de ces grandeurs donnée par le capteur passe par un circuit de conditionnent. Dans notre réalisation, nous avons utilisé la carte d'acquisition NI_DAQ 6008 USB qui intègre des circuits de conditionnement pour tous les signaux d'entrées.

Chaque paramètre est mesuré par un capteur spécifique (tableau 2.3). Dans ce qui suit, nous allons présenter les différents capteurs utilisés dans notre réalisation pour obtenir les résultats souhaités.

2.3.1.1 Mesure de courant

Le capteur de courant permet de mesurer le courant fourni par le champ PV qui est nommé I_{dc_mes}. Pour cela, on utilise un capteur à effet Hall. Ce dernier est un capteur qui exploite l'effet Hall pour produire une tension proportionnelle au courant mesuré. Ce captage de courant est basé sur le capteur ACS712 (figure 2.4), qui peut acquérir les signaux de courant alternatif (AC) ou continue (DC) [87]. Les valeurs du courant (maximal et minimal) qui peuvent être mesurées par le capteur sont de ±20A et le signal de courant actuel peut être lu via un port analogue.

La tension fournie a une valeur positive qui contient un décalage dans l'intervalle de [0 4.5 V] Volts avec une caractéristique linéaire.

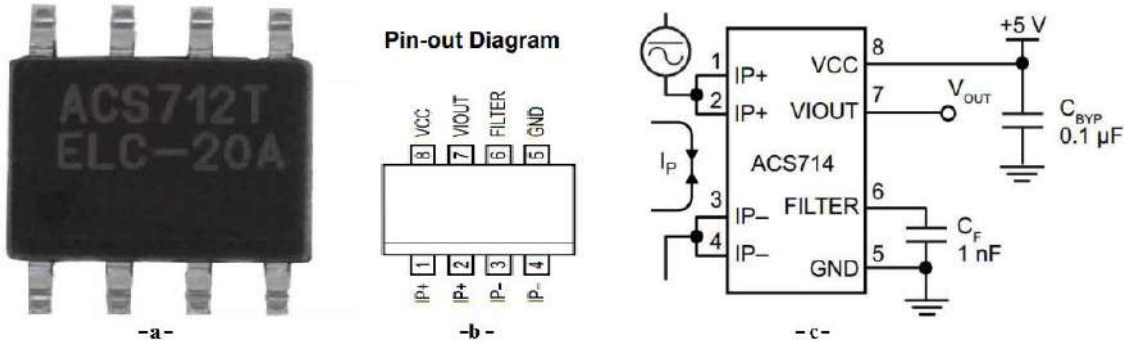


Figure 2.4 Capteur ACS712: -a- la conception du capteur, -b- le brochage du capteur -c- Schéma de brochagede du capteur [**87**].

2.3.1.2 Mesure de tension

Pour acquérir la tension V_{dc_mes} délivrée par le générateur PV qui peut aller jusqu'à 105V, le choix s'est porté sur le capteur LEM LV25-P (figure 2.5) [**88**]. Le principe de fonctionnement du capteur est basé sur l'effet Hall. Ainsi, il fournit une isolation galvanique entre le côté haute tension et le côté de mesure. Il est connecté en parallèle avec la plage de mesure et réglée en choisissant une valeur de la résistance sur son entrée nommée R_e. La résistance doit être choisie de sorte que le courant égale à 10 mA au maximum pour une tension nominale.

Le courant est alors converti avec un rapport de conversion de 2500:1000, un courant 10mA dans le terminal correspond à un courant 25mA sur la borne de sortie [**88**].

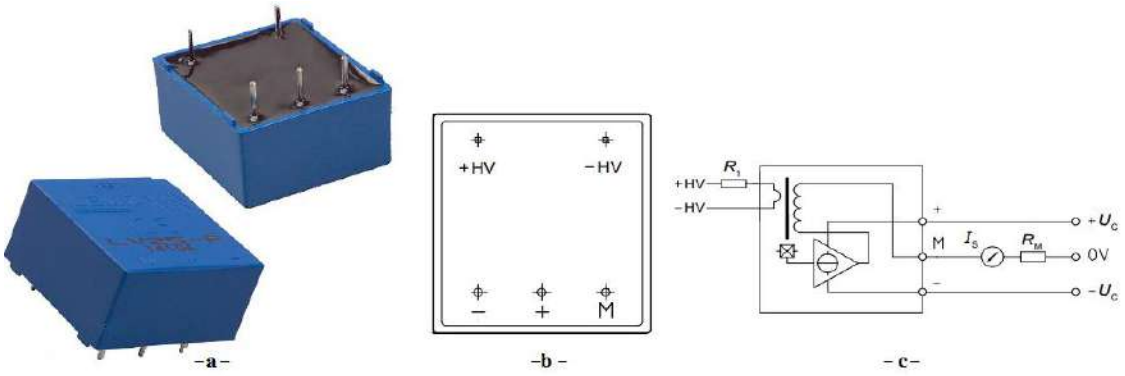


Figure 2.5 Capteur LEM LV 25-P: -a- la conception du capteur, -b- le brochage du capteur, -c- Schéma de brochagede du capteur [**88**].

En outre, une autre résistance nommée R_s doit être choisie, telle que la tension induite dans le circuit secondaire ne dépasse pas le maximum 10 V, la tension autorisée sur les entrées de carte d'acquisition.

2.3.1.3 Mesure de l'ensoleillement sur le plan incliné

Pour la mesure de l'irradiation inclinée le capteur SPEKTRON 200 (figure 2.6) a été utilisé [89]. Le capteur d'irradiation SPEKTRON 200 offre les possibilités d'évaluer l'irradiation entre 0 et 1500 W/m ² et une tension proportionnelle à l'intensité de l'irradiation solaire. En outre, la tension mesurée peut être convertie en unité d'irradiation (W/m ²), en utilisant la valeur d'étalonnage imprimée sur le capteur.

Le capteur SPEKTRON 200 est fixé sur le support de la grille de montage du système PV. Il faut veiller à ce que le capteur ait la même inclinaison et l'orientation que le système PV à examiner. Notons que des écarts mineurs peuvent entraîner des erreurs de mesure.



Figure 2.6 Capteur de rayonnement SPEKTRON 200 [89].

Un étage d'amplification a été ajouté après le capteur SPEKTRON 200 car le signal généré par ce dernier est très faible, donc pour un meilleur signal d'acquisition il est préférable d'ajouter ce circuit.

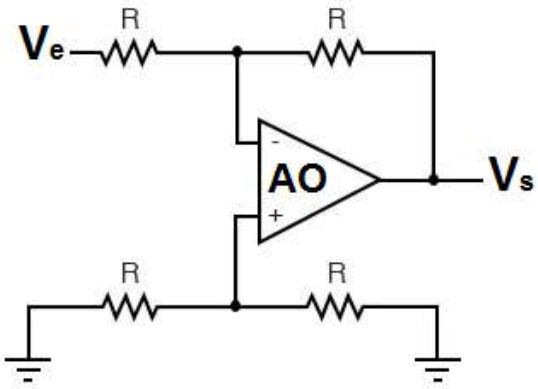


Figure 2.7 Circuit de l'amplificateur différentiel.

Afin d'amplifier la tension mesurée fournie par les capteurs, un circuit amplificateur différentiel classique est ajouté à la sortie de chaque capteur. Ce circuit est constitué d''amplificateur opérationnel et quatre résistances. La figure (2.7) montre le circuit utilisé pour amplifier le signal mesuré.

2.3.1.4 Mesure de la température ambiante

La mesure de la température ambiante peut être effectuée avec plusieurs types de capteurs analogiques ou numériques selon la nature des entrées de la carte d'acquisition. Dans notre réalisation, on a opté pour le semi-conducteur LM35 [**90**].

Ce capteur est précis et peut être facilement calibré, ainsi qu'il est capable de mesurer des températures allant de -55°C à +150°C dans sa version la plus précise et avec le montage adéquat. La sortie analogique du capteur est proportionnelle à la température. Il suffit de mesurer la tension en sortie du capteur pour en déduire la température. Chaque degré Celsius correspond à une tension de +10mV. Le branchement du ce capteur est indiqué dans la figure 2.8.



Figure 2.8 Capteur de température LM35 [**90**].

2.3.2 Circuit d'alimentation

Afin de fournir la tension d'alimentation pour les capteurs utilisés, un circuit d'alimentation a été réalisé .Il est composé d'un transformateur abaisseur 220/24V, d'un redresseur et de régulateurs. Cette alimentation fournit ±12 volts pour le capteur de tension LEM LV 25-P et aussi pour la polarisation de l'amplificateur opérationnel utilisé dans les circuits de conditionnement du signal. La tension de 5 volts est également fournie pour l'alimentation du capteur de courant ACS712 et pour le capteur de température LM35. La figure 2.9 montre le circuit d'alimentation réalisé. La carte de mesure du courant, de la tension, de la température et du circuit d'alimentation est indiquée sur la figure 2.10.

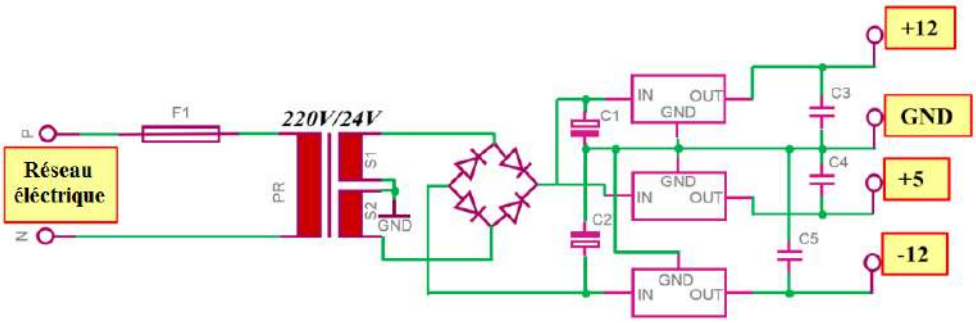


Figure 2.9 Conception du circuit d'alimentation.

2.3.3 Appareil d'acquisition de données

L'acquisition des données a été réalisée à l'aide du multiplexeur NI-USB 6008 fabriqué par National Instruments. Ce module contient huit canaux de signaux asymétriques et quatre entrées différentielles dans lesquelles l'interface avec le PC a été réalisée à l'aide du langage LabVIEW via port USB.

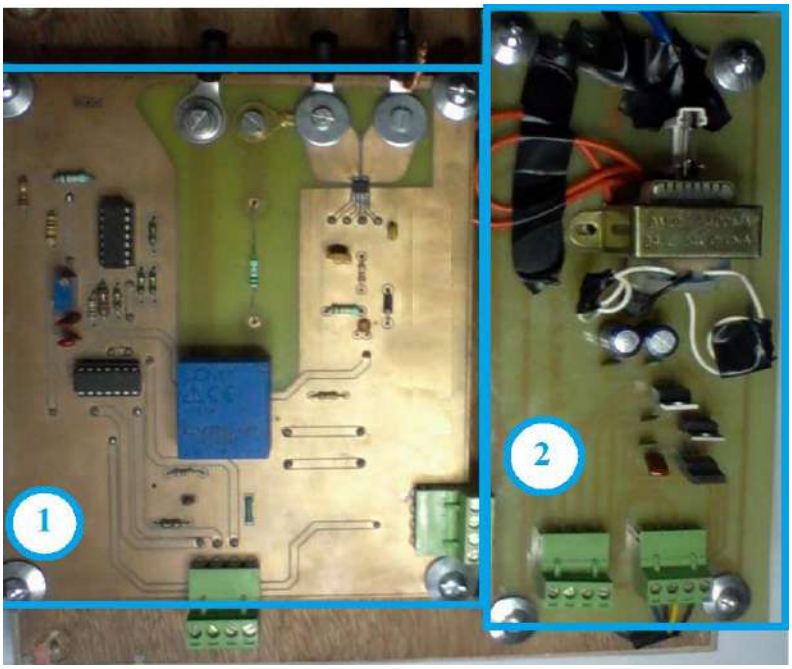


Figure 2.10 (1) Carte de mesure, et (2) carte d'alimentation.

La figure 2.11 montre le multiplexeur NI-USB 6008 utilisé pour l'acquisition de données. Pour plus de détails sur cette carte, la fiche technique est fournie dans [**91**].



Figure 2.11 Multiplexeur d'acquisition de données NI-USB 6008 [**91**].

2.4 Récupération de données et calibration sous LabVIEW

L'acquisition du logiciel a été réalisée avec le langage LabVIEW à l'aide de l'assistant DAQ. Cet outil est adaptable avec le multiplexeur NI-USB 6008 pour fournir un signal analogique fourni par l'entrée de ce multiplexeur sous forme de mesures. Dans la fenêtre de l'assistant DAQ illustrée à la figure 2.12, le type de lecture et l'entrée utilisée sont sélectionnés. Une fois que les signaux de mesures sont prêts à la lecture dans le langage LabVIEW, l'étalonnage est effectué pour que les mesures soient affichées à l'aide d'un simple cadre de calcul. La capture d'écran de la figure 2.13 montre le programme VI de l'acquisition de données.

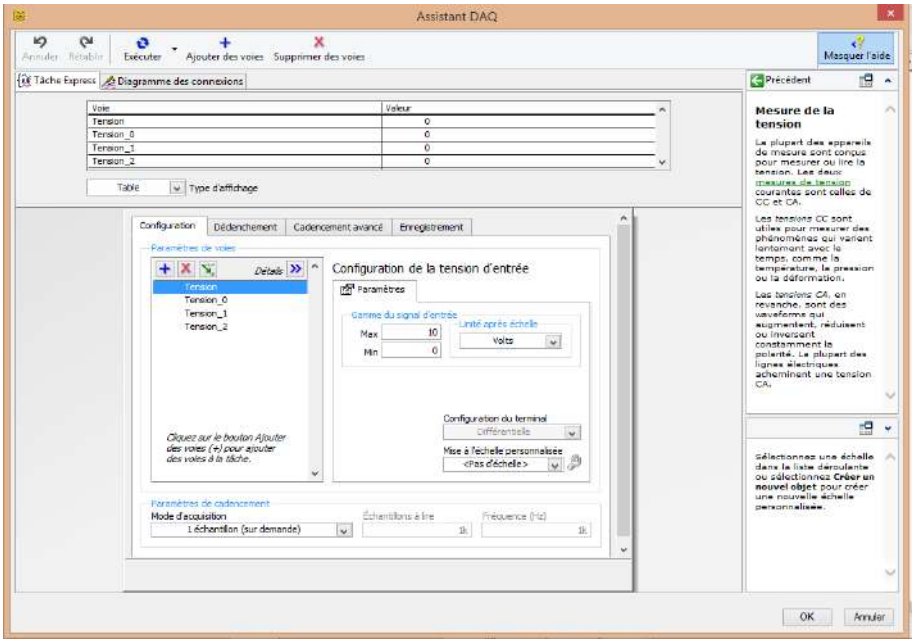


Figure 2.12 La fenêtre assistant DAQ.

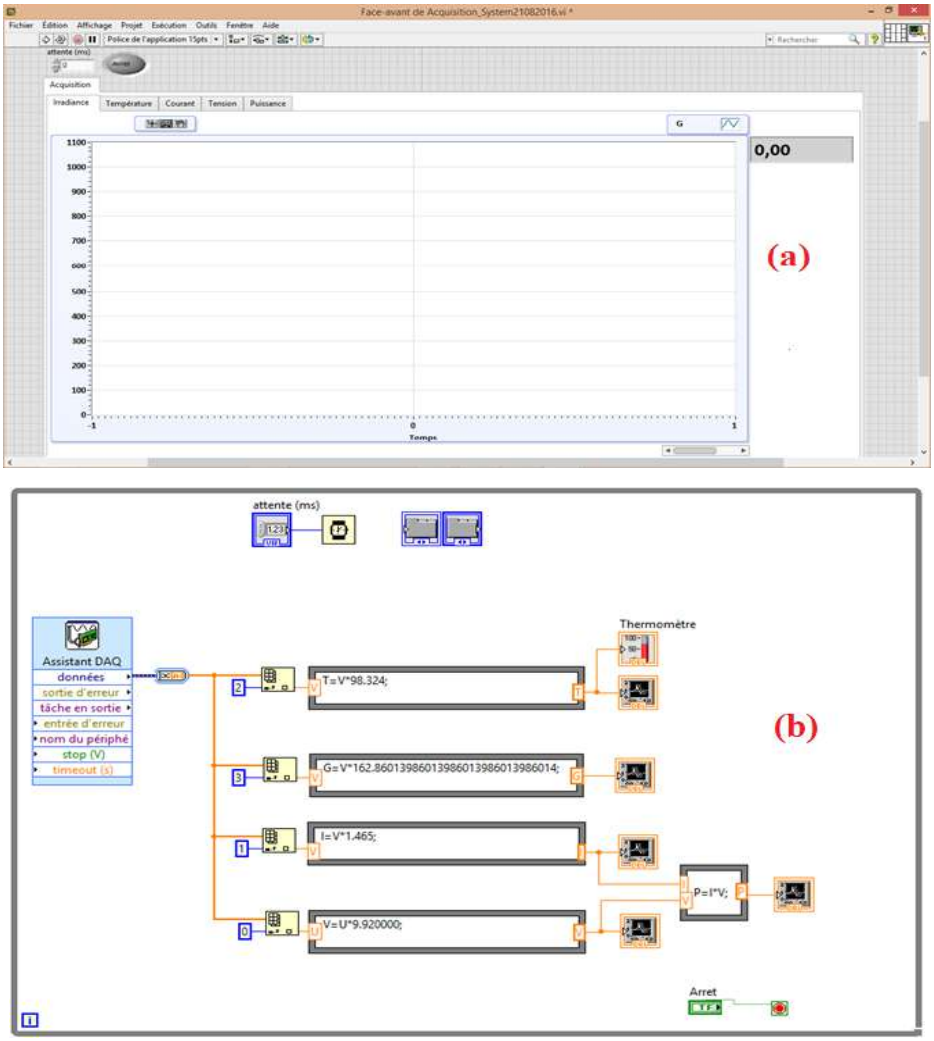


Figure 2.13 (a) Face avant du programme d'acquisition LabVIEW **(b)** Diagramme LabVIEW VI pour l'acquisition.

Le tableau 2.3 résume toutes les caractéristiques des capteurs avec le facteur de calibration en LabVIEW.

Tableau 2.3 La description des capteurs avec leurs valeurs de calibration.

Paramètre mesuré		Symbole	Unité	Type de capteur	Référence de capteur	Précision	La valeur de calibration
Electrique	Courant DC	I_{dc}	A	Effet de Hall	Allegro ACS712	±0.9 %	1.453
	Tension DC	V_{dc}	V		LEM LV 25-P	1.5%	9.8
Météorologique	Température ambiante	T	°C	Température du semi-conducteur	LM35	0.5°C	4.95024
	Irradiation globale inclinée	G	W/m²	Cellule de référence PV	SPEKTRON 200	±5 %	139.86

2.5 Résultats d'acquisition de données

Les figures 2.14 à 2.18 montrent les résultats fournis par le système d'acquisition pour SPVCR qui sont respectivement l'irradiation, la température ambiante, le courant DC, la tension DC et la puissance DC.

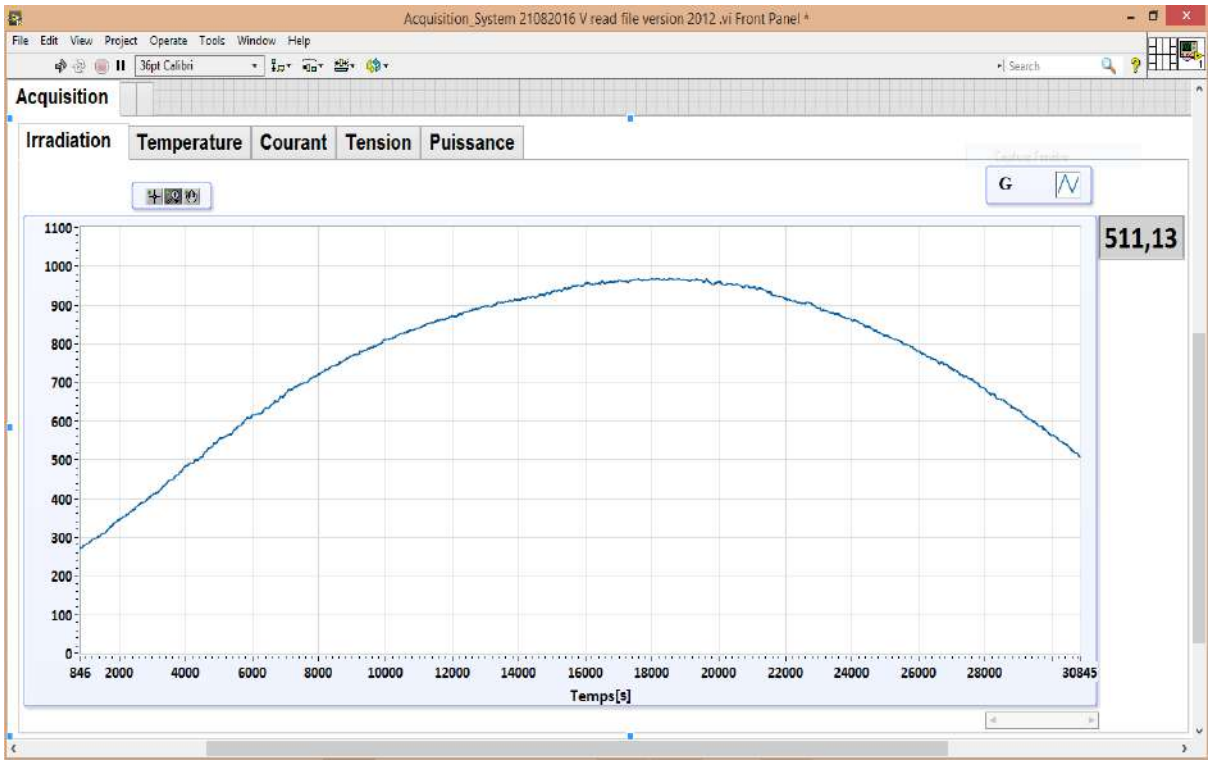


Figure 2.14 L'irradiation mesurée.

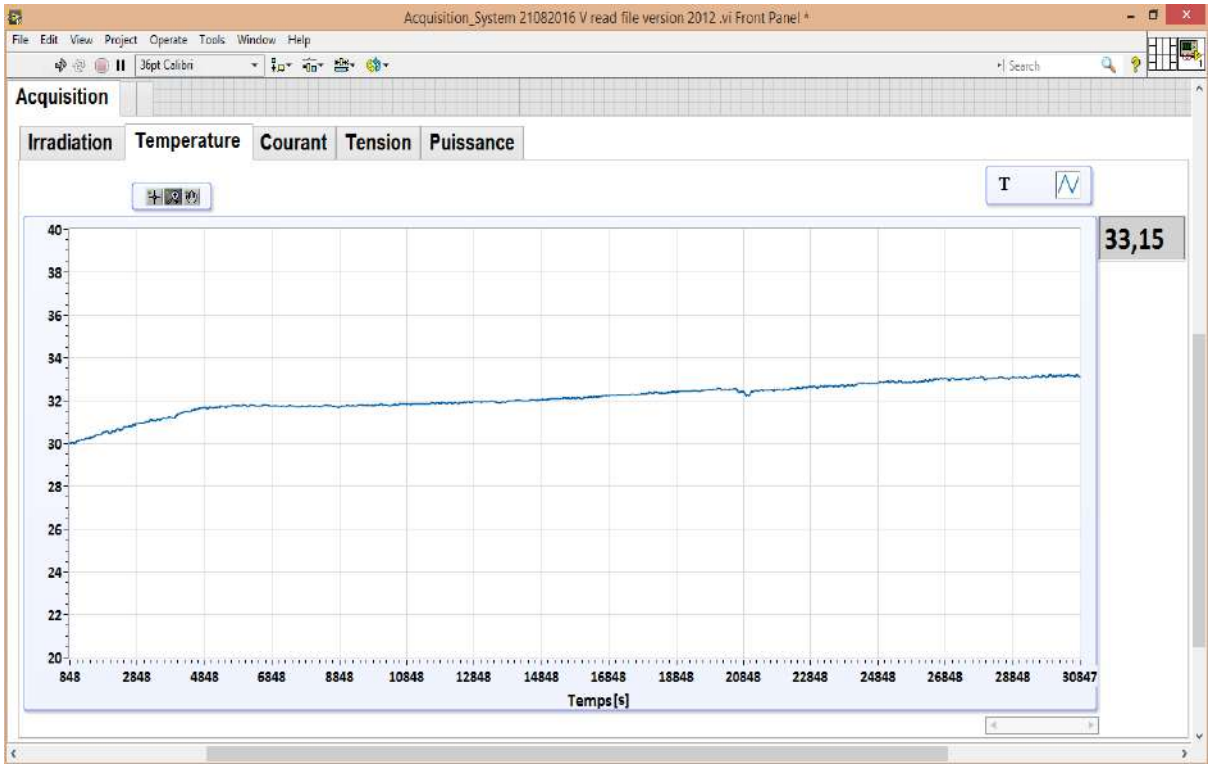


Figure 2.15 La température ambiante mesurée.

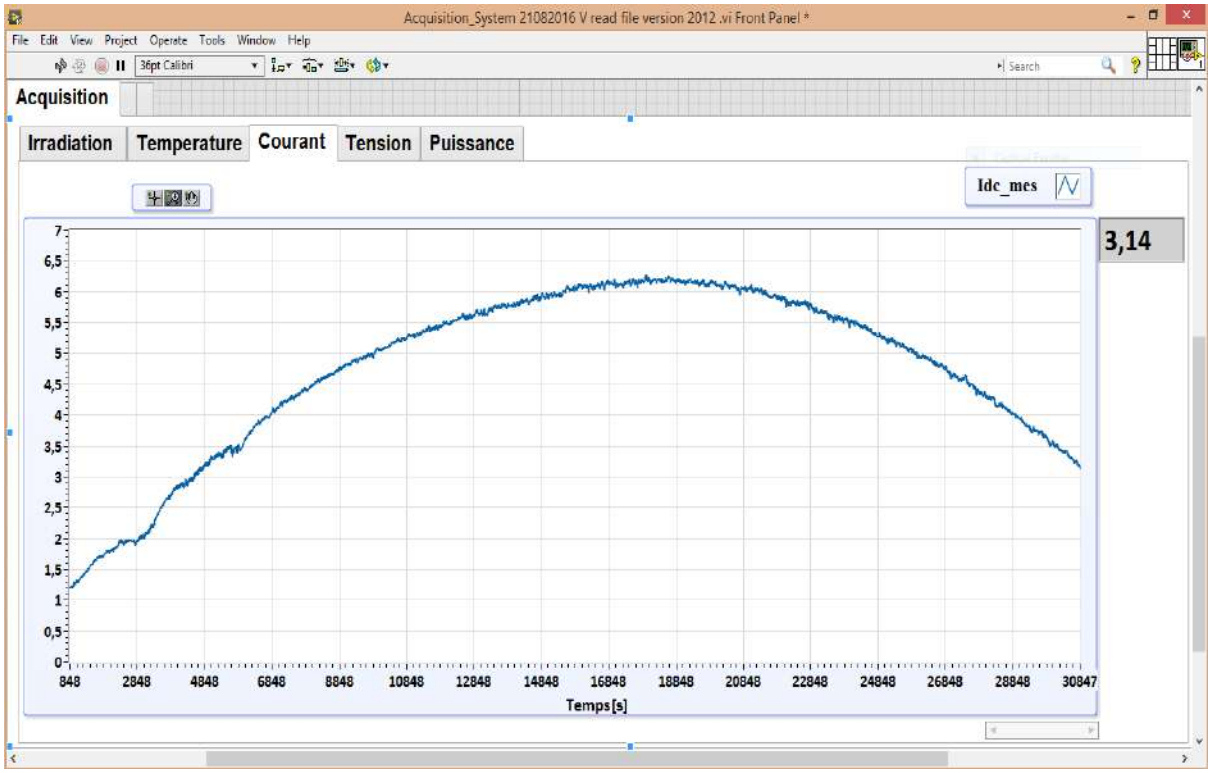


Figure 2.16 Le courant DC mesuré.

2.6 Conclusion

Dans ce chapitre, une brève description du système PV considéré dans le cadre de cette thèse connecté au réseau est présentée. Ce système nécessite un système de surveillance et d'acquisition de données. Le système d'acquisition intègre une carte d'acquisition qui contient quatre capteurs. Ces capteurs permettent de mesurer quatre grandeurs, le courant, la tension, la température ambiante et l'irradiation.

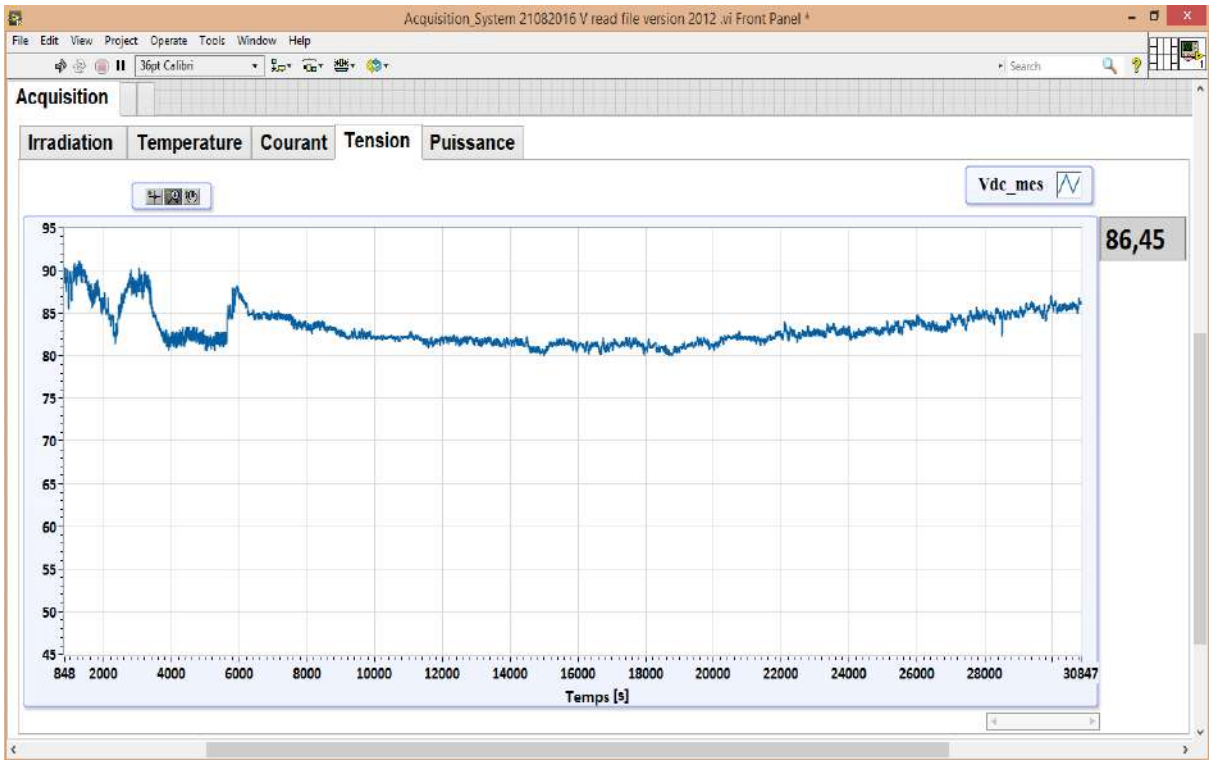


Figure 2.17 La tension DC mesurée.

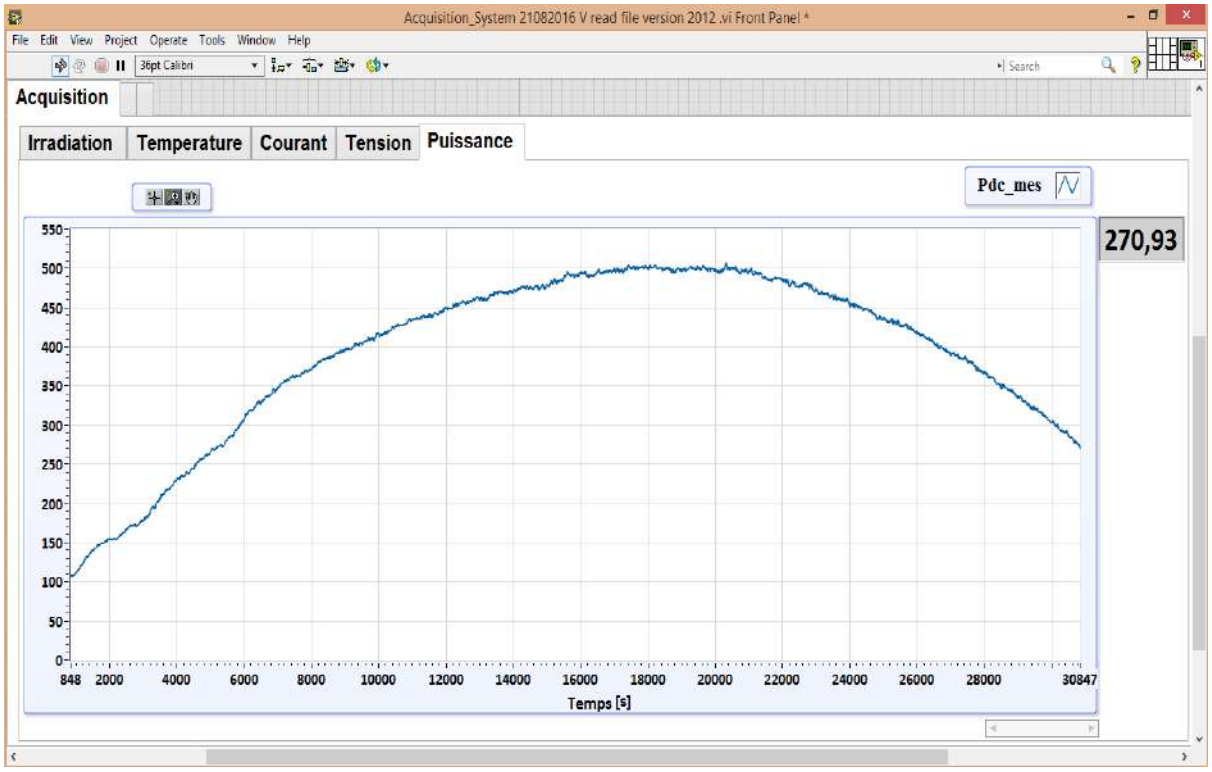


Figure 2.18 La puissance DC mesurée.

CHAPITRE 3

MODELISATION D'UN SYSTEME PV CONNECTE AU RESEAU

Chapitre 3

Modélisation de Système PV Connecté au Réseau

3.1 Introduction

Plusieurs modèles de circuits électriques équivalents ont été proposés dans la littérature pour décrire le comportement de la cellule PV [92]. Les modèles les plus utilisés sont les modèles à une seule diode (MSD) et à double diodes (MDD). Une description plus détaillée de ces modèles est donnée dans la section suivante.

3.2 Modélisation de module PV

Dans cette section, les modèles équivalents de cellules PV utilisés sont décrits en détail et qui sont: le MSD et le MDD. Ces modèles possèdent respectivement cinq et sept variables inconnues.

3.2.1 Modèle à une Seule Diode (MSD)

Comme la figure 3.1 ci-dessus le montre, le modèle à une seule diode est composé d'une

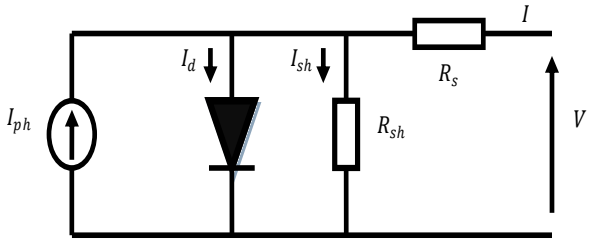


Figure 3.1 Circuit électrique équivalent d'un modèle à une seule diode pour une cellule PV.

source de courant qui représente proportionnellement l'intensité de courant générée par le rayonnement incident. Cette source de courant est en parallèle avec une diode et deux résistances série et parallèle R_s et R_{sh} respectivement. Ces deux résistances représentent les phénomènes de dissipation dans la cellule [93].

Selon le modèle électrique donné précédemment, la cellule PV possède une caractéristique I-V non linéaire qui représente la variation du courant fourni par la cellule en fonction de la tension telle qu'elle est décrite dans l'équation (3.4). Dans cette équation,

I est le courant généré par la cellule solaire ; I_{ph}, I_d et I_{sh} sont respectivement le courant photo-généré, le courant de la diode et le courant de court-circuit (shunt). K est la constante de Boltzmann, q est la charge électrique, T_c est la température de la cellule, n est le facteur d'idéalité de la diode; I_0 est le courant de saturation dû à la diffusion et à la recombinaison.

$$I = I_{ph} - I_d - I_{sh} \tag{3.1}$$

$$I_d = I_0 \left[\exp\left(\frac{q(V + R_s I)}{nKT_c} \right) - 1 \right] \tag{3.2}$$

$$I_{sh} = \frac{V + R_s I}{R_{sh}} \tag{3.3}$$

On remplace les équations (3.2) et (3.3) dans l'équation (3.1) ce qui donne l'équation (3.4) du courant de sortie de la cellule.

$$I = I_{ph} - I_0 \left[\exp\left(\frac{q(V + R_s I)}{nKT_c} \right) - 1 \right] - \frac{V + R_s I}{R_{sh}} \tag{3.4}$$

I_{ph}, I_0, R_S, R_{sh} et n sont les paramètres inconnus (à identifier).

Le modèle à une seule diode ignore l'effet de recombinaison de la diode. Alors, il est efficace pour la simulation mais pas le modèle le plus précis. C'est pourquoi on préfère utiliser le modèle à double diode pour l'étude des caractéristiques des cellules solaires à des points critiques.

3.2.2 Modèle à double diode (MDD)

Le modèle à double diode est dérivé du même circuit équivalent du modèle à une seule diode avec une principale différence que le courant de recombinaison est divisé en I_{01} et I_{02} qui représentent la recombinaison en surface et en masse.

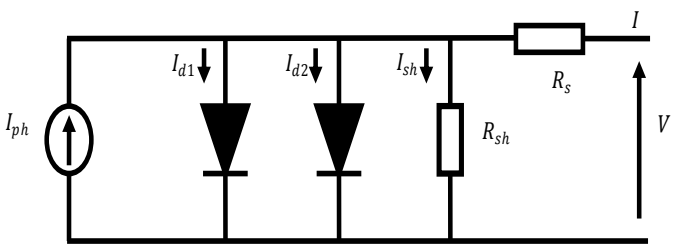


Figure 3.2 Circuit électrique équivalent du modèle à double diode pour une cellule PV.

Ceci peut être ajouté au sous-circuit de la cellule solaire en ajoutant simplement une deuxième diode, comme le montre la figure 3.2 avec n = 2 [**93**].

I_{01} et I_{02} sont les courants de saturation; n_1 et n_2 sont les facteurs d'idéalité de la première et la deuxième diode respectivement. Les sept paramètres inconnus sont I_{ph}, I_{01}, I_{02}, n_1, n_2, R_s et R_{sh}

$$I = I_{ph} - I_{01}\left[\exp\left(\frac{q(V + R_s I)}{n_1 K T_c}\right) - 1\right] - I_{02}\left[\exp\left(\frac{q(V + R_s I)}{n_2 K T_c}\right) - 1\right] - \frac{V + R_s I}{R_{sh}} \qquad (3.5)$$

3.2.3 Fonction de coût

Le but du processus d'optimisation est de trouver les meilleures valeurs pour les paramètres qui donne une erreur minimale entre le modèle avec les paramètres extraits et les données mesurées, pour deux fonctions de coût, la première est pour le MSD et la seconde est pour le MDD. En outre, l'erreur utilisée dans notre cas est l'erreur quadratique moyenne (RMSE) donnée par l'expression (3.6) représentant la fonction objective. L'erreur entre les données mesurées et le modèle MSD est calculée à partir de l'équation (3.7) et l'erreur entre les données mesurées et le modèle MDD est calculée à partir de l'équation (3.8).

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k} f\left(I_m, V_m, X\right)^2} \qquad (3.6)$$

K est le nombre de données I-V mesurées ; chaque solution est représentée par le vecteur X, pour le modèle à une seule diode X = [I_{ph}, I_0, n, R_s, R_{sh}] et pour le modèle à double diode X = [I_{ph}, I_{01}, I_{02}, n_1, n_2, R_s, R_{sh}] tel que défini dans les équations (3.7), (3.8).

$$f\left(I_m, V_m, X\right) = I_{ph} - I_0\left[\exp\left(\frac{q(V_m + R_s I_m)}{nKT_c}\right) - 1\right] - \frac{V_m + R_s I_m}{R_{sh}} - I_m \qquad (3.7)$$

$$f\left(I_m, V_m, X\right) = I_{ph} - I_{01}\left[\exp\left(\frac{q(V_m + R_s I_m)}{n_1 K T_c}\right) - 1\right] - I_{02}\left[\exp\left(\frac{q(V_m + R_s I_m)}{n_2 K T_c}\right) - 1\right] - \frac{V_m + R_s I_m}{R_{sh}} - I_m \qquad (3.8)$$

3.3. Algorithme de Pollinisation des Fleurs (FPA)

La plupart des plantes sont des plantes à fleurs et il y a plus de 250 000 espèces de plantes à fleurs dans le monde, où la pollinisation représente la stratégie principale de reproduction des plantes [**94, 95**]. La pollinisation est un processus de transfert de pollen d'une fleur à une autre par le vent ou par les pollinisateurs tels que les insectes, les papillons, les abeilles, les oiseaux et les chauves-souris.

Les plantes à fleurs ont les caractéristiques de produire le nectar et d'attirer les pollinisateurs ainsi pour assurer la pollinisation [**96**]. En outre, certains pollinisateurs et les espèces végétales telles que les colibris et les plantes à fleurs ornithophiles forment une constance de fleur co-évolutionnaire [**94,95**]. Basé sur les caractéristiques principales de la pollinisation, l'algorithme de la pollinisation des fleurs a été développé [**97**].

3.3.1 FPA dans le contexte de l'optimisation : source inspirée de la nature

Avant de décrire l'algorithme de la pollinisation des fleurs en détail, on passe brièvement en revue la forme de base de la pollinisation chez les angiospermes. Le processus de la pollinisation possède deux formes de base : biotiques ou abiotiques.

i. Pollinisation biotique

Parmi les principales formes de pollinisation, on trouve la pollinisation biotique, également appelée pollinisation croisée, qui est due aux pollinisateurs tels que les insectes, les oiseaux ou d'autres visiteurs de la fleur. Près de 90 % des angiospermes utilisent cette forme de pollinisation. Comme pollinisateurs, ils se déplacent et même volent avec différentes allures et vitesses. Le mouvement du pollen peut être à une distance assez longue. Telle pollinisation peut également être considérée comme pollinisation globale avec des propriétés potentielles de vols de Lévy [**97,98**]. Si le pollen est codé comme un vecteur de la solution, cette action peut être équivalente à la recherche globale.

ii. Pollinisation abiotique

La pollinisation abiotique, également appelée autopollinisation, ne nécessite pas de pollinisateurs. On estime qu'environ 10 % des plantes à fleurs prennent cette forme de la pollinisation. La pollinisation a tendance à être une autopollinisation ou une pollinisation locale qui prend lieu par le vent, la gravité ou parfois par la pluie [**96,97**]. La distance

parcourue par ce mouvement local est généralement courte, et cette action peut donc être considérée comme une recherche locale.

iii. Constance de la fleur

Parfois, il est avantageux pour les plantes et les pollinisateurs comme les colibris de former un partenariat afin d'économiser de l'énergie avec succès garanti. Par conséquent, la constance de la fleur a évoluée. Dans ce cas, les pollinisateurs seulement visitent un ensemble fixe de types de fleurs sans perte d'énergie pour explorer de nouveaux types de fleurs, tandis que les plantes fleuries évoluent pour fournir comme récompense suffisamment de nectar aux pollinisateurs afin d'encourager les visites fréquentes des pollinisateurs et donc favoriser leur succès de la reproduction [**94, 96**].

Les caractéristiques ci-dessus ont été utilisées pour concevoir un algorithme d'optimisation, appelé algorithme de pollinisation des fleurs (FPA) [**97**]. Les principales caractéristiques et les composantes de l'algorithme du FPA peuvent sont résumés dans le tableau 3.1, qui montrent la relation ou l'équivalence entre les termes de l'optimisation et le contexte de la fleur.

Tableau 3.1 Pollinisation et ses composants d'optimisation.

Pollinisation des fleurs	**Composants d'optimisation (en FPA)**
Pollinisateurs (Insectes, Oiseaux, …)	Mouvements / modification de variables
Biotique	Recherche globale
Abiotique	Recherche locale
Vol de Lévy	Tailles de pas (obéissant à une loi de puissance)
Pollen / fleurs	Vecteurs de solution
Constance de la fleur	Similarité dans les vecteurs de solution
Evolution des fleurs	Evolution itérative des solutions
Reproduction florale optimale	Ensemble de solutions optimales

Avec ces éléments et caractéristiques, nous pouvons maintenant décrire l'algorithme de pollinisation des fleurs standard en détail.

3.3.2 Les lois fondamentales du FPA

Le FPA est un algorithme inspiré de la nature qui imite le comportement principal de la pollinisation des plantes à fleurs. Quatre règles fondamentales étaient utilisées par Yang en 2012 pour le développement de l'algorithme FPA [**97,98**]. Elles peuvent être résumées comme suit :

Règle 1 : La pollinisation globale implique biotiques et pollinisation croisée où les pollinisateurs transportent le pollen issu des vols de Lévy.

Règle 2 : L'autopollinisation et l'autofécondation peuvent être supposées comme une pollinisation locale.

Règle 3 : Constance de fleur peut être considérée comme une probabilité de reproduction qui est proportionnelle à la similitude entre les deux fleurs.

Règle 4 : Probabilité de commutation p ∈ [0,1] peut contrôler l'échange entre la pollinisation locale et la pollinisation globale. La pollinisation locale a un rapport **p** important dans l'ensemble des activités du processus de la pollinisation.

Pour illustrer le mécanisme de l'accord de partenariat basé sur les quatre règles mentionnées précédemment, les trois étapes principales sont décrites dans les trois paragraphes suivants.

3.3.2.1 Recherche globale de FPA (Biotique)

Comme mentionné ci-dessus, les pollinisateurs tels que les oiseaux et les chauves-souris peuvent transférer le pollen sur de longues distances pendant la pollinisation biotique, assurant ainsi la diversité et la pollinisation la plus adaptée à la reproduction. Par conséquent, la première (**règle 1**) et la troisième (**règle 3**) de l'algorithme FPA peuvent être formulées mathématiquement comme suit:

$$x_i^{t+1} = x_i^t + L\left(x_i^t - g_*\right) \tag{3.9}$$

Où x_i^t le vecteur de pollen ou de solution à l'itération t et g∗ est la meilleure solution trouvée parmi toutes les solutions à l'itération actuelle. Le paramètre L est la force de la pollinisation, qui est essentiellement une taille de pas. Parce que les pollinisateurs se déplacent sur de longues distances avec divers intervalles de distance (différentes

longueurs de pas), le vol de Lévy peut être un simulateur efficace pour cette caractéristique [**97,98**].

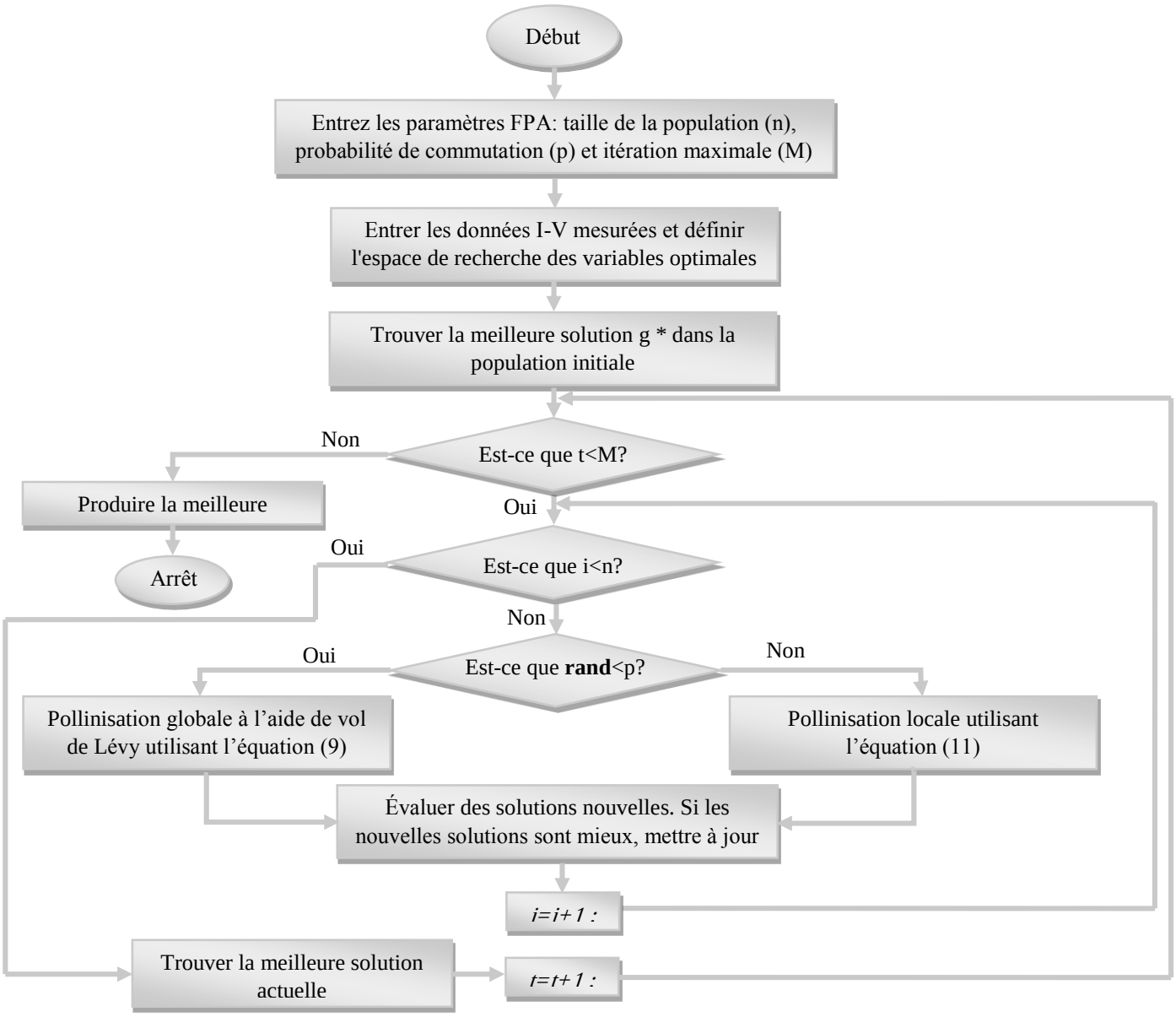


Figure 3.3 L'organigramme de l'algorithme FPA.

Autrement dit, *L* peut être tiré d'une distribution de Lévy comme définie par l'équation (3.10) :

$$L \sim \frac{\lambda \Gamma(\lambda) \sin\left(\frac{\pi\lambda}{2}\right)}{\pi} \frac{1}{s^{1+\lambda}} \left(s \gg s_0 > 0\right) \qquad (3.10)$$

Où Γ (λ) désigne la fonction gamma standard et cette distribution est valable pour les grands pas s > 0. Il est recommandé d'utiliser λ = 1,5 [**97,98**].

3.3.2.2 Recherche locale de FPA (Abiotique)

La pollinisation abiotique se produit par le vent, la gravité, l'eau ou par diffusion sans l'aide des pollinisateurs. La pollinisation locale (**règle 2**) et la constance des fleurs (**règle 3**) peuvent être représentées comme suit:

$$x_i^{t+1} = x_i^t + \in \left(x_j^t - x_k^t \right) \tag{3.11}$$

Où x_i^t et x_k^t sont le pollen des différentes fleurs du même type de plante. Cette équation imite essentiellement la constance de fleur dans un voisinage limité. Mathématiquement, si x_i^t et x_k^tsont de la même espèce qui peut être sélectionnée de la même population, l'équation devient une marche aléatoire locale si nous tirons ε à partir d'une distribution uniforme en [0, 1], et le nouveau vecteur de solution généré ne sera pas trop éloigné des solutions existantes.

3.3.2.3 Probabilité de commutation en FPA

Bien que nous ayons simulé la pollinisation biotique et abiotique, nous n'avons pas pris en compte le pourcentage et la fréquence de chaque type de pollinisation. Pour imiter cette fonctionnalité, nous utilisons une probabilité de commutation (**règle 4**), où la valeur de *p* détermine la modification de la solution qui suit la pollinisation locale ou globale. Bien qu'une valeur naïve de p = 0,5 puisse être utilisée, une valeur plus réaliste et efficace de p = 0,7 donne de meilleures performances pour la plupart des applications [**98**].

La figure 3.3 montre l'organigramme du FPA. Trois étapes clés peuvent être résumées dans le pseudocode FPA présenté dans l'algorithme 1.

1 : **Entrées** : Les agents de fleur Totale N, Nombre d'itérations Maximal M,
Probabilité de commutation P

2 : **Sorties** : la meilleure solution (g∗) et sa valeur de remise en forme.

3 : Initialiser la population de fleurs N au hasard.

4 : Choisir la meilleure solution (g∗).

5 : **Tant que** les critères d'arrêts ne répondent pas **Faire ou Refaire**

6 : **Pour toutes** les fleurs dans l'ensemble de la solution **Faire ou Refaire**

7 : **Si** rand < p

8 : *Concevoir* le vecteur de dimensions L basé sur la distribution de Lèvy.

9 : *Employer* la pollinisation globale sur j solution dans l'équation (3.9).

10 : **Sinon**

11 : *Tirer* ε de la distribution uniforme.

12 : *Sélectionner* les solutions j et k au hasard.

13 : *Exécuter* la pollinisation locale sur la solution i en utilisant les solutions j et k comme dans l'équation (3.11).

14 : **Fin Si**

15 : *Evaluer* la nouvelle solution.

16 : **Si** la nouvelle solution est meilleure que celle en cours

17 : *Remplacer* la solution actuelle **j** par la nouvelle solution.

18 : **Fin Si**

19 : **Fin pour**

20 : *Mise à niveau* de la g∗ de la solution optimale.

21 : **Fin tant que**

22 : *Choisir* la solution optimale.

Dans ce cadre, le FPA est appliqué pour obtenir l'erreur quadratique moyenne la plus basse entre le réel et les courbes I-V estimées. Ces courbes estimées sont obtenues à partir de l'équation (3.4) dans le cas d'un modèle à une seule diode avec cinq paramètres inconnus et de l'équation (3.5) dans le cas d'un modèle à double diode avec sept paramètres inconnus. Le tableau 3.2 ci-dessous montre les valeurs des paramètres réglables pour exécuter le processus d'optimisation par le FPA.

Tableau 3.2 Paramètres de l'algorithme FPA

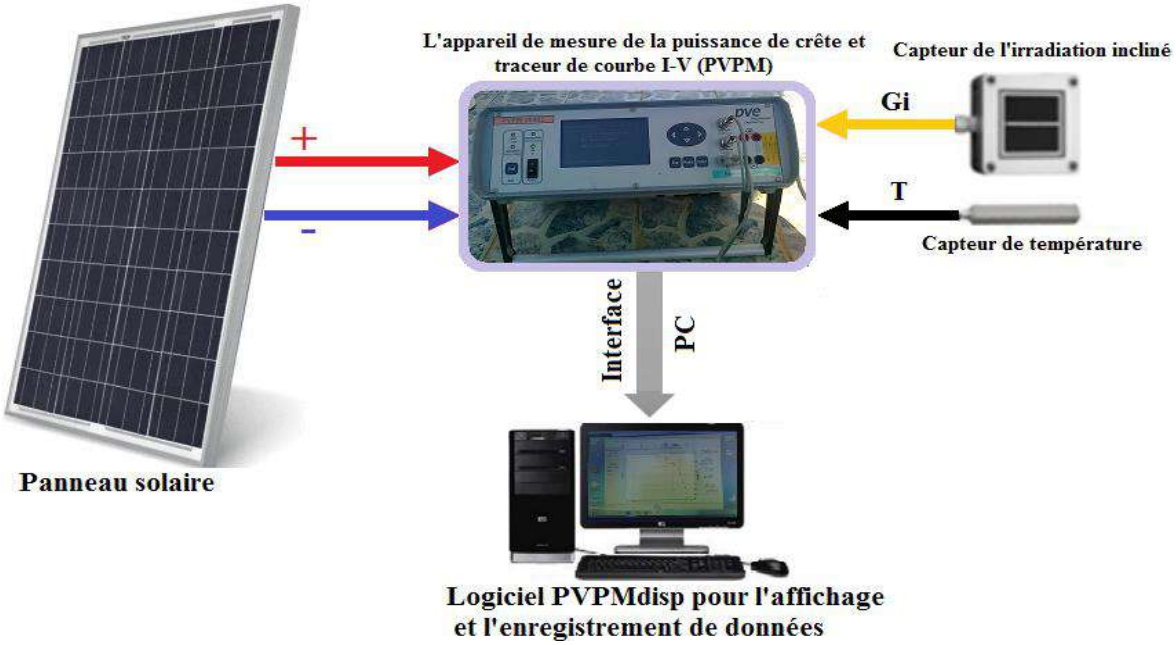
Paramètres de l'algorithme FPA	**Valeur**
Taille de la population (n)	25
Probabilité de commutation (p)	0.7
Maximum nombre d'itération (M)	1000

3.4. Mesure de caractéristique I-V

Un système d'acquisition est utilisé afin d'obtenir les données des caractéristiques I-V mesurées à l'extérieure. Ce système est installé au niveau du Centre de Développement des Energies Renouvelables (CDER). En utilisant sur l'instrument PVPM, une courbe I-V est établie pour un ou plusieurs modules PV (chaine PV) déjà choisis. L'instrument est mis en place sur le lieu de montage du module PV [**99**].

Chaque courbe I-V contient 101 échantillons sur une période de 30 secondes. Ces données permettent un véritable traçage de courbe I-V. Le schéma synoptique du système d'acquisition est représenté sur la figure 3.4.

Le courant, la tension, la température et l'irradiation sont mesurés et enregistrés à partir d'un ensemble de mesures expérimentales à l'extérieur. Les données sont obtenues à partir de quatre modules PV de différents types de technologies en silicium, Monocristallin (Mono-Si) et Polycristallin (Poly-Si). Chaque module PV possède ses propres paramètres. Le tableau 3.3 représente les paramètres électriques fournis par les fabricants de modules PV. Ces modules sont : ISOFOTON 106, SANYO 180, CONDOR, 250 et KYOCERA 125.



Figures 3.4 Schéma synoptique pour l'acquisition de données des caractéristiques I-V.

Le schéma synoptique d'identification des paramètres pour les deux modèles proposés résumé par la figure 3.5. L'espace de recherche des paramètres est parmi les conditions nécessaires de l'exécution du processus d'optimisation. Le tableau 3.4 montre les valeurs de limites inférieures et supérieures de l'espace de recherche. Les valeurs des meilleurs paramètres optimaux obtenus doivent appartenir à ce domaine.

Tableau 3.3 Les paramètres électriques aux conditions d'essai normalisées (STC) pour les modules PV utilisés.

Module PV / Paramètres	ISOFOTON 106	SANYO 180	CONDOR 250P	KYOCERA 125
$V_{oc}(V)$	17,4	66,4	37,24	21,7
$I_{sc}(A)$	6,54	3,65	9,01	8
$V_{mp}(V)$	21,6	54	31,19	17,4
$I_{mp}(A)$	6,10	3,33	8,37	7,2
$P_{mp}(W)$	106	180	250	125
$\beta V_{oc}(\%/°C)$	-0,36	-0.33	-0,36	XX
$\alpha I_{sc}(\%/°C)$	0,06	0.01	0,06	XX

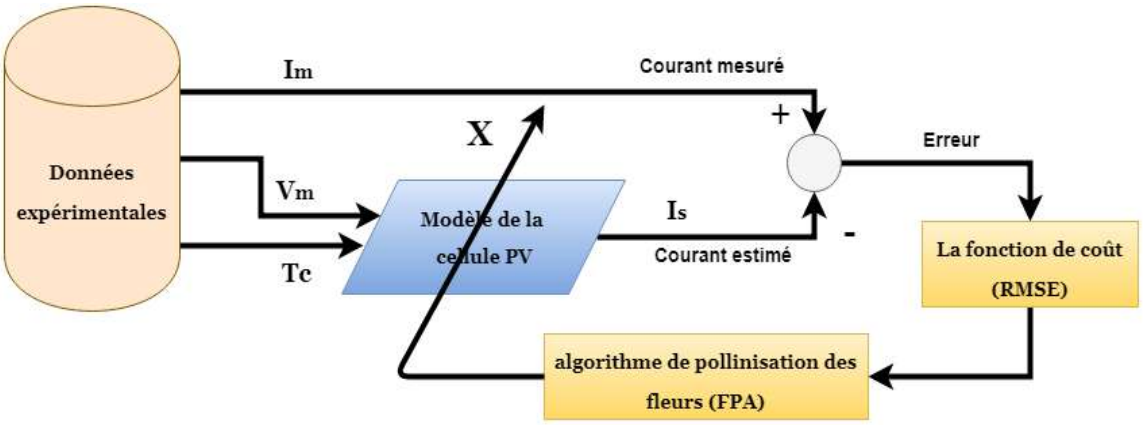


Figure 3.5 Le schéma synoptique du processus d'optimisation.

Le tableau 3.4 contient les valeurs de l'espace de recherche pour chaque paramètre du module ISOFOTON.

Tableau 3.4 L'espace de recherche des paramètres.

Paramètre	Limite inférieure	Limite supérieure
$\boldsymbol{I(A)}$	0	20
$\boldsymbol{R_s(\Omega)}$	0	0.5
$\boldsymbol{R_{sh}(\Omega)}$	0	400
$\boldsymbol{I_{01}(\mu A)}$	0	10^{-2}
$\boldsymbol{I_{02}(\mu A)}$	0	10^{-2}
$\boldsymbol{n1}$	**0**	**120**
$\boldsymbol{n2}$	0	120

3.5 Simulation et résultat expérimental

Dans cette partie, les paramètres seront identifiés pour les deux modèles électriques MSD et MDD; le processus d'identification est basé sur l'algorithme FPA utilisant les données de mesure I-V pour les modules PV déjà choisis (tableau 3.3). Les données collectées contiennent deux types de cellules PV. L'étude fournit un test de FPA pour plusieurs fabricants de modules PV dans diverses conditions météorologiques. En outre, une comparaison avec deux autres algorithmes, à savoir: algorithme de luciole (**FA**) [**98,100**] et recuit simulé (**SA**) [**101,102**] est effectué.

L'algorithme de luciole (**Firefly Algorithm**) est un algorithme méta heuristique basé sur un essaim qui a été introduit par Yang [**98,100**]. L'algorithme imite la façon dont les lucioles interagissent à l'aide de leurs lumières clignotantes. L'algorithme suppose que toutes les lucioles sont unisexes, ce qui signifie que n'importe quelle luciole peut être attirée par n'importe quelle autre luciole ; l'attractivité d'une luciole est directement proportionnelle à sa luminosité qui dépend de la fonction d'objectif. Chaque luciole sera attirée par une autre luciole plus brillante. De plus, la luminosité diminue en fonction de la distance et en fonction de la loi du carré inverse.

L'algorithme de recuit simulé (**Simulated Annealing Algorithm**) est un procédé d'optimisation qui imite le refroidissement lent de métaux, qui se caractérise par une réduction progressive des mouvements atomiques qui réduisent la densité des défauts réticulaires jusqu'à un état d'énergie le plus bas est atteint [**101,102**].

De plus, l'efficacité de la convergence a été testée par l'exécution du processus d'optimisation, 20 fois pour chaque algorithme, et le calcul de la moyenne de l'écart type des erreurs. Ce test a été effectué sur les deux modèles MSD et MDD.

i. Modèle à une seule diode (MSD)

Un test d'efficacité de convergence est exécuté vingt fois avec les mêmes conditions initiales pour tous les algorithmes. La courbe I-V expérimentale du module ISOFOTON est obtenue à une température de 46.08 ° C et à une irradiation de 723,37 W/m². Le tableau 3.5 résume les résultats de comparaison.

Comme le démontrent les résultats dans le tableau 3.5, il est très clair que le FPA donne des meilleurs résultats pour les mêmes conditions initiales. Ceci est exprimé par les trois grandeurs suivantes: RMSE qui est égale à 0,022, la valeur moyenne de toutes les erreurs obtenues qui est égale 0,0223 et enfin l'écart type STD qui est égale à 1, 051E-06.

En outre, le facteur le plus puissant de l'algorithme FPA est sa vitesse de convergence à la solution optimale, le temps de convergence à la meilleure solution par FPA est de 22 secondes pour 4185 itérations. Alors que, les deux autres algorithmes prennent plus de temps et plus d'itération, 30 secondes pour l'algorithme FA pour 6632 itérations et 28 secondes pour l'algorithme SA pour 6421 itérations.

Tableau 3.5 Les résultats obtenus par les trois algorithmes appliqués sur le MSD.

Paramètre	FPA	FA	SA
$R_s(\Omega)$	0,238	0,0924	0,064
$R_{sh}(\Omega)$	400	318,331	348,805
$I_{ph}(A)$	4,710	4,761	4,841
$I_0(A)$	5,171.10^-6	0,0001	0,0002
n	50,643	67,175	71,172
$RMSE$	0,022	0,0457	0,064
$MEAN$	0,0223	0,0502	0,087
STD	1,051.10^-6	0,002	0,037

Par ailleurs, en tenant compte de la RMSE des trois algorithmes, le FPA est plus précis selon les deux autres algorithmes comme il est mis en évidence dans le tableau 3.5. Dans ce dernier, la valeur optimale de R_{sh} extraite à l'aide du FPA est égale à 400 ohms.

Cette valeur est identique à la limite supérieure de l'espace de recherche. Ainsi, l'algorithme fournit une recherche des paramètres optimaux dans tout l'espace de recherche avec un minimum d'itérations et moins de temps. Cela est dû au basculement entre l'équation globale et l'équation locale. De plus, cela augmente la probabilité que l'algorithme converge vers l'optimum global le plus rapidement possible.

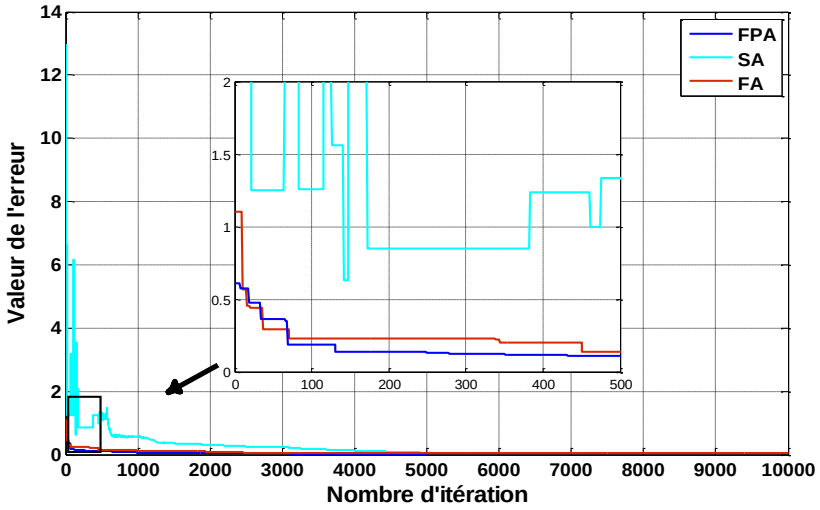
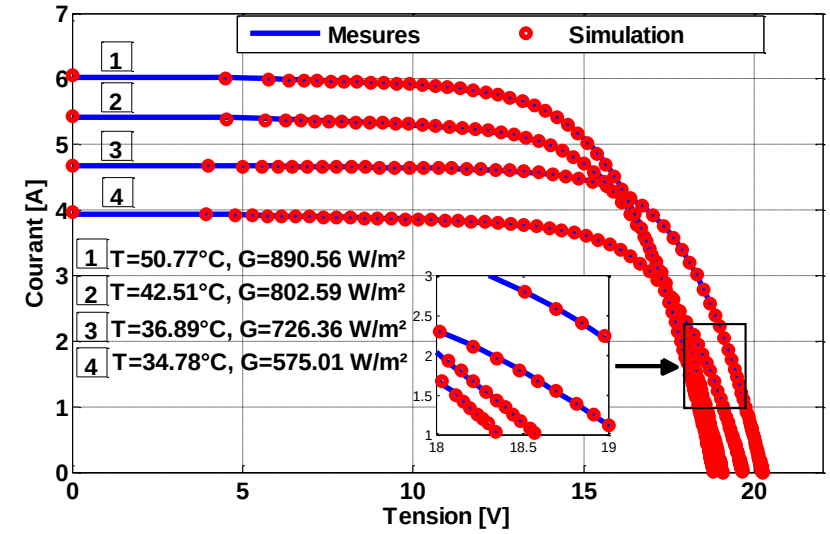


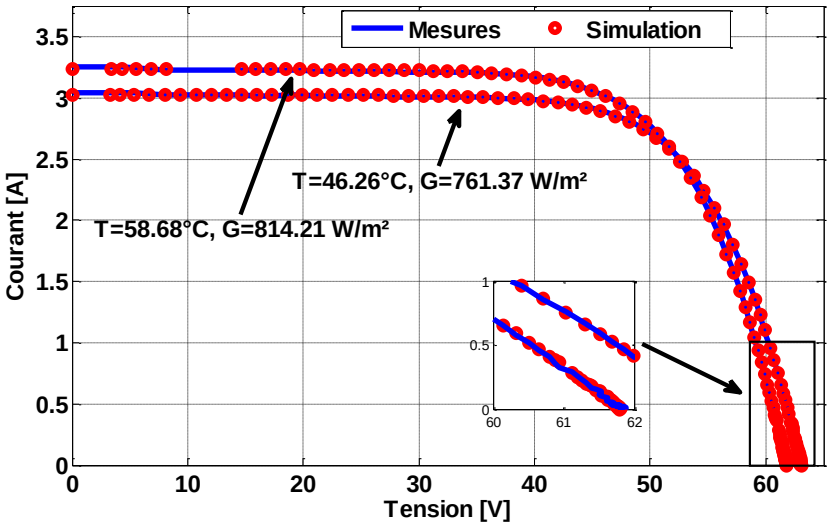
Figure 3.6 Les courbes de la variation des erreurs obtenues pour les trois algorithmes FPA, SA et FA utilisant le MSD.

La variation de l'erreur au cours du processus d'optimisation utilisant le FPA est représentée sur la figure 3.6. Cette figure montre que pour le MSD, le FPA converge vers l'erreur minimale avec une plus grande vitesse que celle des deux autres algorithmes.

Les figures 3.7 et 3.8 montrent une comparaison entre les courbes expérimentales et simulées pour les deux modules PV dans différentes conditions météorologiques (température et irradiation). Elles montrent clairement l'efficacité du FPA dans l'identification des paramètres pour les quatre modules PV dans différentes conditions météorologiques.

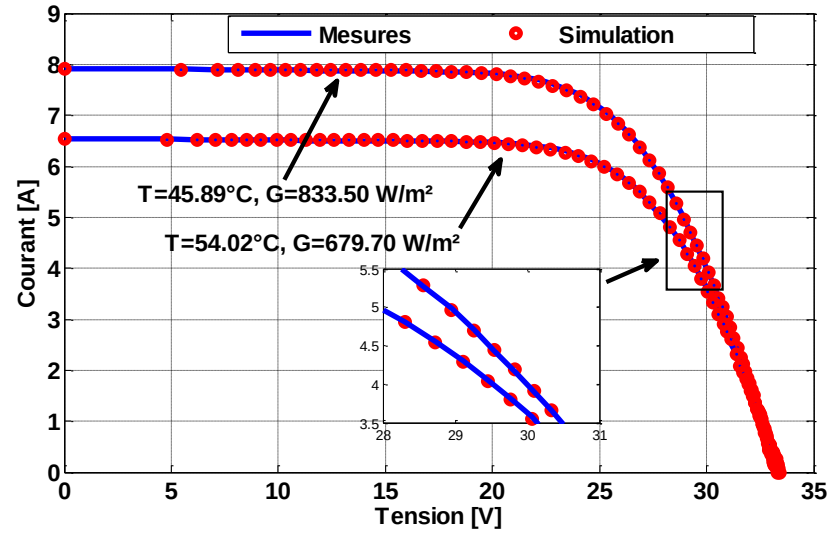


(a)

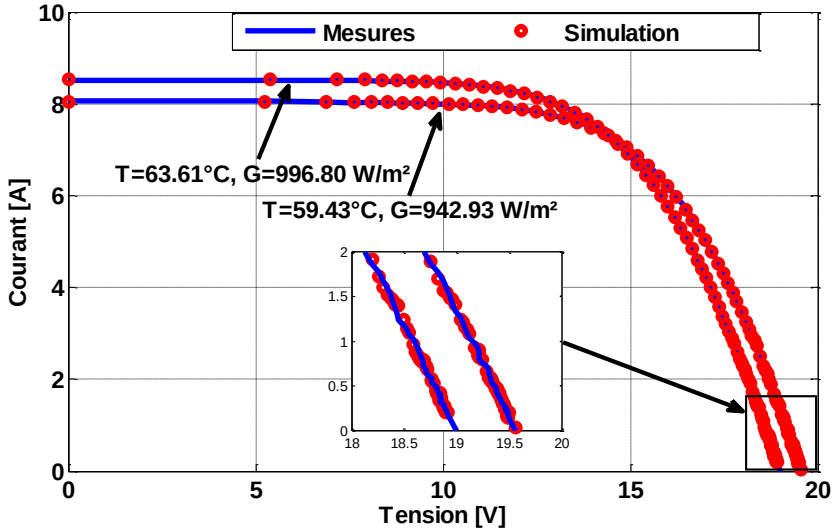


(b)

Figure 3.7 Données mesurées vs simulées avec MSD pour cellule de type Monocristallin **(a)** ISOFOTON 106 **(b)** SANYO.



(a)



(b)

Figure 3.8 Données mesurées vs simulées avec MSD pour cellule de type Polycristallin **(a)** CONDOR **(b)** KYOCERA.

ii. Modèle à double diode (MDD)

Pour assurer l'efficacité de la convergence de l'algorithme, le FPA est testé dans un modèle plus compliqué; donc au lieu de 5 inconnus cette fois-ci l'algorithme est appliqué sur le MDD pour identifier sept paramètres en suivant les mêmes étapes que celles de MSD.

Tableau 3.6 Les résultats obtenus par les trois algorithmes dans le cas de MDD.

Paramètre	FPA	FA	SA
$R_s(\Omega)$	0,166	0,016	0,022
$R_{sh}(\Omega)$	395,010	376,758	259,958
$I_{ph}(A)$	4,674	4,715	4,775
$I_{01}(A)$	1,072.10^-6	7,571.10^-5	0,0001
$I_{02}(A)$	1,743.10^-6	8,085.10^-5	0,0017
n_1	77,007	68,672	73,155
n_2	51,249	111,123	106,507
$RMSE$	0.0186	0.04321	0,062
$MEAN$	0,0224	0,0472	0,099
STD	0,0038	0,0056	0,037

Le tableau 3.6 montre clairement que, pour les mêmes conditions initiales, le FPA donne de meilleurs résultats par rapport aux deux autres algorithmes. Après avoir calculé le temps de la convergence vers l'optimum globale pour les trois algorithmes on a obtenu les résultats suivants: le FPA prend 27 secondes pour 7356 itérations, alors que les deux autres algorithmes prennent plus de temps, donc plus d'itérations. 40 secondes pour l'algorithme FA pour 9745 itérations et 36 secondes pour l'algorithme SA pour 9342 itérations.

Les courbes de la variation de l'erreur durant le processus d'optimisation pour ses algorithmes sont données par la figure 3.9.

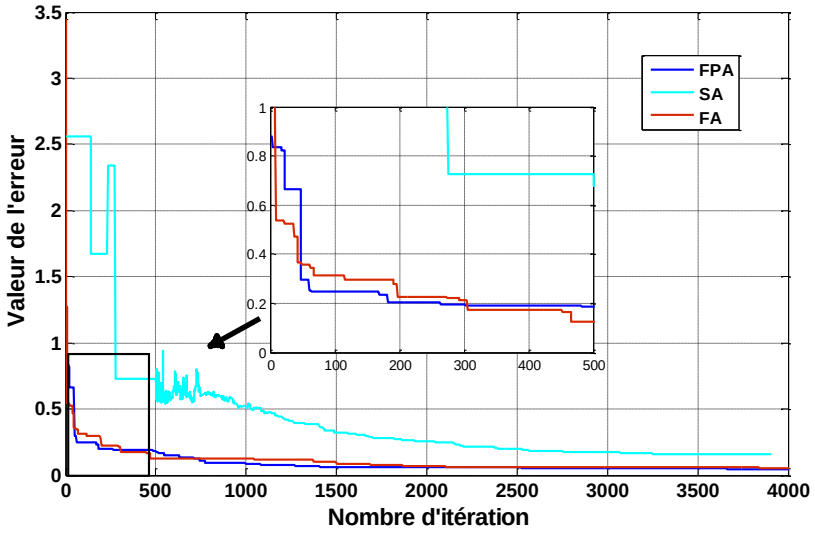
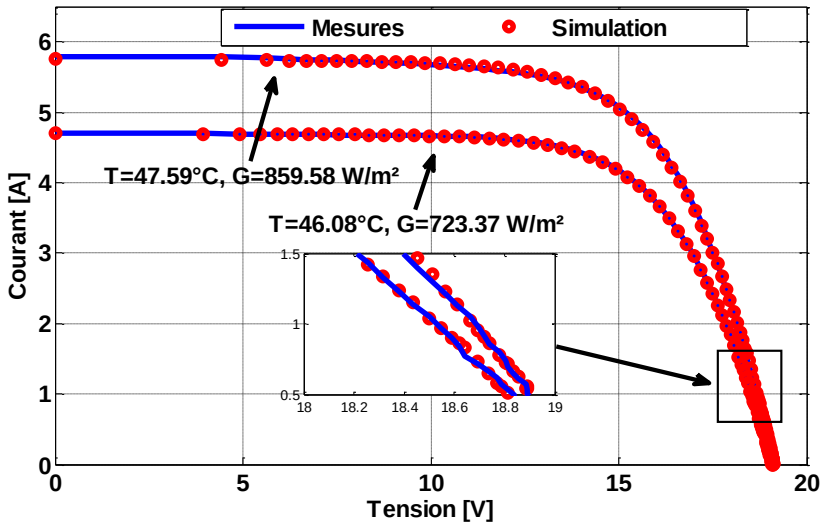


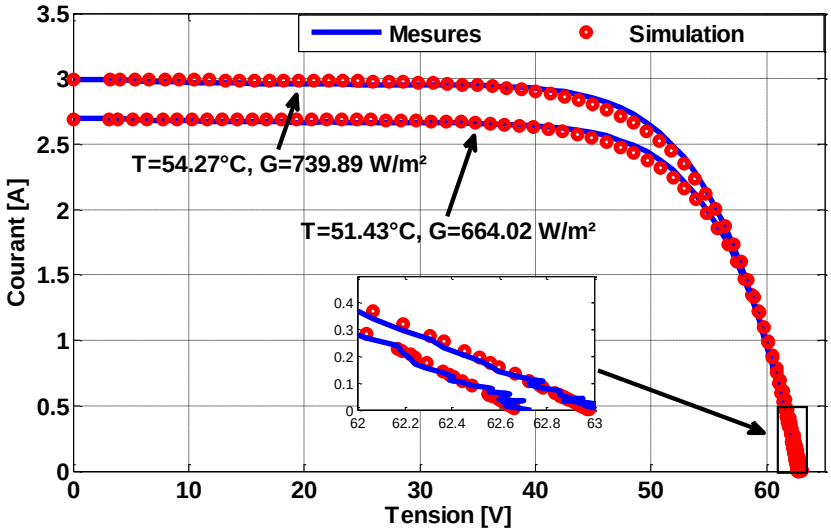
Figure 3.9 Les courbes de la variation des erreurs obtenues pour les trois algorithmes FPA, FA et SA dans le cas du MDD.

Cette figure montre l'efficacité du FPA et sa convergence vers la meilleure solution avec erreur minimale et une grande vitesse dans le cas du MDD en le comparant avec les deux autres algorithmes.

Les figures 3.10 et 3.11 montrent également une comparaison entre les courbes expérimentales et celles simulées pour les quatre modules PV pour différentes conditions météorologiques (température et irradiation).



(a)



(b)

Figure 3.10 Données mesurées vs simulées avec MDD pour cellule de type Monocristallin **(a)** ISOFOTON 106 **(b)** SANYO.

D'après ces figures, le FPA montre son efficacité dans l'extraction des sept paramètres du MDD cités ci-dessus.

3.6 L'estimation de MPP

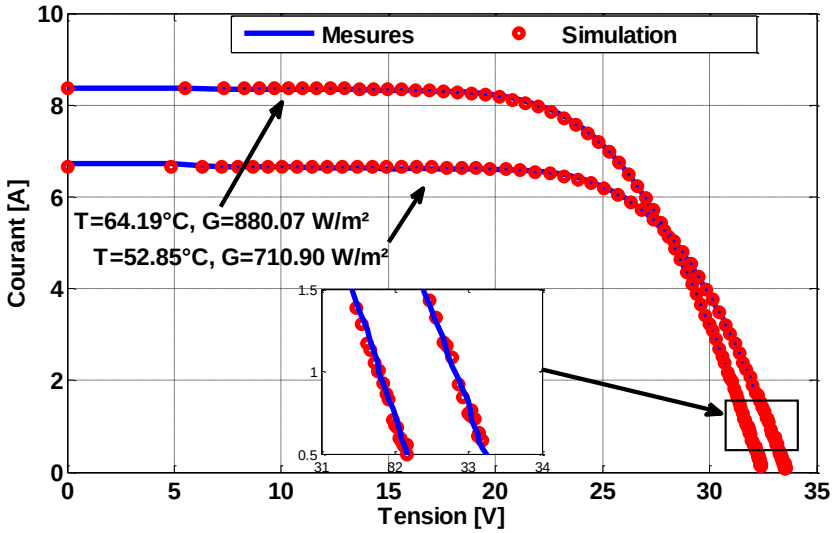
Chaque courbe I-V du module PV peut être représentée par une fonction équivalente de nature non linéaire telle que le MSD ou le MDD. Cette fonction a un point de puissance maximale. En outre, chaque courbe I-V a son propre point qui peut être représenté par un couple courant/tension au maximum. Ces derniers sont désignés respectivement par I_{mpp} et V_{mpp}. La puissance au maximum est désignée par P_{mpp}. Le passage de la courbe I-V statique vers la courbe dynamique est basé sur une méthode d'estimation [**99**]. Cette méthode utilise les informations données par le fabricant du module PV avec les paramètres du MSD déjà identifiés.

Les équations (3.12), (3.13) et (3.14) sont utilisées pour estimation de la valeur de l'I_{mpp}, V_{mpp} et P_{mpp}:

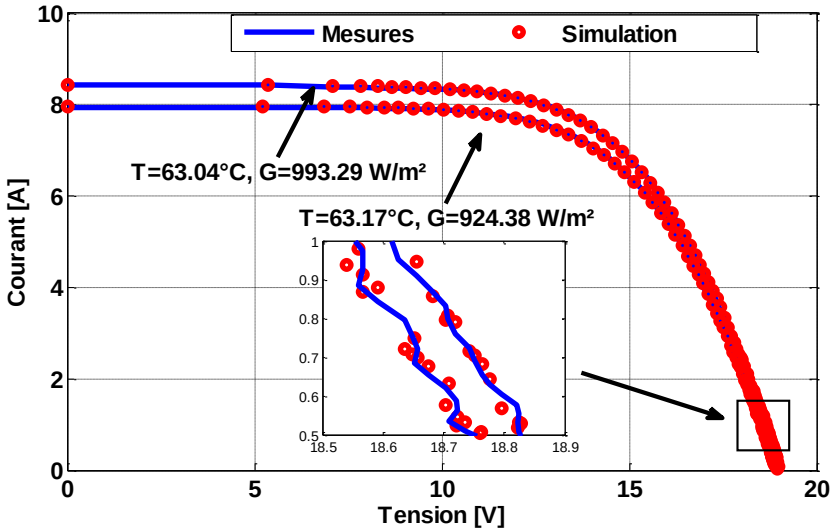
$$V_{P\max 0} = \frac{V_{mpp}}{1+C_T\left(T_j+T_{j0}\right)} + V_T \frac{T_{j0}}{T_j}\ln\left(\frac{G_0}{G_{eff}}\right) - I_{mpp}R_s\left(\frac{G_0}{G_{eff}}-1\right) \qquad (3.12)$$

$$I_{P\max 0} = I_{mpp}\frac{G_0}{G_{eff}} \qquad (3.13)$$

$$P_{mpp} = I_{mpp} \times V_{mpp} \qquad (3.14)$$



(a)



(b)

Figure 3.11 Données mesurées vs simulées avec MMD pour cellule de type Polycristallin **(a)** CONDOR **(b)** KYOCERA

Où: G_0 et T_{j0} représentent, respectivement, l'irradiation et la température dans les conditions de test standard (STC) avec G_0 = 1000 W/m² et T_{j0} = 25 °C. Rs est la résistance en série du MSD. C_T est le coefficient de la température de puissance, tandis que la valeur de T_j est calculée par l'équation (3.15):

$$T_j\left(G_{eff},T_{amb}\right)=T_{amb}+\left(NOCT-T_{ambN}\right)\left(\frac{G_0}{G_{eff}}\right) \tag{3.15}$$

Dans l'équation (3.15), NOCT représente la relation entre la température de la cellule PV et la température de la cellule PV dans les conditions de fonctionnement nominal. Cette

valeur peut être trouvée dans les informations données par le fabricant (document technique du module). Nous trouvons que NOCT = 48 °C. G_N et T_{ambN} sont l'irradiation et la température ambiante dans les conditions de fonctionnement normal. Leurs valeurs sont respectivement égales à 800 W/m² et 20 °C. V_T est la tension thermique qui est exprimée par l'équation (3.16):

$$V_T = \frac{nKT}{q} \tag{3.16}$$

V_{Pmax0} et I_{Pmax0} sont la tension et le courant au MPP dans des conditions nominales et leurs valeurs sont données par le fabricant du module PV.

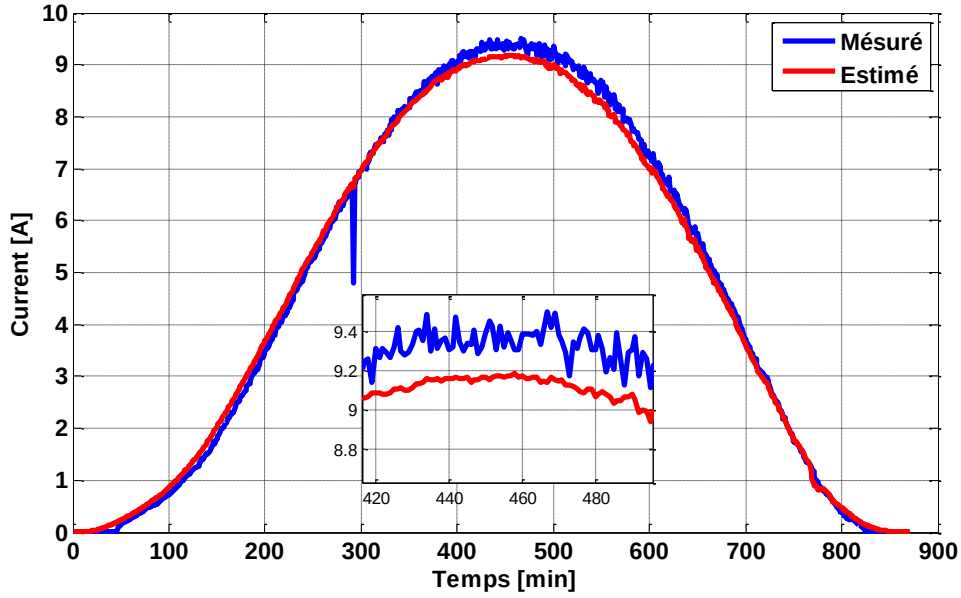
Les valeurs de paramètres du MSD trouvées précédemment permettent de prédire le MPP en utilisant les équations (3.12) et (3.13) mentionnées ci-dessus. Les données utilisées sont obtenues pour deux jours. Ces jours étant caractérisés par des conditions météorologiques différentes telles que ciel clair et ciel nuageux.

En outre, les données expérimentales sont collectées à partir d'un système PV connecté au réseau (SPVCR) installé dans le CDER à Alger [**35**]. Le système génère une puissance égale à 3,18 KW à partir d'un champ PV qui contient 30 modules PV structurés en deux chaînes. Chacune compte 15 modules en série et le type du module de ce système est Monocristallin (ISOFOTON). Pour évaluer les résultats obtenus, les valeurs de prédictions et les valeurs expérimentales sont tracées dans les mêmes figures ci-dessous. Ces figures montrent les courbes de la variation dynamique pour les trois variables le courant, la tension et la puissance au MPP. Chaque grandeur électrique ayant deux valeurs: la première est prédite et l'autre est mesurée à partir du SPVCR. Les figures montrent que la courbe prédite et celle mesurée ont une bonne correspondance. Par conséquent, les paramètres SDM extraits par le FPA ont évidemment une grande précision.

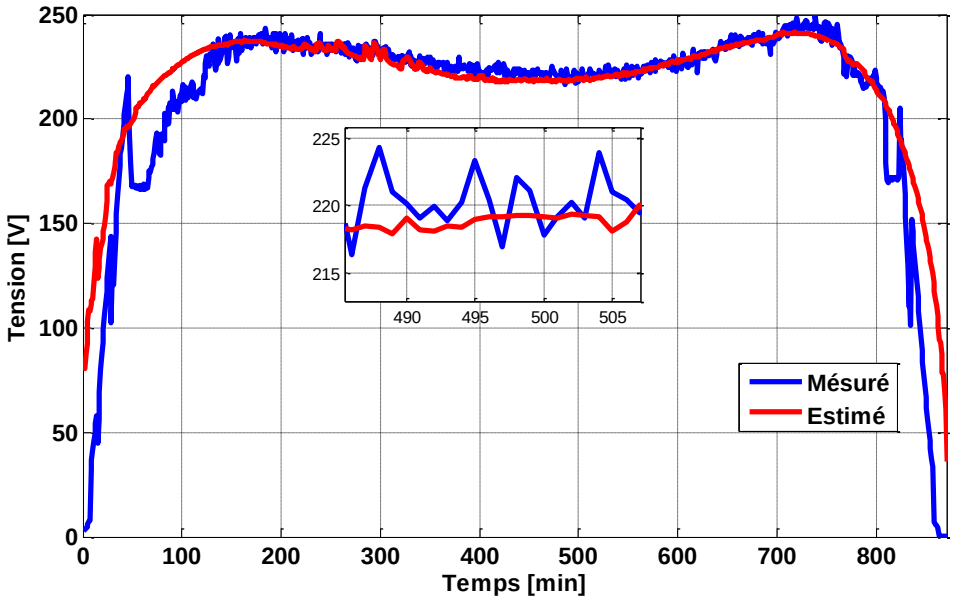
3.7 Conclusion

Dans ce chapitre, l'algorithme de pollinisation des fleurs a été utilisé pour extraire les paramètres des modèles MSD et MDD. L'algorithme FPA est utilisé pour identifier les paramètres de modules PV pour deux fabricants différents. Nous avons également comparé les résultats obtenus par le FPA avec deux autres algorithmes qui sont l'algorithme luciole (FA) et l'algorithme de recuit simulé (SA). Plusieurs données expérimentales des courbes I-V sont comparées avec d'autres données simulées. Les résultats de cette comparaison

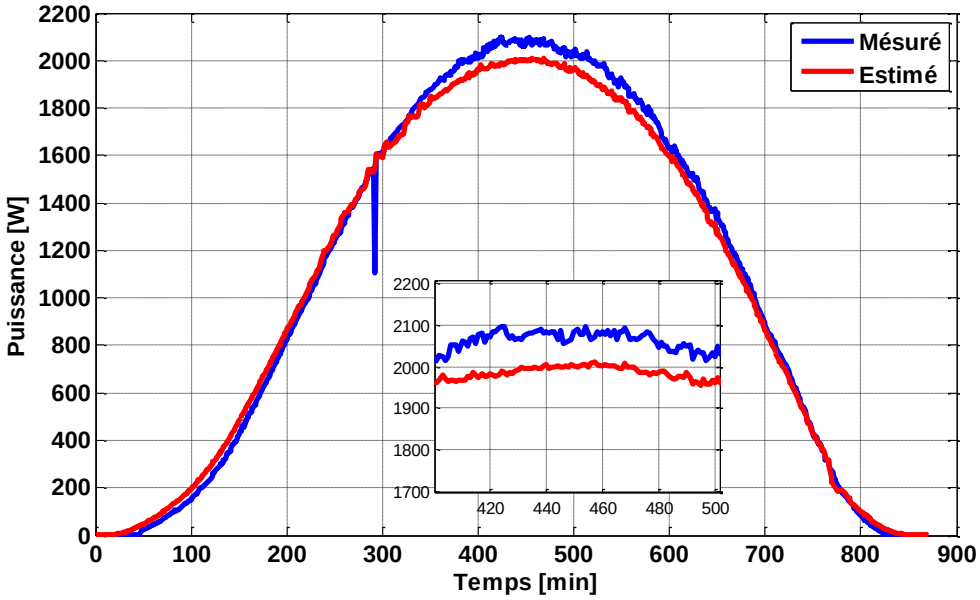
montrent que le FPA proposé donne de meilleures performances. Il est donc clair que le FPA est plus efficace que les deux autres algorithmes pour l'extraction des paramètres optimaux des modules PV. A la fin du chapitre, nous avons vu comment les paramètres extraits par le FPA sont utilisés pour estimer le point de puissance maximale et les résultats ont été comparés aux MPP mesurés du système photovoltaïque connecté au réseau. Encore une fois, le FPA a donné de meilleures performances.



(a)

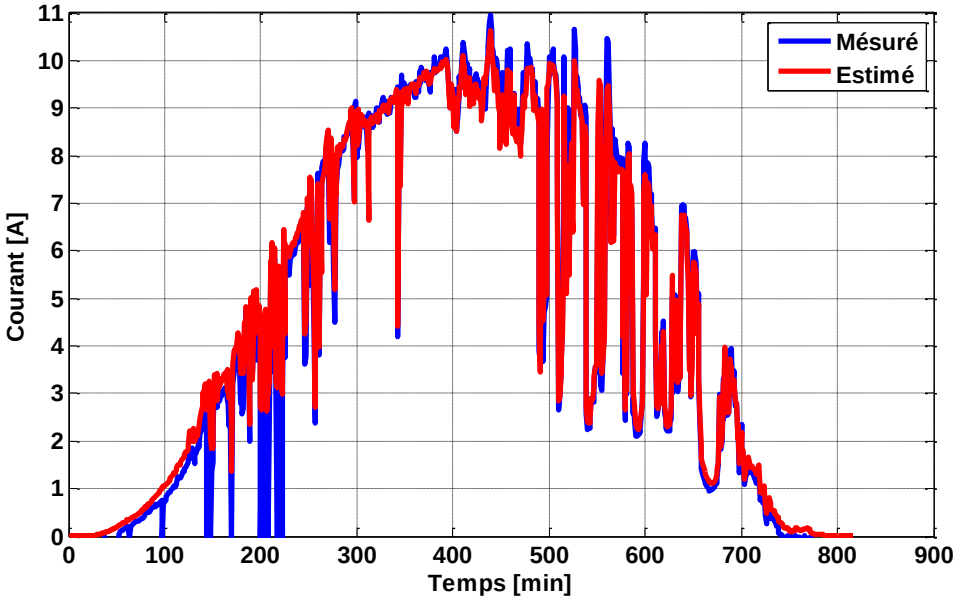


(b)

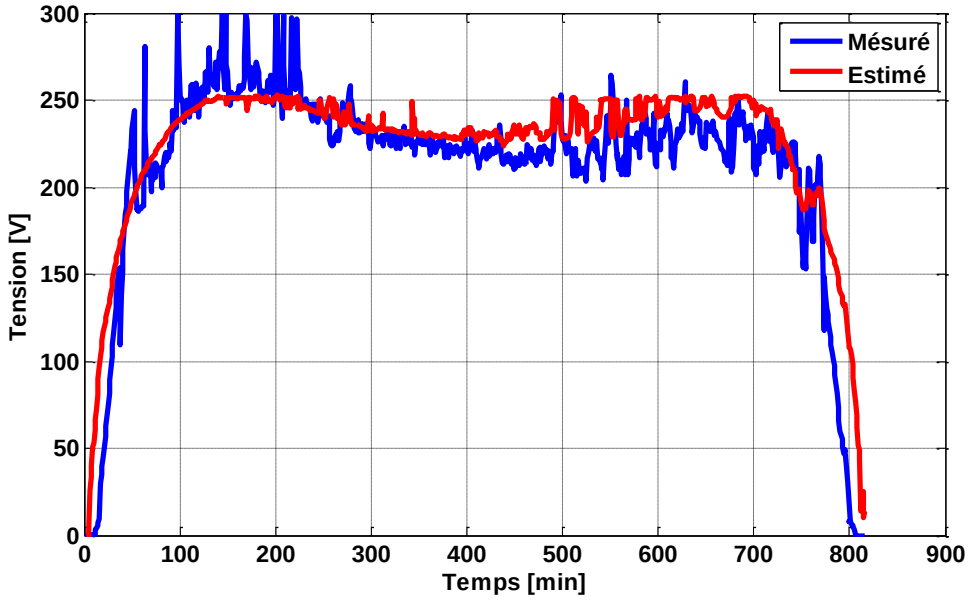


(c)

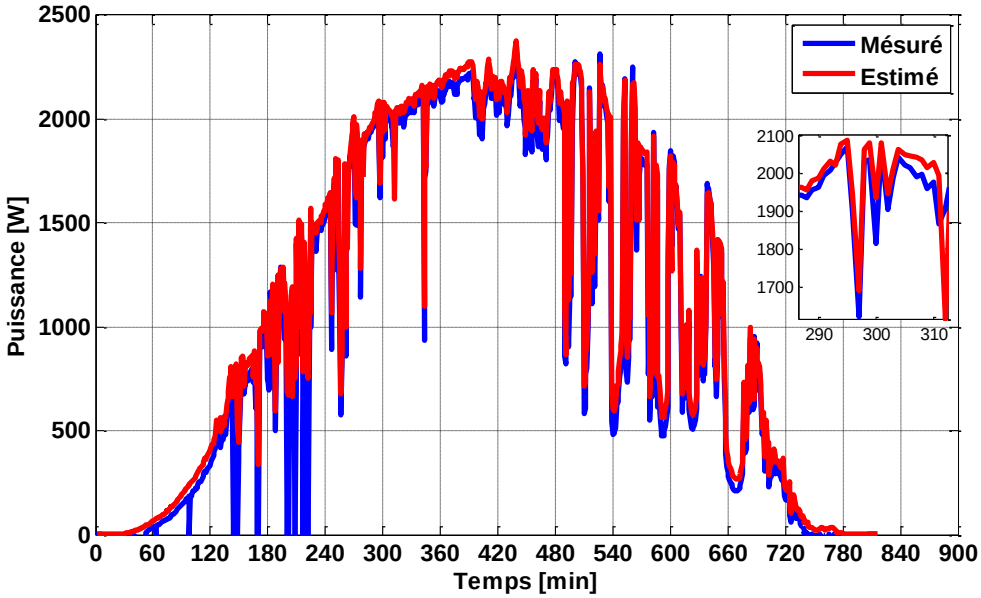
Figure 3.12 Comparaison entre les valeurs réelles et estimées au MPP du ciel clair pour (a) courant, (b) tension et (c) puissance.



(a)



(b)



(c)

Figure 3.13 Comparaison entre les valeurs réelles et estimées au MPP du ciel nuageux pour (a) courant, (b) tension et (c) puissance.

CHAPITRE 4

DETECTION ET DIAGNOSTIC DES DEFAUTS PAR L'ARBRE DE DECISION

Chapitre 4

Détection et Diagnostic des Défauts par l'Arbre de Décision

4.1 Introduction

Dans ce chapitre, une nouvelle approche basée sur l'algorithme d'arbre de décision pour détecter et diagnostiquer les défauts dans le système photovoltaïque connecté au réseau (SPVCR) est proposée. Un modèle non paramétrique pour prédire l'état du SPVCR par une tâche d'apprentissage est utilisé; un ensemble de données est recueilli à partir du SPVCR par le système d'acquisition dans plusieurs conditions météorologiques. Trois attributs numériques et deux cibles sont choisis pour les données utilisées. Ces attributs sont : la température ambiante, l'irradiation et le rapport de puissance calculé à partir de la puissance mesurée et estimée. La première cible représente l'état du système qui possède deux cas, soit sain soit défectueux (pour le modèle de détection). La seconde cible contient quatre étiquettes de classes nommées sain, défaut de chaîne, défaut de court-circuit ou défaut de ligne-ligne (c'est pour le modèle de diagnostic). Le modèle SANDIA est appliqué pour estimer la puissance générée par le SPVCR en fonctionnement sain (sans aucun défaut).

L'ensemble de données a été divisé en deux parties; 66 % ont été utilisées pour l'apprentissage et le reste pour le test. Par la suite, des nouvelles données obtenues pendant cinq jours sont utilisées afin d'évaluer la robustesse et l'efficacité des deux modèles. Les résultats des tests indiquent que les modèles ont une haute performance de prédiction avec la précision est supérieure à 99%.

4.2 Modèle de SANDIA

La modélisation du champ PV peut être réalisée par plusieurs approches soit un modèle paramétrique par exemple dans le cas d'une seule diode ou double diode, et par fois par des modèles non paramétriques comme le réseau neurone ou des modèles basés sur les neurones-fuzzy [**103**]. Parfois, on utilise les paramètres de modèle MSD pour l'estimation

des variables électriques du système en régime dynamique (courant/tension en MPP). En raison de ses performances dynamiques satisfaisantes, le modèle empirique élaboré par le laboratoire SANDIA est également pris en compte pour l'étude dynamique [**104**]. Ce modèle prend en compte tous les comportements électriques, thermiques, optiques et photovoltaïques et s'adapte avec n'importe quelle technologie de panneau PV (Monocristallin ou Poly-cristallin), en plus de sa pertinence pour le grand nombre de champ PV associés [**105**]. En outre, ce modèle peut prédire les performances du module PV au cours de son fonctionnement en raison de sa simplicité. Les variables I_{mpp}, V_{mpp} et P_{mpp} sont données par les équations (4.1-4.5).

$$I_{mpp} = I_{mpp_STC}\,(C_0 E_e + C_1 E_e^2)(1+\alpha_{mpp}\,(T_c - T_{STC})) \tag{4.1}$$

$$V_{mpp} = V_{mpp_STC} + C_2 N_s \cdot \delta(T_c) \cdot \ln(E_e) + C_3 N_s\,(\delta(T_c)\ln(E_e))^2 + \beta_{mpp}\,(T_c - T_{STC}) \tag{4.2}$$

$$E_e = \frac{G}{E_0} \tag{4.3}$$

$$\delta(T_c) = n \cdot k \cdot (T_c + 273.15)/q \tag{4.4}$$

$$P_{mpp} = I_{mpp}.V_{mpp} \tag{4.5}$$

Où I_{mpp_STC}, V_{mpp_STC} et T_{STC} représentent respectivement le courant maximal, la tension maximale et la température à la condition de test standard (STC). K est la constante de Boltzmann et q la charge d'électrons. T_c est la température de la cellule et δ (T_c) est la tension thermique. E_e, E_0 et G sont respectivement les irradiations : effective, nominale et mesurée. α_{mpp} est le coefficient de température à la I_{mpp}, et β_{mpp} est le coefficient de température à la V_{mpp}. De C_0 à C_3 et n sont des paramètres empiriques qui seront identifiés. En effet, la température mesurée par le système d'acquisition est la température ambiante, et comme mentionnée précédemment (voir les équations (4.1) et (4.2)), la température de la cellule est nécessaire pour permettre d'estimer la puissance à l'aide du modèle SANDIA, alors que la température de la cellule doit être estimée. Ainsi, plusieurs approches sont proposées dans la littérature pour estimer la température cellulaire. La plus populaire étant

le modèle thermique de Ross qui utilise la température ambiante et l'irradiation solaire [**106**]. Le modèle de Ross est donné par l'équation (4.6).

$$T_c = T + (T_{NOCT} - 20℃) E_e \qquad (4.6)$$

Où T est la température ambiante. T_{NOCT} est la température en fonctionnement nominale de la cellule qui est égale à 48 ° C. Cette valeur est donnée par le fabricant du module solaire.

4.2.1 Identification des paramètres de modèle de SANDIA

Pour identifier les paramètres du modèle SANDIA susmentionné, on a utilisé l'algorithme FPA [**7, 97,98**]. Des mesures réelles du courant I_{dc} et de la tension V_{dc} permettant de calculer la puissance P_{dc} générée par le générateur photovoltaïque pendant une journée. Sept paramètres ont été concernés par cette identification. Fondamentalement, le processus consiste à minimiser la fonction de coût exprimée par les équations (4.7) et (4.8) qui représentent l'erreur quadratique moyenne entre la courbe de puissance mesurée et la courbe de puissance simulée. Les paramètres extraits sont répertoriés dans le tableau 4.1.

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k} f\left(P_{dc}, P_{mp}(\theta)\right)} \qquad (4.7)$$

$$f\left(P_{dc}, P_{mp}(\theta)\right) = P_{dc} - N_s * N_p * P_{pm}(\theta) \qquad (4.8)$$

Où N_s et N_p sont respectivement les nombres de modules en série et en parallèle dans le champ PV. P_{dc} est la puissance mesurée au point de puissance maximale générée par le système au cours de son fonctionnement. Elle est calculée par le produit de I_{dc} et V_{dc} ; ces derniers sont le courant et la tension mesurée à côté DC de l'onduleur connecté au réseau. K est le nombre total d'échantillons mesurés. P_{mp} (θ) est la puissance prédite par le modèle de SANDIA dont le vecteur de paramètres inconnus θ, c'est-à-dire θ = [α_{mpp} β_{mpp} C_0 C_1 C_2 C_3 n].

Tableau4.1 Les valeurs des paramètres identifiés.

Paramètre	**α_{mpp}**	**β_{mpp}**	**C_0**	**C_1**	**C_2**	**C_3**	**n**
valeur	-0.0001	-0.0600	0.9211	0.0096	-0.0009	-18.502	1.8624

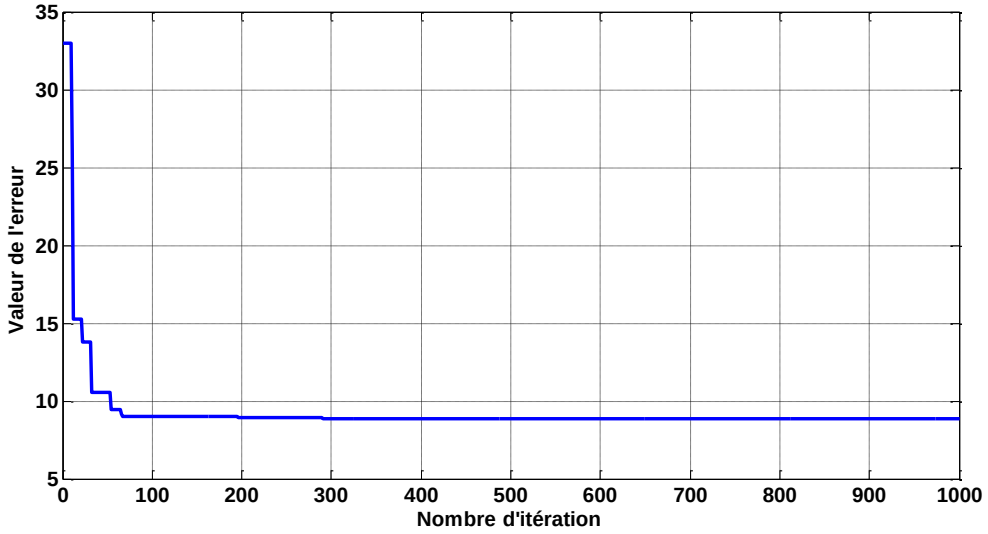


Figure 4.1 La courbe de la variation d'erreur durant le processus d'identification des paramètres.

La variation d'erreur au cours du processus d'optimisation est donnée par la figure 4.1. La courbe indique que la valeur minimale de RMS est de 8,82 Watt. La corrélation entre les deux puissances est représentée sur la figure 4.2, on constate que le coefficient de corrélation est égal à 0,9911. La figure 4.3 montre également la correspondance entre la puissance mesurée et la puissance simulée.

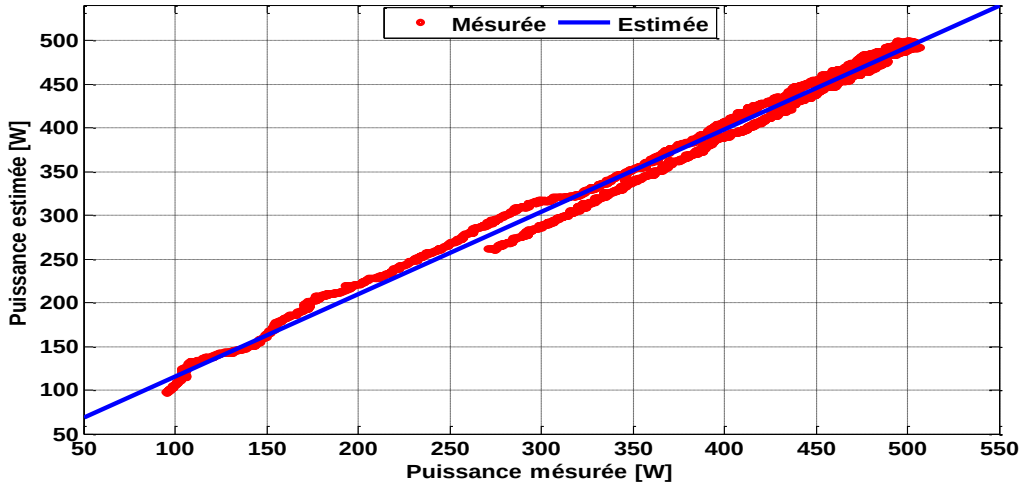
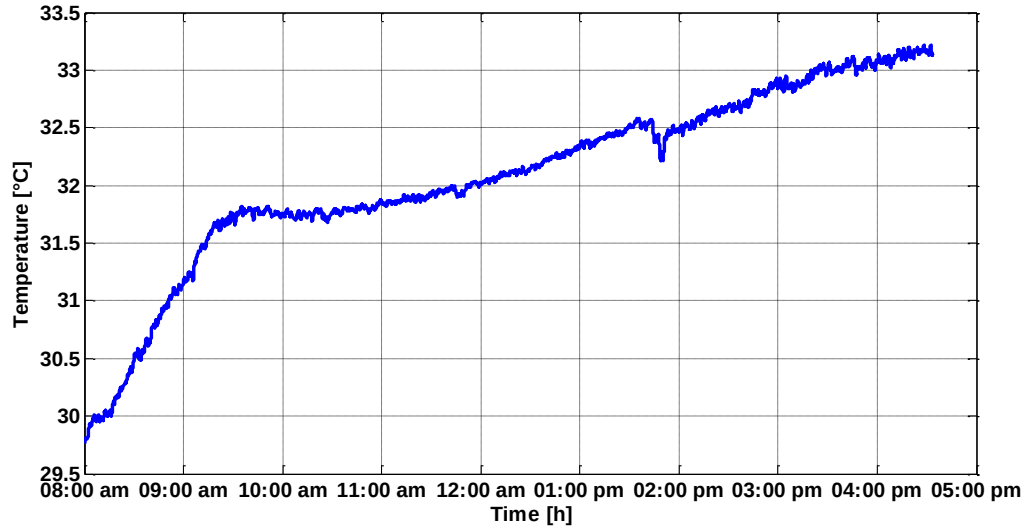
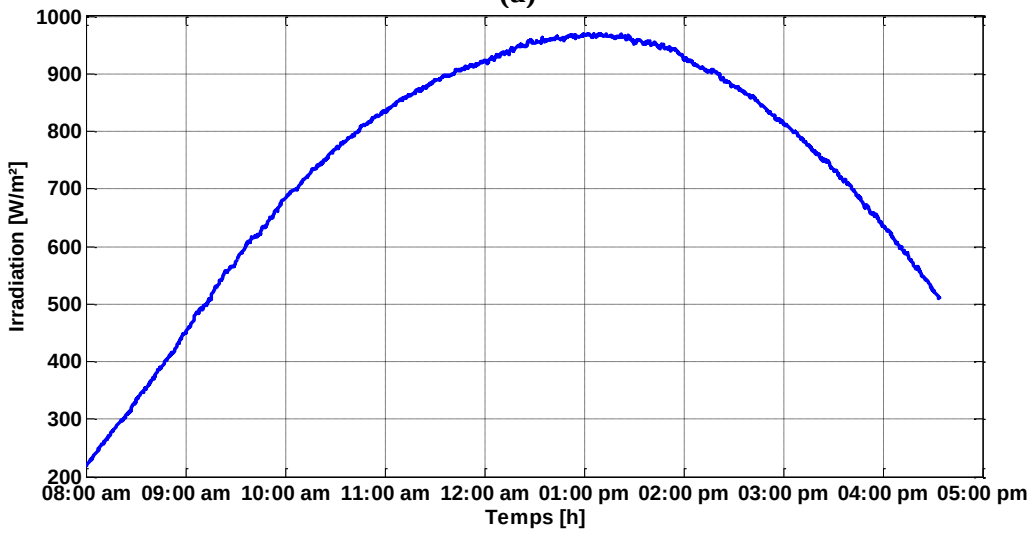


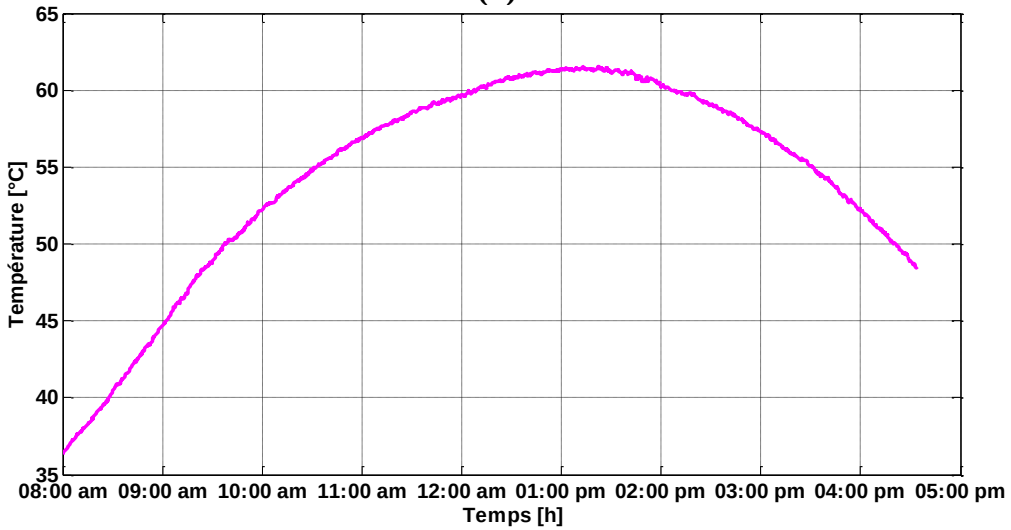
Figure 4.2 Représentation de dispersion de la puissance mesurée et la puissance estimée.


Temperature [°C]
33.5
33
32.5
32
31.5
31
30.5
30
29.5
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm 05:00 pm
Time [h]

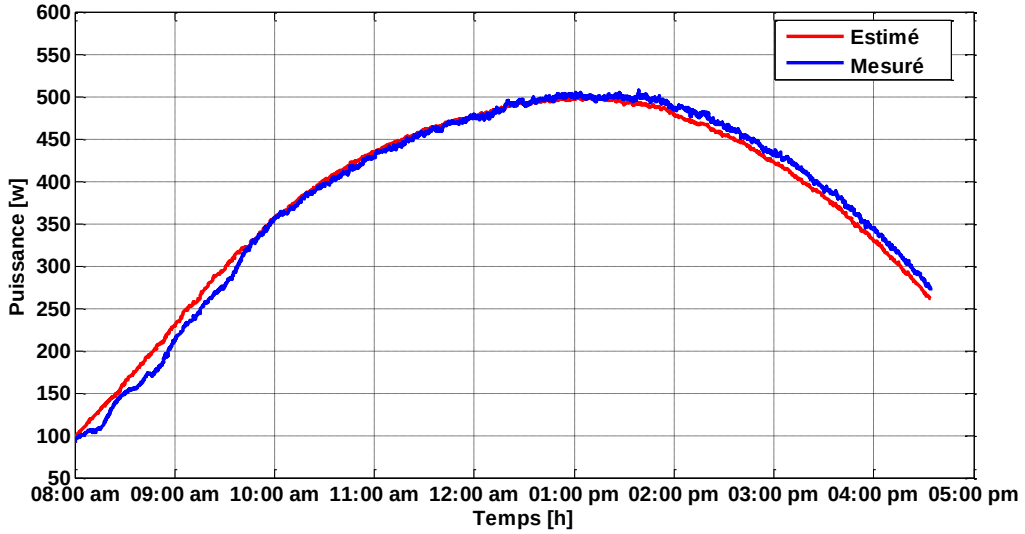
(a)


Irradiation [W/m²]
1000
900
800
700
600
500
400
300
200
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm 05:00 pm
Temps [h]

(b)


Température [°C]
65
60
55
50
45
40
35
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm 05:00 pm
Temps [h]

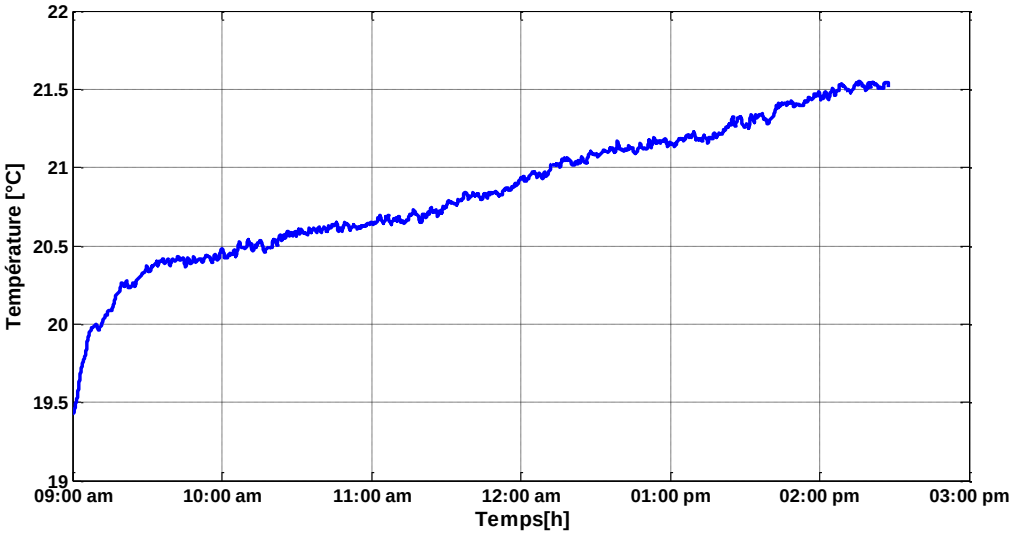
(c)



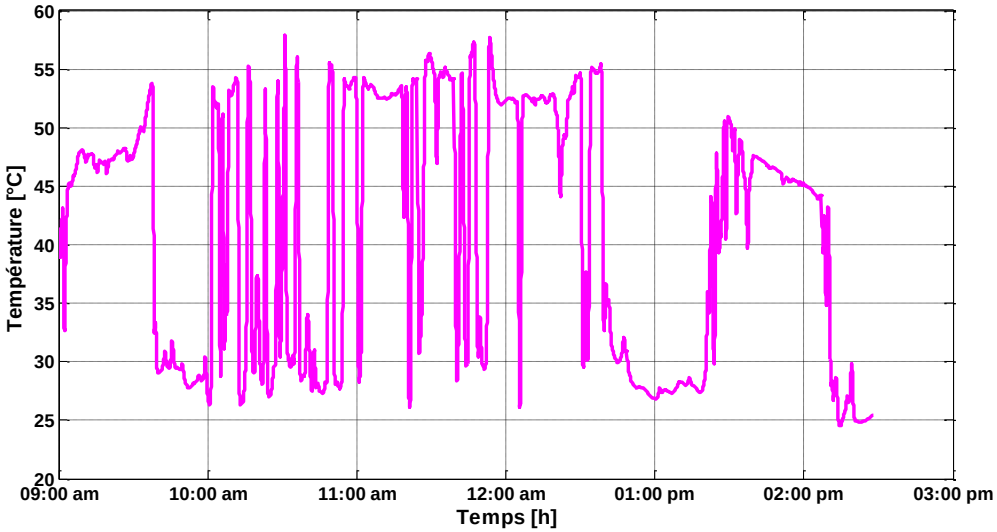
(d)

Figure 4.3 Profil d'un jour à ciel clair: (a) température ambiante; (b) l'irradiation; (c) température de cellule estimée; (d) puissance estimée VS mesurée du SPVGC (20 juillet 2017).

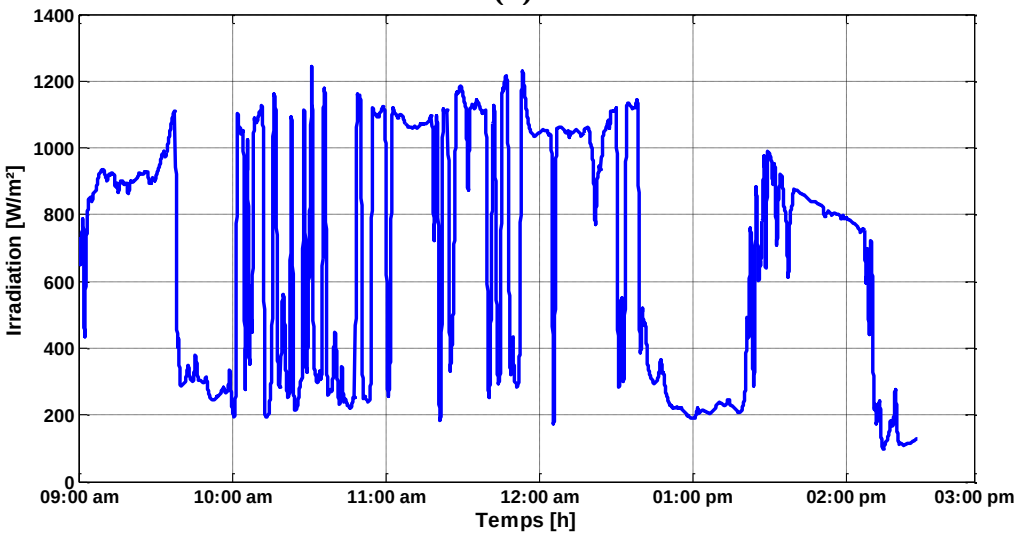
De plus, le modèle SANDIA obtenu est testé avec un autre profil qui a été enregistré pour différentes conditions météorologiques, comme représenté à la figure 4.4. La valeur de RMS est égale à 10,44 watts pour cette dernière. La figure 4.4 montre la similitude entre les puissances mesurées et prédites, ce qui est également significatif du fait que le modèle SANDIA a une bonne fiabilité.



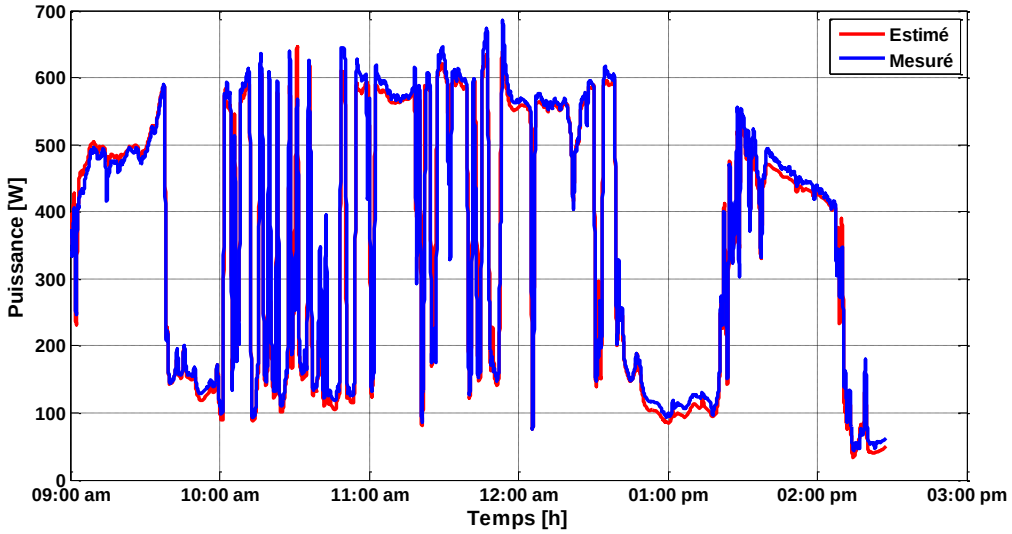
(a)


Température [°C]
60
55
50
45
40
35
30
25
20
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(b)


Irradiation [W/m²]
1400
1200
1000
800
600
400
200
0
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(c)



(d)

Figure 4.4 Profil d'un jour à ciel nuageux (a) température ambiante; (b) température estimée de la cellule; (c) irradiation; (d) puissance estimé VS la puissance mesurée du SPVCR (4 novembre 2017).

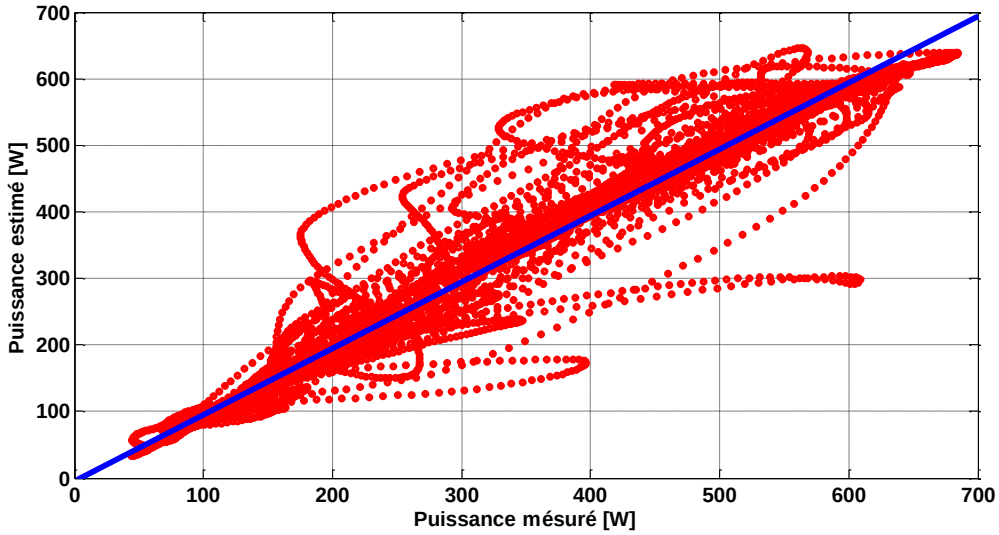


Figure 4.5 Représentation de dispersion de la puissance mesurée et la puissance estimée d'une journée à ciel nuageux.

4.3 Les défauts considérés

Dans le présent travail, nous avons introduit intentionnellement des défauts dans la centrale PV afin de vérifier notre approche. Les défauts considérés sont: chaîne ouverte (S_f), court-circuit d'un module (SC_f) et enfin le cas d'un défaut de ligne avec une

résistance de valeur de 7 ohm (LL_f), le système à l'état sain symbolisé par Ff. Dans le schéma ci-dessous nous donnons une illustration graphique de ces défauts.

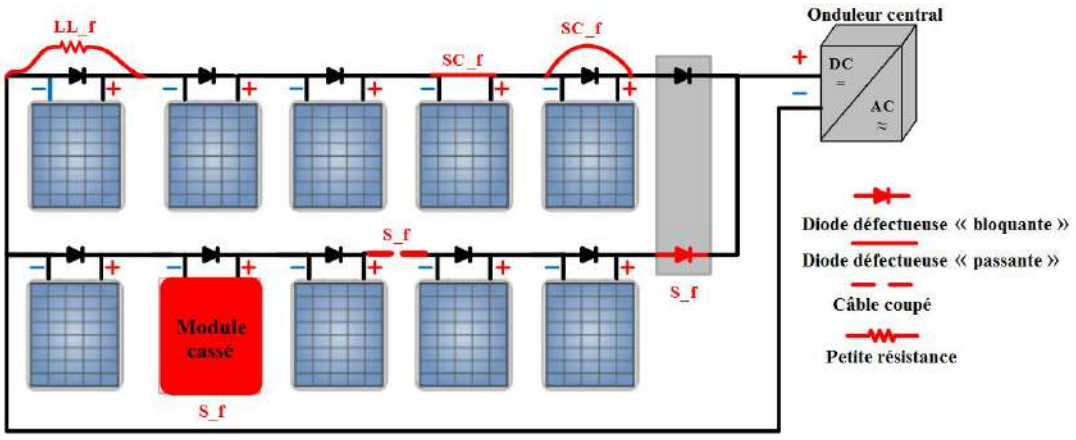


Figure 4.6 Les Différents types de défauts considérés dans le système PV.

Les défauts sont créés artificiellement dans le système PV, où le défaut de chaîne ouverte est simplement obtenu en déconnectant une série de modules, tel que cinq modules PV en série sont débranchés. Ce défaut est conservé dans le système pendant six jours, les 25, 26, 27 et 28 juillet et les 10 et 17 août 2017. De plus, le défaut de court-circuit est provoqué par le court-circuit d'un module PV dans une chaîne présélectionnée. Cet état est maintenu dans le système pendant quatre jours, les jours choisis sont les 30, 31 juillet et 18,19 août 2017. Finalement, le défaut de ligne est obtenu en retirant un module PV et en le remplaçant par une résistance de 7 ohms, et le système reste avec ce défaut pendant cinq jours sont les 11 et 12 juillet et les 22, 23 et 24 août 2017. Pour le fonctionnement normal, six jours ont été sélectionnés, à savoir les 19 et 20 juillet et les 2, 3, 26 et 27 août 2017.

4.4 La base de données de l'entrainement

La construction d'un modèle d'arbre de décision nécessite, comme toutes les méthodes d'intelligence artificielle, la formation d'une base de données (base de connaissances) collectant et regroupant des observations des périodes de fonctionnement sain, ainsi que des observations en présence des différents défauts (les défauts qui seront étudiés). Dans cette section, l'ensemble de données utilisé pour l'apprentissage et l'évaluation est décrit.

4.4.1 Les attributs électriques

Pour élaborer l'approche proposée, un ensemble de données a été recueilli auprès du SPVCR au cours de son exploitation par un système de surveillance. Le processus consiste à enregistrer un jeu de données contenant à la fois les conditions météorologiques (température ambiante et irradiation) et les paramètres électriques (Idc, Vdc et Pdc). Sachant que les entrées du système photovoltaïque ont une variation aléatoire et qu'elles sont indépendantes les unes des autres, un bon jeu de données est celui qui couvre le maximum de scénarios possibles tout au long du fonctionnement du système. Par conséquent, une base de données contenant 471 518 échantillons en plusieurs jours (jours de ciel nuageux et clairs) a été enregistrée et couvre une grande variation de température et d'irradiation. En outre, une tâche de prétraitement est effectuée pour nettoyer les données de mesures erronées. Par conséquent, 464 092 échantillons ont été obtenus après ce processus.

L'ensemble de données qui contient 464 092 échantillons est divisé en deux sous-ensembles, le premier sous-ensemble est collecté à l'état de fonctionnement normal (sain), le second à l'état de fonctionnement défaillant, de même que les données de défauts contiennent les ensembles de trois défauts. En outre, ces ensembles portent 129584 échantillons en fonctionnement sain, 126 639 échantillons de défaut de chaîne ouverte, 93 886 échantillons de défaut de court-circuit et 113 983 échantillons de défaut de ligne. La figure 4.7 représente l'histogramme de l'ensemble de données.

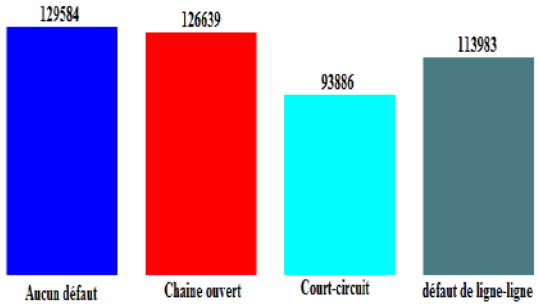


Figure 4.7 Histogramme de l'ensemble de données.

De plus, la température ambiante mesurée varie de 28 °C à 38 °C; alors que la plage de variation d'irradiation est de [200 W/m², 1000 W/m²]. Cette richesse en scénarios nous

permet d'obtenir un modèle de décision généralisé en termes d'adéquation de la prédiction avec l'état réel. Pour éviter les échantillons insignifiants dus à une erreur de mesure, une tâche de prétraitement est introduite avant le processus d'enregistrement.

La figure 4.8 montre l'évolution du Pdc en fonction du Vdc pour un fonctionnement normal et en présence de défauts dans plusieurs conditions météorologiques. Il est visible que le jeu de données présente des zones de chevauchement, ce qui augmente la complexité de la classification des différents états. Par conséquent, d'autres attributs sont nécessaires pour que l'approche proposée puisse distinguer et prédire correctement les états du système. Ainsi, trois attributs ont été utilisés, à savoir: la température, l'irradiation et le rapport de puissance.

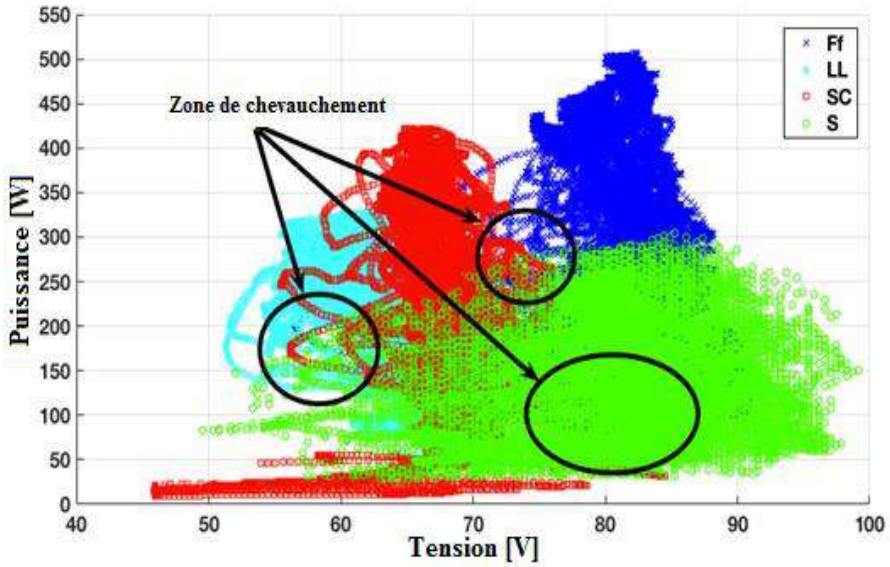


Figure 4.8 Les caractéristiques P-V au MPP dans diverses conditions météorologiques.

4.4.2 Les attributs sélectionnés

Dans l'approche proposée, deux autres attributs sont utilisés, en l'occurrence la température et l'irradiation, car ces deux attributs ont une relation profonde avec les performances du système. En d'autres termes, la température a un effet sur la tension du système PV et les pertes thermiques, de même que l'irradiation sont directement liées à l'intensité du courant généré [**104**]. En outre, pour créer un bon modèle de classification, il est préférable de choisir des attributs indépendants les uns des autres afin de fournir plus d'informations sur le système et d'avoir une corrélation élevée avec la cible. Par

conséquent, la raison pour laquelle la température ambiante a été attribuée à la température de la cellule est due aux informations données par cette dernière, et qui dépendent déjà de l'irradiation, telle qu'elle est exprimée dans l'équation (4.6). En effet, les attributs météorologiques sont insuffisants pour permettre de classer les défauts. À cette fin, un autre attribut doit être ajouté à l'ensemble d'apprentissage. Entre autre, un attribut appelé rapport de puissance est introduit. Par ailleurs, le rapport de puissance a une forte corrélation avec les performances du système, de même que les défauts affectent la puissance qui engendre une perte de l'énergie de production. Il est également pris en compte pour augmenter la précision de la classification. Cet attribut est calculé à partir de la puissance estimée par le modèle SANDIA et celle mesurée, tandis que la puissance prédite par le modèle SANDIA est considérée comme la puissance produite par le SPVCR en fonctionnement sain. La figure 4.9 représente un tracé 3D de ces attributs.

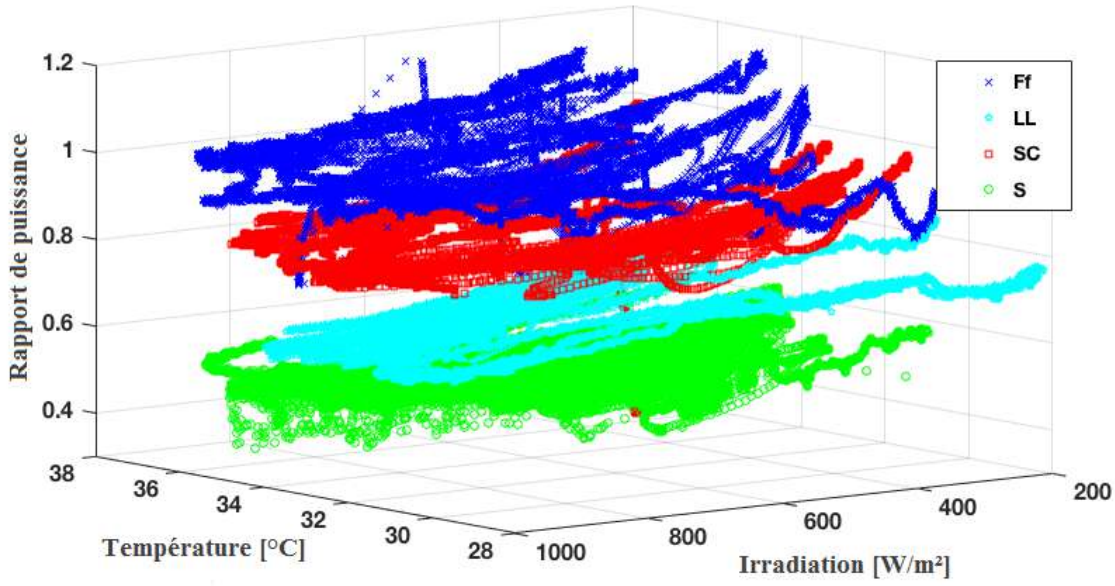


Figure 4.9 Tracé 3D du modèle d'entrées DT proposé.

À partir de la figure 4.9, chaque instance de l'ensemble d'apprentissage est composée de trois attributs et d'une étiquette de classe, d'où la classe représente l'état du système soit sain soit défectueux pour le modèle de détection. Ou bien qu'elle soit un système sain, défaut de chaîne ouverte, défaut de court-circuit ou défaut ligne de ligne pour le modèle de diagnostic. Le but est de construire deux modèles capables de prédire avec une grande précision l'état du système et la classification du type de défaut s'il existe. Par conséquent,

basés sur ces données, deux modèles non-paramétriques sont construits à l'aide de l'algorithme d'arbre de décision C4.5.

4.5 L'Arbre de Décision (AD)

Parmi les méthodes couramment utilisées dans la littérature pour détecter et diagnostiquer les défauts dans le SPVCR, on retrouve celles qui reposent sur un modèle de prédiction en fonctionnement normal et comparé aux valeurs mesurées. Lorsqu'un écart important apparaît, un signal d'alarme se déclenche [**62,74**]. Néanmoins, ces méthodes ont montré des limites pour identifier avec précision certains défauts.

Dans le présent travail, une nouvelle approche intelligente est proposée pour détecter et classer les défauts avec une précision de prédiction élevée. L'induction par arbre de décision est l'un des algorithmes les plus utilisés pour résoudre le problème de classification, considéré comme l'un des algorithmes les plus sophistiqués dans ce domaine.

Un arbre de décision consiste à déterminer par des stratégies de diviser et conquérir (**Divide-and-Conquer Strategy**) effectuées sur l'ensemble d'apprentissage, une représentation d'une procédure de décision pour déterminer la classe d'une donnée [**107**]. Généralement, la forme du modèle possède plusieurs nœuds reliés les uns aux autres par des branches et descendant jusqu'aux nœuds terminaux dans une forme pyramidale. Par conséquent, une voie du nœud supérieur au nœud terminal correspond à une série d'attributs (ou questions) avec leurs valeurs (réponses) [**108**]. La structure de l'arbre de décision peut être représentée par des règles traduites par « *si – alors* », ce qui permet une prédiction en suivant l'un de ces chemins. De plus, les règles de décision sont très similaires à celles naturellement manipulées par les humains.

4.5.1 Notions de base

L'apprentissage inductif consiste à extraire un modèle à partir d'un ensemble d'exemples appelé ensemble d'apprentissage **[109]**. Chaque exemple de cet ensemble est représenté par un vecteur d'attributs et chaque attribut prend un ensemble de valeurs (Voir la figure 4.10).

Si la classe de chaque vecteur est donnée, on se situe dans le cadre de l'apprentissage supervisé. On cherche à utiliser les exemples fournis déjà classés pour apprendre au modèle comment déterminer la classe de tout nouvel exemple rencontré.

Si aucune classe n'est disponible, on se situe dans le cadre de l'apprentissage non-supervisé qui cherche à regrouper les exemples (clustering) en mesurant la similarité entre eux **[110].**

L'apprentissage par arbre de décision s'inscrit dans le cadre de l'apprentissage supervisé, où la classe de chaque exemple dans la base est donnée. Le but est de construire un modèle à partir d'un ensemble d'exemples associés aux classes. Une fois ce modèle construit, on peut extraire un ensemble de règles de classement. Ce modèle ou les règles extraites sont ensuite utilisées pour classer de nouveaux exemples dont la classe est inconnue (Voir la figure 4.10).

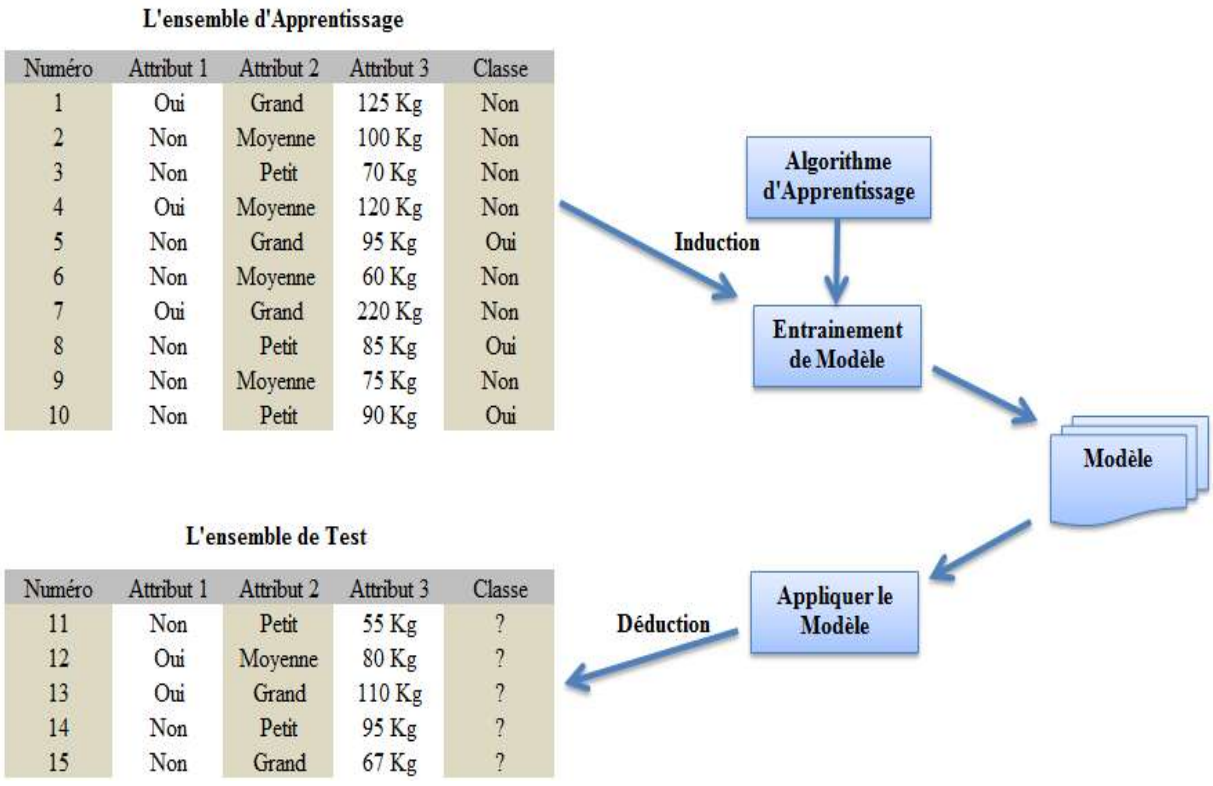


Figure 4.10 L'induction par l'arbre de décision [**110**].

4.5.2 Types de données

Nous pouvons diviser le type d'un attribut en deux grandes catégories à savoir **[20]:**

- **Quantitative** (Numérique) : Si l'ensemble des valeurs qu'il peut prendre est un ensemble de nombres, fini ou infini, ou un intervalle des valeurs réelles. Un attribut « *X* » numérique peut être discret ou continu selon sa nature :
    - **Continu** : Si l'ensemble des valeurs qu'il peut prendre est un intervalle réel, il s'agit donc d'un ensemble infini non dénombrable: on ne peut pas énumérer systématiquement l'ensemble de tous les points d'un intervalle. Par exemple, « *X* » peut être l'âge d'une personne prise au hasard, sa taille, son poids, …etc.
    - **Discret** : Si l'ensemble des valeurs qu'il peut prendre est un ensemble numérique fini (comprenant un nombre fini d'éléments) ou un ensemble infini dénombrable (comprenant une infinité de nombres que l'on peut énumérer).
- **Qualitative** (nominale) : Si l'ensemble des valeurs qu'il peut prendre est non numérique, l'attribut « *X* » peut être par exemple la couleur des yeux d'une personne prise au hasard, sa région de naissance, son sexe, …etc.

D'autre part, une donnée numérique ou nominale peut être ordinale si ses valeurs sont ordonnées. Par exemple, l'attribut dont les valeurs sont {bien, très-bien, excellente} est un attribut ordinal nominal ; l'attribut dont les valeurs sont {1, 2, 3, 4, 5} est un attribut ordinal numérique (discret).

De plus, si les valeurs d'un attribut discret ou nominal sont binaires, on parle d'un attribut binaire, par exemple l'attribut nominal sexe qui prend les valeurs {masculin, féminin} ou un attribut discret qui prend les valeurs {0,1}.

Dans notre travail, nous traitons les attributs nominaux et les attributs discrets. Si les attributs dans la base d'apprentissage sont continus, on applique des méthodes de discrétisation pour les rendre discrets.

4.5.3 Exemple introductif à l'arbre de décision

Si on considère un défaut (court-circuit) dans un champ PV, généralement les caractéristiques I-V changent et affectent le point de puissance maximale (MPP). Donc le MPP Normal et Défaut varient dans la courbe I-V comme illustré dans la figure 4.11. Un arbre de décision est composé d'un *nœud racine*, de *nœuds internes* et de *nœuds terminaux* ou de *feuilles*, il est utilisé pour détecter le MPP Défaut ou le MPP Normal (Voir la figure 3.12). Chaque nœud a un test (question) qui correspond à un attribut (V_{PV} et I_{PV}

correspondant à chaque MPP) et chaque feuille correspond à une classe (Normal ou Défaut).

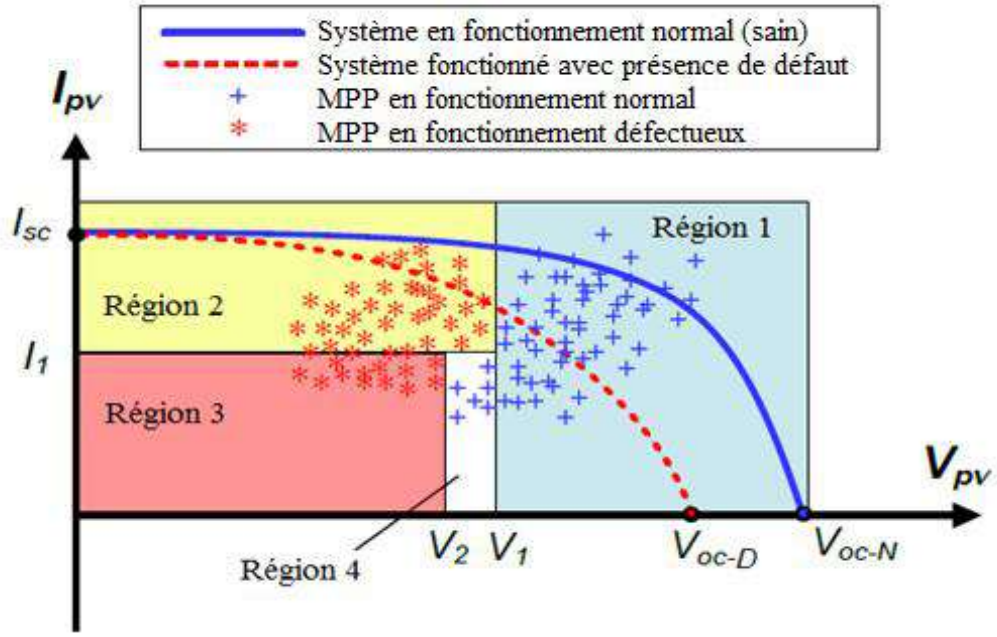


Figure 4.11 Caractéristiques I-V en fonctionnement Normal et Défaut [12].

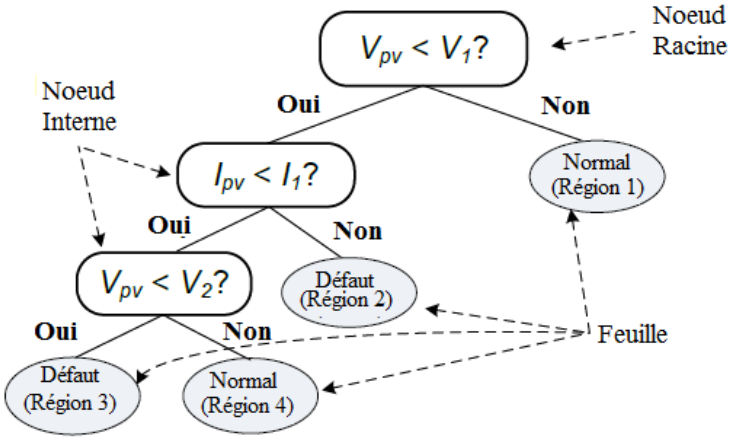


Figure 4.12 Modèle d'arbre de décision pour détection de défaut [12].

Notons que le modèle simple dans la figure 4.12 utilise deux attributs seulement (V_{PV} et I_{PV}) pour détecter le défaut. Mais dans le cas réel, le MPP Normal et défaut peuvent se chevaucher dans la courbe I-V, donc il est utile que d'autre attributs soient pris en considération (puissance, température, radiation,…etc.). Ci-après, des définitions de base relatives à l'arbre de décision.

- **Définition de base**

Un arbre de décision est une représentation d'une procédure de décision pour déterminer la classe d'un objet donné, construit à partir d'un ensemble d'apprentissage. Un chemin de la racine à un nœud correspond à une série d'attributs (questions) avec leurs valeurs (réponses). La structure arborescente d'un arbre est équivalente à un ensemble de règles « *si – alors* ». Ceci justifie une prise d'une ou de plusieurs décisions en suivant l'un de ces chemins. Cette structure de connaissances est très proche de celle manipulée naturellement par un être humain **[111]**.

4.5.4 Construction d'un arbre de décision

Le problème de la construction d'un arbre de décision peut être exprimé d'une manière récursive. Tout d'abord, sélectionnez un attribut à placer au nœud racine et créer une branche pour chaque valeur possible. Cela divise l'ensemble d'exemples en sous-ensembles, un pour chaque valeur de l'attribut. Le processus peut être répété de manière récursive pour chaque branche en utilisant uniquement les instances qui atteignent réellement la branche. Si à tout moment toutes les instances d'un nœud ont la même classification, arrêter le développement de cette partie de l'arbre. De plus, pour diviser l'ensemble d'apprentissage, on choisit des attributs qui vont minimiser l'impureté dans les sous-arbres; autrement dit, qui maximise l'information apportée par les réponses. C'est-à-dire pour chaque attribut qui n'a pas encore été utilisé, on calcule l'impureté qui reste après son utilisation. Celui qui laisse le moins de désordre est choisi comme étant le prochain nœud de l'arbre de décision. On répète le processus sur chaque nouveau nœud. Le processus s'arrête quand les feuilles de l'arbre ainsi obtenu contiennent des exemples d'une seule classe ou quand aucun test n'apporte plus d'amélioration **[20, 110, 111].**

Dans toutes les méthodes de construction d'un arbre de décision, on trouve les trois opérateurs suivants expliqués ci-après **[20]**:

- Décider si un nœud est terminal : tous les exemples (un ou plus) appartiennent à la même classe (il y a moins d'erreurs) ou (il y a moins d'un certain nombre d'erreurs).
- Sélectionner un test pour l'associer à un nœud.
- Affecter une classe à une feuille. On attribue la classe majoritaire à une feuille.

Formulation mathématique

Nous adoptons les notations suivantes :

$X = \{X_1, X_2, \dots, X_i, \dots, X_n\}$; L'ensemble des attributs, avec n le nombre d'attributs ;

$C = \{C_1, C_2, \dots, C_k\}$; L'ensemble de classes, avec k le nombre de classes ;

$|C_j|$, j = 1,2, … , k ; Le nombre des exemples qui appartiennent à la classe C_j ;

D; L'ensemble des exemples ;

$|D|$; Le nombre total des exemples.

La construction d'un arbre de décision se base sur le principe d'induction de haut en bas (*Top-Down Induction*) **[112]**. On commence par construire le nœud racine de l'arbre en continuant jusqu'à la feuille. A chaque étape un test sur le nœud courant est choisi où le choix d'un attribut test X_i se fait en mesurant l'impureté (*l'incertitude*). Une des fonctions utilisée à cette fin est L'entropie **[110].**

- **Entropie de Shannon** (*Information*) : c'est la quantité d'information nécessaire pour identifier la classe d'un élément de D, définie comme suit :

$$Info(D) = -\sum_{j=1}^{k} \frac{|C_j|}{|D|} \log_2 \left(\frac{|C_j|}{|D|} \right) \qquad (4.9)$$

S'il n'y a qu'une seule classe, l'entropie est nulle donc par convention $\log_2(0) = 0$.

Si les *k* classes sont équiprobables, l'entropie est maximale et vaut $\log_2(k)$. Pour partitionner l'ensemble d'apprentissage *D* sur la base des valeurs d'un attribut X_i en sous-ensembles D_i où *i=1,…, n* (Voir la figure 4.13), on va calculer l'information gagnante par cet attribut en utilisant l'entropie conditionnelle.

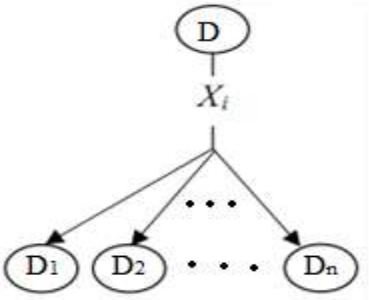


Figure 4.13 Structure de l'arbre de décision pour l'ensemble D.

- **Entropie conditionnelle** : c'est la quantité d'informations nécessaires pour identifier la classe d'un élément de *D* sachant les valeurs de l'attribut X_i, il est donné par l'expression donnée ci-dessous:

$$Info(X_i, D) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} . Info(D_i) = -\sum_{i=1}^{n} \left(\frac{|D_i|}{|D|} . \sum_{j=1}^{k} \frac{|C_j|}{|D_i|} \log_2 \left(\frac{|C_j|}{|D_i|} \right) \right) \quad (4.10)$$

- **Gain d'Entropie** pour l'attribut X_i : ceci est défini par l'expression suivante :

$$Gain(X_i, D) = Info(D) - Info(X_i, D) \quad (4.11)$$

Cela représente la différence entre l'information nécessaire pour identifier un élément de *D* et l'information nécessaire pour identifier un élément de *D* sachant que la valeur de l'attribut X_i est obtenue.

Ce critère permet de choisir l'attribut le plus important. L'attribut choisi est celui qui a le gain le plus grand par rapport aux autres attributs. Après avoir choisi un attribut test, on doit partitionner l'ensemble en sous-ensembles selon les valeurs possibles de cet attribut. On crée les branches correspondant à chaque valeur de l'attribut, on crée des nœuds pour chaque sous-ensemble non vide et on répète le processus pour chaque nouveau nœud jusqu'à ce qu'on arrive à un critère d'arrêt. Dans ce cas, on crée une feuille et on associe à cette feuille la classe la plus probable **[110].**

L'algorithme de construction d'un arbre de décision est résumé ci-après **[107]:**

```
Construire-arbre (D, X): (TDIDT)
Entrées :
D: un ensemble d'apprentissage.
X: liste d'attributs.
Sortie: un arbre de décision.
Si un critère d'arrêt est vérifié
Créer une feuille avec la classe la plus probable.
Sinon:
```

```
Choisir le meilleur attribut A dans X (Qui maximise le gain
d'information).
Partitionner D en n sous-ensembles selon les valeurs de A.
Créer un nœud pour A.
Pour toutes les valeurs de A : construire-arbre (Di, X-A)
Fin si.
```

L'ensemble de données citées dans le tableau 4.2 est utilisé pour expliquer le processus de construction de l'algorithme l'arbre de décision décrit plus haut **[112]**. Ces données pour jouer le tennis selon les conditions climatiques qui sont prend comme des données type pour illustration le déroulement de l'algorithme d'arbre de décision ID3 et sont appelées données météorologiques **[112]**.

Tableau4.2. Données météorologiques pour jouer le tennis [**112**].

Temps	**Température**	**Humidité**	**Vent**	**Jouer**
Ensoleillé	Elevée	Haute	Faux	Non
Ensoleillé	Elevée	Haute	Vrai	Non
Couvert	Elevée	Haute	Faux	Oui
Pluvieux	Moyenne	Haute	Faux	Oui
Pluvieux	Basse	Normale	Faux	Oui
Pluvieux	Basse	Normale	Vrai	Non
Couvert	Basse	Normale	Vrai	Oui
Ensoleillé	Moyenne	Haute	Faux	Non
Ensoleillé	Basse	Normale	Faux	Oui
Pluvieux	Moyenne	Normale	Faux	Oui
Ensoleillé	Moyenne	Normale	Vrai	Oui
Couvert	Moyenne	Haute	Vrai	Oui
Couvert	Elevée	Normale	Faux	Oui
Pluvieux	Moyenne	Haute	Vrai	Non

La première chose qu'il faut savoir est comment déterminer sur quel attribut se fendre, étant donné un ensemble d'exemples avec différentes classes. Il existe quatre possibilités pour chaque division, et au niveau supérieur, ils produisent les arbres de la figure 4.14 [**112**].

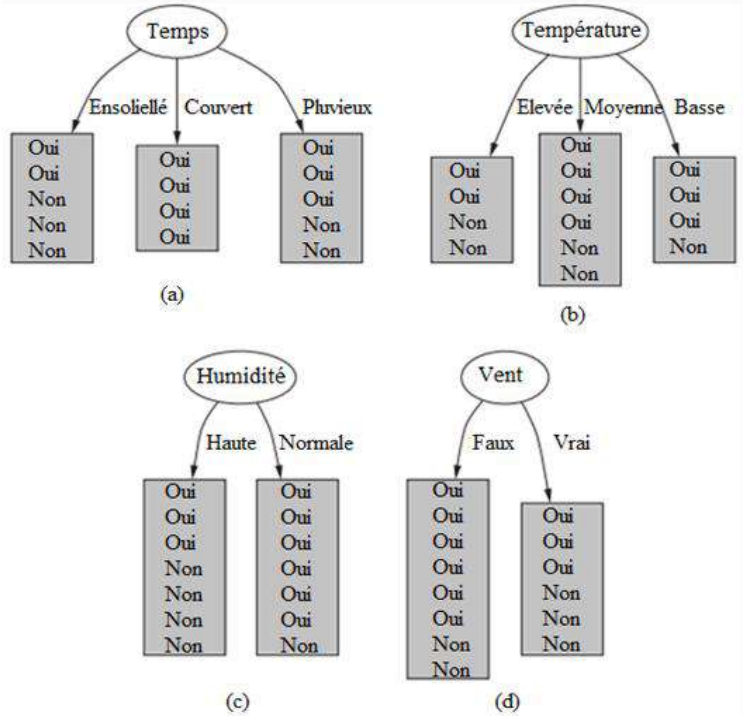


Figure 4.14 Les sous-ensembles associés aux valeurs des quatre attributs [**112**].

On va chercher l'attribut le plus pertinent qui maximise le gain d'information (Equation 4.11): Pour l'attribut Temps, le nombre de Oui et Non selon les valeurs : Ensoleillé, Couvert et Pluvieux sont respectivement [2, 3], [4,0], et [3, 2] (figure 4.14 (a)), l'entropie correspondante est :

$$Info\ ([2,3]) = \left(\frac{2}{5}\right) * log_2\left(\frac{2}{5}\right) + \left(\frac{3}{5}\right) * log_2\left(\frac{3}{5}\right) = 0.971\ bits.$$

$$Info\ ([4,0]) = \left(\frac{4}{4}\right) * log_2\left(\frac{4}{4}\right) + \left(\frac{0}{4}\right) * log_2\left(\frac{0}{4}\right) = 0.0\ bits.$$

$$Info\ ([3,2]) = \left(\frac{3}{5}\right) * log_2\left(\frac{3}{5}\right) + \left(\frac{2}{5}\right) * log_2\left(\frac{2}{5}\right) = 0.971\ bits.$$

L'entropie conditionnelle sera :

$$Info\ ([2,3],[4,0],[3,2]) = (\tfrac{5}{14}) * 0.971 + (\tfrac{4}{14}) * 0 + (\tfrac{5}{14}) * 0.971$$

$$= 0.693 \text{ bits.}$$

La base d'apprentissage contient neuf *Oui* et cinq *Non* dont l'entropie est :

$$Info(D) = Info([9,5]) = 0.940\ bits.$$

Info(D) correspond à l'entropie de l'ensemble d'exemples.

Finalement, le gain d'information apportée par l'attribut Température est :

$$\begin{aligned} \text{Gain(Temps)} &= \text{Info(D)} - \text{Info (Temps, D)} \\ &= \text{Info ([9, 5])} - \text{Info ([2, 3], [4, 0], [3, 2])} \\ &= 0.940\text{-}0.693 = 0.247 \text{ bits.} \end{aligned}$$

On refait la même chose avec les autres attributs, et on obtient :

Gain(Temps) = 0.247 bits.

Gain(Température) =0.029 bits.

Gain(Humidité) = 0.152 bits.

Gain(Vent) = 0.048 bits.

L'attribut qui maximise le gain d'information est ***Temps***. Donc, Temps est la racine de l'Arbre de Décision comme le montre la figure ci-dessous (figure 4.15.)

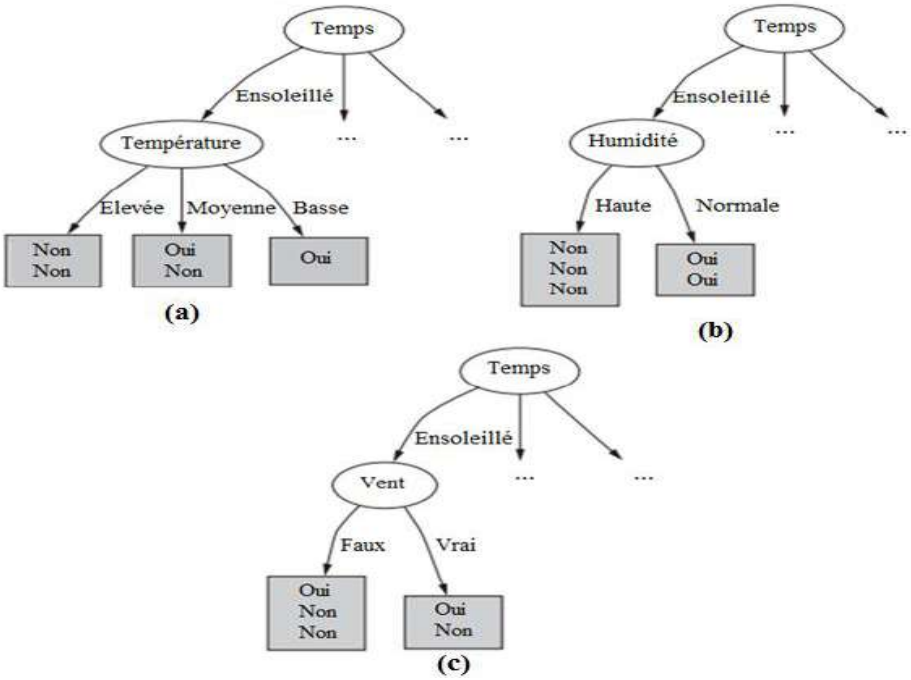


Figure 4.15 Les sous-ensembles correspondant à la valeur ensoleillée de l'attribut Temps [**112**].

Pour continuer la construction de l'arbre, on recommence le processus en choisissant d'autres attributs pertinents pour chaque sous-ensemble d'apprentissage (Voir la figure 4.15). L'attribut le plus pertinent est *Humidité*. Donc, le sous-ensemble est partitionné en deux sous-ensembles selon les valeurs haute et normale de l'humidité. Dans le sous-ensemble correspondant à la valeur *pluvieux* de l'attribut *Temps*, c'est l'attribut *Vent* qui maximise le gain, et le sous-ensemble est partitionné en deux sous-ensembles selon les valeurs *vrai* et *faux* du *Vent*.

L'arbre de décision final obtenu est donné par la figure 4.16.

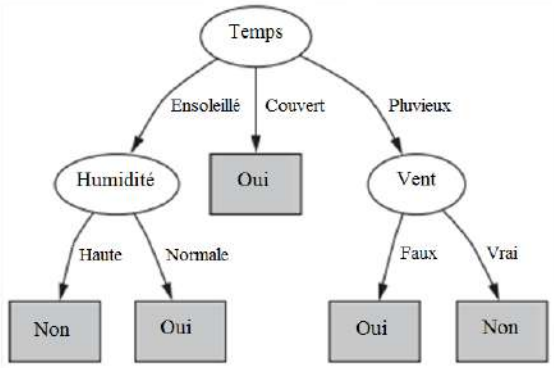


Figure 4.16. L'arbre de décision final pour la base météo [**112**].

4.5.5 Choix de l'algorithme

L'algorithme décrit précédemment n'est pas fonctionnel pour les raisons suivantes:

- Ne prend pas en considération les attributs numériques;
- Les valeurs manquantes (inconnus) dans les attributs ne sont pas prises en compte;
- Ne traite pas l'étape d'élagage ;
- Ne traite pas les données de dimensions élevées.

4.5.6 L'algorithme C4.5

L'algorithme C4.5 est un algorithme proposé par Quinlan en 1993 qui est l'extension de l'algorithme de base d'ID3 **[113]**. Elle est l'un des algorithmes d'induction d'AD les plus connus et les plus utilisés aujourd'hui **[114-116]**. Elle comporte plusieurs améliorations parmi lesquelles, on peut citer :

- La construction d'AD selon la même méthodologie expliquée dans l'algorithme précédent.

- L'élagage se base sur une estimation de l'erreur réelle à partir de l'ensemble d'apprentissage.
- Utilisation du ***rapport de gain*** au lieu du ***gain d'entropie***, défini comme suit :

$$Rapport_Gain(X_i, D) = \frac{Gain(X_i, D)}{Split_Info(X_i, D)} \quad (4.12)$$

Avec

$$Split_Info(X_i, D) = -\sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \left(\frac{|D_i|}{|D|} \right) \quad (4.13)$$

- Les valeurs numériques se traitent comme suit : (soit X_i un attribut numérique)
- Toutes les valeurs de cet attribut sont classées dans un ordre croissant.
- pour chacune des valeurs S_j on partitionne les enregistrements en deux catégories:
  - ceux qui ont une valeur inférieure ou égale à S_j.
  - ceux qui ont une valeur supérieure à S_j.

Pour chacune de ces partitions, on calcule le rapport de gain, et on choisit la partition qui maximise ce dernier.

4.5.7 Critères d'arrêt de construction d'un AD

Il y a plusieurs critères d'arrêt possible lors de la construction d'un arbre de décision. Ils sont énumérés ci-dessous:

- Tous les exemples appartiennent à la même classe ***ou*** il n'y a plus d'attribut à utiliser.
- La profondeur de l'arbre atteint une limite fixée.
- Le nombre de feuilles atteint un maximum fixé.
- Le nombre d'instances par nœud est inférieur à un seuil fixé.
- Le gain d'information maximum obtenu est inférieur à un seuil fixé.
- La qualité de l'arbre n'augmente plus de façon sensible. Par exemple, aucun attribut n'améliore la qualité de l'arbre.

Le fait de construire l'arbre jusqu'au bout selon le premier critère ci-dessus, jusqu'à ce qu'aucun attribut n'est utilisé ou tous les exemples appartiennent à la même classe, favorise

les feuilles ayant peu d'objets et diminue la fiabilité de l'arbre lors du classement d'un nouvel exemple. Par conséquent, le taux d'erreurs de classement augmente. Le premier critère n'aide donc pas à construire l'arbre de décision le plus simple. Les autres critères aident à stopper la construction lorsqu'un seuil est atteint, comme le nombre d'instances par feuille, la profondeur de l'arbre, le nombre de feuilles dans l'arbre, etc…. Ces critères se situent dans le cadre de l'*élagage* et plus précisément dans son premier type appelé *pré-élagage.*

4.5.8 Élagage

Le sur-apprentissage est en général provoqué par un mauvais dimensionnement de l'arbre de décision en raison de :

- Le bruit présent dans les données d'apprentissage ;
- Les données ne contiennent pas certains exemples importants, ou lorsque les exemples sont trop spécifiques.

Pour limiter, ce genre de problèmes on utilise l'élagage qui consiste à réduire la taille de l'arbre par suppression de quelques branches **[117]**.

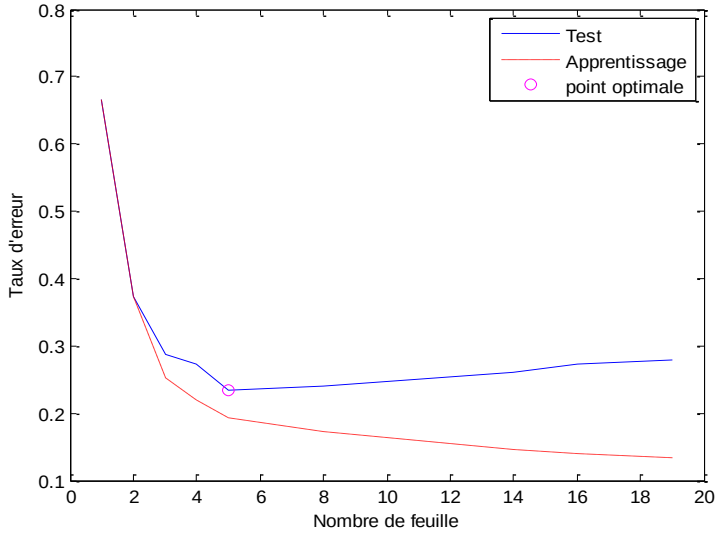
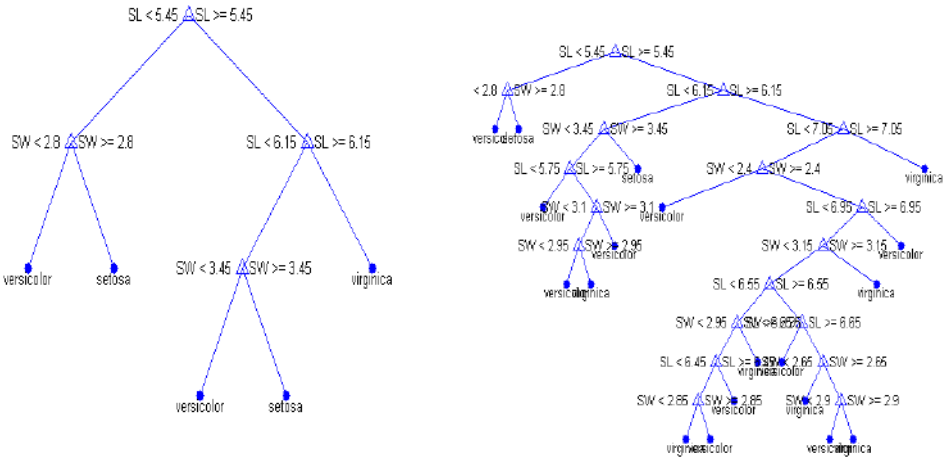


Figure 4.17 Evolution de taux d'erreur en apprentissage et en test [**118**].

Dans la figure 4.17 ci-dessus le graphe en rouge, représente l'erreur sur l'ensemble d'apprentissage. En bleu, l'erreur sur l'ensemble de test.

Si l'erreur de test augmente alors que l'erreur d'apprentissage continue à diminuer, il y a un sur-apprentissage.

Dans le point optimal le commencement de sur-apprentissage apparait, et on commence à élaguer depuis le repère indiqué sur le graphe (point optimal) (figure 4.17). Pour cela l'élagage se fait au niveau de 5 feuilles. Les deux structures de l'arbre avec et sans élagage sont montrées sur la figure 4.18.



(a) Arbre de décision à 5 feuilles après élagage **(b)** Arbre de décision sans élagage à 19 feuilles

Figure 4.18 Arbres de décision avec et sans élagage [118].

4.5.8.1 Pré-élagage

Le pré-élagage a pour but d'arrêter la construction de l'AD à l'avance même si les feuilles ne sont pas pures ; c'est-à-dire qu'on décide ou non de continuer à développer un certain nœud en utilisant par exemple les critères d'arrêts **[114]**.

Il existe une autre technique proposée par Quinlan et appelée « ***Chi-square pruning*** », citée dans **[114, 119, 120]**.

4.5.8.2 Post-élagage

Dans cette approche, l'arbre de décision est simplifié en supprimant un ou plusieurs de ses sous-arbres et en les remplaçant par des feuilles. Il existe plusieurs méthodes. Nous détaillerons la méthode implémentée en C4.5 :

- **Error-based pruning « EBP»**

L'idée de cette méthode est la suivante: si N_T est le nombre total d'instances parvenues à un nœud ou une feuille et E le nombre d'objets mal classés sur cette feuille, alors le taux d'erreur optimiste observé est $f = E/N_T$ (observer E événements dans N_T essais) **[114]**. Quinlan a utilisé l'intervalle confiance *cf* de la distribution Binomiale pour estimer le taux d'erreur prédit *e*. Ensuite par l'approximation de la distribution Binomiale en distribution Normale Centré Réduite pour trouver le seuil de confiance Z. L'expression pour estimer la valeur de *e* est :

$$e\left(N_T, f, \alpha\right) = \frac{f + \frac{Z_{\alpha/2}^2}{2N_T} + Z_{\alpha/2}\sqrt{\frac{f}{N_T} - \frac{f^2}{N_T} + \frac{Z_{\alpha/2}^2}{4N_T^2}}}{1 + \frac{Z_{\alpha/2}^2}{N_T}} \qquad (4.14)$$

Avec α est le niveau de confiance. Si $\alpha = 0{,}25$ la valeur de $Z_{\alpha/2} = 1.15$ (**voir le tableau 6 sur l'annexe A**). Pour décider si on élague un sous-arbre ou pas on doit d'abord :

- Calculer la probabilité d'erreur prédite sur chaque feuille de ce sous-arbre $e(N_j, f_j, \alpha)$; d'où j est le $j^{\text{ème}}$ feuille;
- Calculer le nombre d'erreurs prédites sur chaque feuille qui est égal au nombre de cas sur cette feuille × le taux d'erreur prédit : $N_j \times e(N_j, f_j, \alpha)$;
- Calculer le nombre total d'erreurs prédites sur toutes les feuilles du sous-arbre (la somme de toutes les erreurs des feuilles);$\sum_j N_j \times e(N_j, f_j, \alpha)$
- Calculer le nombre d'erreurs prédites sur le nœud père de ces feuilles ; qui est le nombre de cas sur ce nœud multiplié par le taux d'erreur prédit sur ce nœud; $N_T \times e(N_T, f_T, \alpha)$
- Comparer les valeurs obtenues lors des deux étapes précédentes: si l'erreur donnée par le nœud père est inférieure à l'erreur donnée par les feuilles, on élague le sous-arbre de ce nœud père.

Ce processus est répété d'une manière itérative de bas en haut (bottom up). La description de l'algorithme d'élagage est le suivant :

```
Tant qu'il existe un sous-arbre que l'on peut remplacer par une
feuille sans faire croître l'erreur réelle estimée
```

Alors élaguer ce sous-arbre.

Dans l'exemple ci-dessous, on considère un arbre de décision sans élagage, et on montre dans ce qui suit comment peut-on élaguer le sous-arbre.

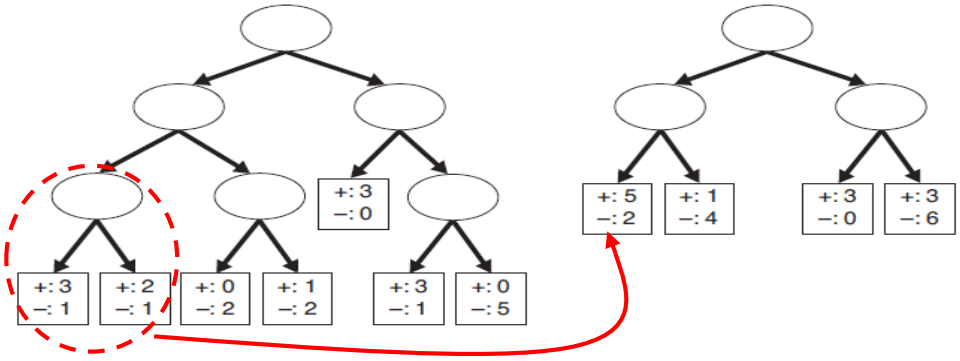


Figure 4.19 Elagage d'un sous-arbre [**110**].

On fait une étude illustrative sur le sous-arbre de la figure 3.15, pour la feuille gauche $N_T = 4$ et $E = 1$ avec un intervalle de confiance $cf = 25\%$ donc le taux d'erreur estimé est le suivant :

$e(4;1/4;25\%) = 0{,}537$ pour la feuille droit $N_T = 3$ et $E = 1$ alors $e(3{,}1/3;25\%) = 0{,}650$, l'erreur globale de ces feuilles est $4 \times 0{,}537 + 3 \times 0{,}650 = 4{,}098$, par contre l'erreur estimée par le nœud père de cette sous-arbre est $e(7{,}2/7;25\%) = 0{,}503$ et l'erreur globale de ce nœud sera $7 \times 0{,}503 = 3{,}521$, le taux d'erreur de nœud père est inférieur à l'erreur de ces feuilles donc on élague ce sous-arbre comme démontré sur la figure 4.19. On s'intéresse aux deux feuilles situées à gauche de l'AD comme détaillé dans la figure 4.19

- L'erreur prédite de la feuille gauche est : $e(4;1/4;25\%) = 0{,}537$
- L'erreur prédite de la feuille droite est : $e(3{,}1/3;25\%) = 0{,}650$
- L'erreur globale de ces feuilles est : $4 \times 0{,}537 + 3 \times 0{,}650 = \mathbf{4{,}098}$
- L'erreur prédite par le nœud père est : $e(7{,}2/7;25\%) = 0{,}503$
- Et l'erreur globale pour ce nœud père est : $7 \times 0{,}503 = \mathbf{3{,}521}$

Il est remarquable que le taux d'erreur de nœud père est inférieur à l'erreur de ces feuilles donc on élague ce sous-arbre et on le remplace par la classe majoritaire comme illustré sur la figure 4.19.

4.5.9 Évaluation du modèle

Pour évaluer un modèle, nous commençons par le tester sur plusieurs bases réelles. Ensuite, nous évaluons sa performance en utilisant la ***Matrice de Confusion***.

4.5.9.1 Matrice de Confusion

La matrice de confusion contient des informations concernant le classement actuel dans la base de test ainsi que le classement prédit par le modèle de classement utilisé **[20, 111]**. Le tableau 4.3 présente la matrice de confusion pour deux valeurs de classe (positive, négative). Ainsi,

- a est le nombre de classements corrects des instances de classe **négative**.
- b est le nombre de classements incorrects des instances de classe **négative**.
- c est le nombre de classements incorrects des instances de classe **positive**.
- d est le nombre de classements corrects des instances de classe **positive**.

Tableau 4.3 : La matrice de Confusion.

		Classe prédite	
		Négative	Positive
Classe réelle	Négative	a	b
	Positive	c	d

À partir de la matrice de confusion ci-dessus, plusieurs mesures appelées mesures d'exactitude par classe, le plus important parmi eux est défini comme suit :

- *Taux de bon classification* (exactitude - taux de bon apprentissage) : la diagonale de la matrice de confusion sur tous les éléments

$$Taux_de_bonne_classification = \frac{a+d}{a+b+c+d} \qquad (4.15)$$

4.6 Résultats expérimentaux et discussions

Dans cette section, les résultats expérimentaux sont présentés afin d'évaluer les modèles obtenus. L'ensemble de données est divisé en deux sous-ensembles choisis au hasard. Le premier est utilisé pour l'apprentissage. Cette partie contient 66% de l'ensemble de données

globales, tandis que les données restantes sont destinées au test. Un autre test a été effectué à l'aide d'un nouvel ensemble de données jamais utilisé auparavant pour permettre de valider ces deux modèles.

4.6.1. Résultats de l'entrainement

L'ensemble de données a été enregistré en état sain (état normal) et à l'état défectueux. Ce dernier contient trois défauts différents à savoir : le défaut de chaîne ouverte (S_f), le défaut de court-circuit (SC_f) et le défaut ligne (LL_f). Afin de classer l'état du système au cours de son fonctionnement, deux modèles ont été construits et basés sur l'algorithme C4.5 à l'aide du logiciel WEKA pour la détection et le diagnostic [**121**]. Bien que l'exactitude est calculée à partir du rapport entre le nombre d'instances correctement classées et le nombre total d'instances utilisées par ce modèle de prédiction tel qu'exprimée par l'équation (4.15).

Pour exécuter l'algorithme, certains paramètres doivent être initialisés, tels que l'intervalle confiance, le nombre minimum d'instances par feuille et l'option d'utilisation du processus d'élagage (activer ou désactiver).

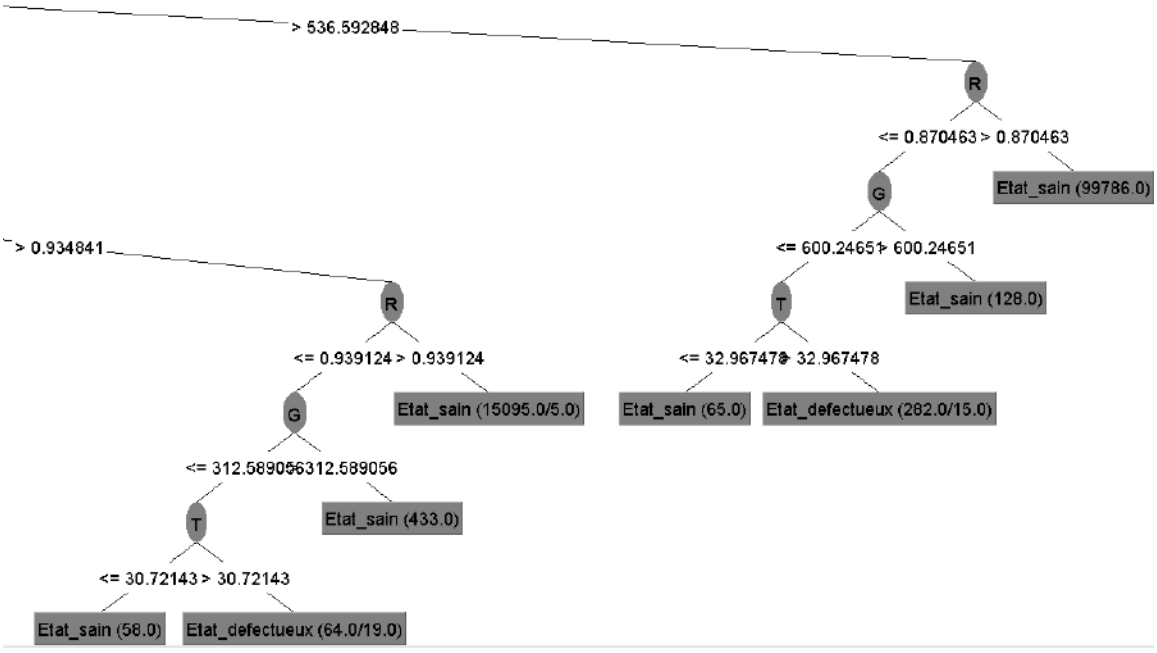


Figure 4.20 Une partie de l'arbre du modèle pour la détection de défauts.

```
|  |  |  R > 0.934841
|  |  |  |  R <= 0.939124
|  |  |  |  |  G <= 312.589056
|  |  |  |  |  |  T <= 30.72143: Etat_sain (58.0)
|  |  |  |  |  |  T > 30.72143: Etat_defectueux (64.0/19.0)
|  |  |  |  |  G > 312.589056: Etat_sain (433.0)
|  |  |  |  R > 0.939124: Etat_normal (15095.0/5.0)
|  G > 536.592848
|  |  R <= 0.870463
|  |  |  G <= 600.24651
|  |  |  |  T <= 32.967478: Etat_sain (65.0)
|  |  |  |  T > 32.967478: Etat_defectueux (282.0/15.0)
|  |  |  G > 600.24651: Etat_sain (128.0)
|  |  R > 0.870463: Etat_sain (99786.0)
```

Figure 4.21 Partie des règles du modèle AD pour la détection de défauts.

Dans l'approche proposée, les paramètres de réglage sont respectivement égales à 0.25, 40 et permet d'utiliser l'élagage. Par la suite, les figures 4.20 et 4.21 illustrent une partie de la forme d'AD avec un modèle de détection et son script de sortie associé.

La figure 4.20 montre une partie du modèle de sortie de WEKA utilisée pour la détection de défaut. Le modèle AD construit est composé de plusieurs sous-arbres qui contiennent des nœuds connectés entre eux et qui sont terminés par des feuilles; L'arbre est ramifié d'une manière binaire où le bord de chaque branche possède un seuil. En outre, le modèle impliqué dans de nombreuses décisions est celui où chaque chemin, commence à partir de la feuille et atteint le nœud racine qui peut être remplacé par une règle de décision telle qu'elle est présentée dans la figure 4.21. De plus, les chemins de décision sont différents les uns des autres. La comparaison entre chaque trajectoire a conclu qu'il existe des chemins courts, moyens et longs, et aussi les règles de décision à la même caractéristique. Par conséquent, le temps de prédiction a également une relation avec la longueur de la règle.

La figure 4.21 décrit la sortie de script pour une partie de modèle d'AD (Figure 4.20) par WEKA. La valeur entre parenthèse devant chaque étiquette d'une classe du modèle représente le nombre des instances passées par la feuille en cours, la première étant le nombre d'instances correctement prédites et le second est le nombre d'instances mal prédites.

L’AD utilisé pour le diagnostic, une partie du modèle est illustrée dans la figure 4.22. Comme mentionné précédemment, ce modèle comporte quatre étiquettes de classe qui aussi basé sur une division binaire, pour la même partie de modèle comporte les règles de script illustrées à la figure 4.23, les deux figures proviennent de la sortie du logiciel WEKA.

Les figures présentées du modèle de diagnostic montrent que le modèle ait plusieurs chemins, ainsi que chaque chemin possède une taille différente liée à la voie prise. Le modèle construit a été testé par l'ensemble de données restant. Les tableaux 4.4 et 4.5 montrent les résultats de ce processus.

Comme il ressort des tableaux 4.4 et 4.5, les modèles construits prédissent avec une grande précision les classes de données de test. L’évaluation de modèle est conclue en calculant les différents paramètres tels que : l'erreur absolue moyenne, l'erreur quadratique moyenne, l'erreur absolue relative et l'erreur quadratique moyenne relative et ainsi de suite. En outre, l'étape de construction a pris un temps de calcul relativement faible (12,64 secondes), ça montre la fiabilité d'utiliser cet algorithme pour les applications de construction en temps réel.

L'évaluation des deux modèles obtenus peut être traduite par des matrices de confusion, qui contiennent le nombre d'échantillons représentant chaque classe, soit pour la détection du défaut, soit pour la classification du type de ce défaut. Le tableau de matrice de confusion récapitulé (voir tableau (4.4)) représente le nombre d'échantillons prédits corrects et incorrects. De plus, le nombre d'échantillons classés corrects étant situé dans la diagonale de la matrice. Ainsi, le taux de bonne classification (accuracy) qui représente le rapport des instances prédites correctes données par l'équation (4.22) peut être déterminé à partir de cette matrice.

La construction du modèle AD pour la détection de défaut est basée sur un ensemble de données contenant 464092 instances enregistrées par le système d’acquisition. 66% des données sont utilisées pour l’apprentissage, les données restantes étant destinées au test.

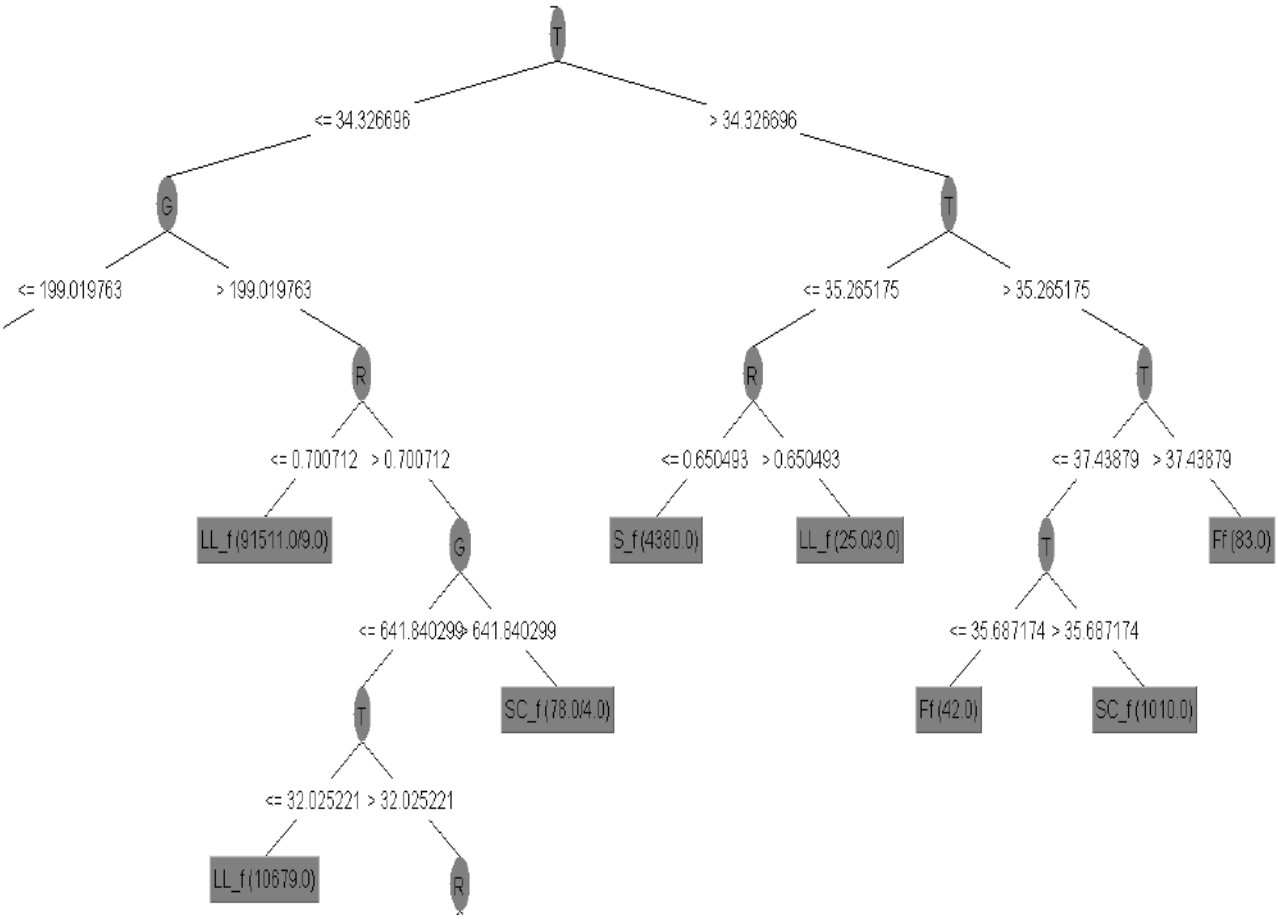


Figure 4.22 Une partie de l'arbre du modèle pour diagnostic de défaut.

Comme le montre le tableau (4.5), le temps de construction du modèle est égal à 12,64 secondes; de plus, le modèle obtenu possède 171 nœuds et 86 feuilles. La valeur du temps de construction nous permet de constater que l'algorithme C4.5 est très rapide au stade de la formation. De plus, la taille de l'arbre est acceptable et le modèle peut être lu en transférant le modèle selon les règles « *si – alors* ».

```
|  |  |  G > 199.019763
|  |  |  |  R <= 0.700712: LL_f (91511.0/9.0)
|  |  |  |  R > 0.700712
|  |  |  |  |  G <= 641.840299
|  |  |  |  |  |  T <= 32.025221: LL_f (10679.0)
|  |  |  |  |  |  T > 32.025221
|  |  |  |  |  |  |  R <= 0.73717: LL_f (6459.0/85.0)
|  |  |  |  |  |  |  R > 0.73717
|  |  |  |  |  |  |  |  T <= 33.369323: SC_f (49.0)
|  |  |  |  |  |  |  |  T > 33.369323: LL_f (1083.0/8.0)
|  |  |  |  |  G > 641.840299: SC_f (78.0/4.0)
|  |  T > 34.326696
|  |  |  T <= 35.265175
|  |  |  |  R <= 0.650493: S_f (4380.0)
|  |  |  |  R > 0.650493: LL_f (25.0/3.0)
|  |  |  T > 35.265175
|  |  |  |  T <= 37.43879
|  |  |  |  |  T <= 35.687174: free_fault (42.0)
|  |  |  |  |  T > 35.687174: SC_f (1010.0)
|  |  |  |  T > 37.43879: free_fault (83.0)
```

Figure 4.23 Une partie des règles du modèle AD pour le diagnostic de défaut.

Tableau 4.4 *Matrice de confusion du modèle de détection*

Classé comme	**a**	**b**
a = Etat_sain	**44123**	92
b = Etat_défectueux	117	**113459**

Tableau 4.5 *Résultat de test du modèle de détection*

Paramètres	**Valeur**
Temps pris pour construire le modèle	*12.64 secondes*
Temps pris pour tester le modèle	*0.25 seconde*
Taille de l'arbre	*171*
Nombre des feuilles	*86*
Instances correctement classées (157582)	***99.8675 %***
Instances mal classées (209)	*0.1325 %*
Statistique Kappa	*0.9967*
Erreur absolue moyenne	*0.0022*
Erreur quadratique moyenne	*0.0341*
Erreur absolue moyenne relative	*0.5495 %*
Erreur quadratique moyenne relative	*7.5994 %*
Nombre total des Instances	*157791*

Tableau 4.6 *Matrice de confusion du modèle de diagnostic*

Classé comme	a	b	c	d
a = Etat sain	**44117**	0	80	18
b = Défaut de chaîne ouverte	0	**42853**	0	21
c = Défaut de court-circuit	125	1	**31790**	23
d = Défaut de ligne	5	7	23	**38728**

En outre, une étape de test est nécessaire après la construction de modèle d'AD. Ce processus est effectué à l'aide de données non considérées auparavant et qui n'ont jamais été incluses dans le processus d'apprentissage afin de permettre l'évaluation des performances du modèle. Ce modèle a été testé par 34% de l'ensemble global de données qui sont choisies au hasard. La précision du modèle de détection obtenu à partir de la matrice de confusion a atteint 99,8675%, ce qui implique une bonne prédiction de l'étiquette de la classe d'instances et de la classification de l'état du système.

Tableau 4.7 *Résultat de test du modèle de diagnostic*

Paramètres	Valeur
Temps pris pour construire le modèle	*14.1 seconde*
Temps pris pour tester le modèle	*0.72 seconde*
Taille de l'arbre	*207*
Nombre des feuilles	*104*
Instances correctement classées (157488)	*99.808 %*
Instances mal classées (303)	*0.192 %*
Statistique Kappa	*0.9974*
Erreur absolue moyenne	*0.0015*
Erreur quadratique moyenne	*0.0284*
Erreur absolue moyenne relative	*0.409 %*
Erreur quadratique moyenne relative	*6.5739 %*
Nombre total des Instances	*157791*

Les résultats des tests des tableaux 4.6 et 4.7 montrent clairement que le modèle de classification offre de bonnes performances en termes de précision avec une taille d'arbre appropriée et un temps de construction rapide.

Alors que les valeurs de différents résultats auraient pu être plus intéressantes, le modèle a été construit en 14,1 secondes avec une taille de l'arbre composée de 207 nœuds et parmi ces nœuds, il y a 104 feuilles. Le modèle aussi possède une grande précision de prédiction. Par conséquent, la plupart des instances de l'essai sont correctement classées, où 157488

échantillons sont correctement classés. Par ailleurs, 303 échantillons sont faussement prédits. Ce test montre clairement l'efficacité du modèle dont la précision est de 99,808%. De plus, le temps de développement du modèle est égal à 0,72 seconde.

4.6.2 Résultats de validation

Bien que les deux modèles donnent de bons résultats dans le processus de test, tels que la précision et la taille de l'arbre, une validation reste nécessaire pour confirmer leur efficacité en termes de généralisation, de robustesse et de complexité. Les tests de validation sont effectués; tous les états du système sont pris en compte par les deux modèles pour garantir des résultats précis. De cette façon, pour la détection et le diagnostic des défauts, un jeu de données est enregistré sur cinq jours où quatre jours correspondent chacun à un état de fonctionnement et le dernier jour contient tous les états du système. Donc, le test tient compte de cinq situations de fonctionnement principales, à savoir : le jour sans défaut (système sain), un jour à un défaut de chaine ouverte, un jour de défaut de court-circuit, un jour de défaut de ligne, le dernier jour contient les quatre situations l'un après l'autre.

4.6.2.1 Jour sans défaut

En général, le SPVGC fonctionne à la majorité du temps dans un état sain et peut être soumis à l'ombrage naturel, ce qui conduit à une faible production d'énergie. Pour cette raison, les modèles obtenus doivent être testés pour ces conditions climatiques. La figure 4.24 représente la température ambiante, la température de cellule estimée, l'irradiation, le courant, la tension, la puissance prédite par le modèle SANDIA avec la puissance mesurée et le rapport entre eux respectivement pendant ce jour.

La figure 4.24.e montre les courbes de la puissance mesurée et de la puissance estimée. On peut voir que les deux courbes sont très proches ce qui confirme que pendant ce jour, le système fonctionne dans un état sain. Comme indiqué dans la section précédente, le rapport entre les deux puissances est utilisé comme un attribut qui présente clairement une forte corrélation avec l'état du système.

Les résultats obtenus à partir du test ce jour par le modèle de détection sont représentés dans les tableaux 4.8 et 4.9.

Tableau 4.8 *Matrice de confusion du modèle de détection (Jour sans défaut)*

Classé comme	a	b
a = Etat_sain	**26263**	3
b = Etat_défectueux	0	**0**

Tableau 4.9 *Résultat de test du modèle de détection*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.36 seconde*
Instances correctement classées (26263)	*99.9886 %*
Instances mal classées (3)	*0.0114 %*
Nombre total des Instances	*26266*

La matrice de confusion du tableau 4.8 montre que le modèle proposé démontre une bonne généralisation, ce qui permet de prédire les états du système avec une grande précision. Par conséquent, la précision est de 99,9886%, 26263 échantillons étant correctement prédits à partir de 26266 échantillons. De plus, seulement 0,36 seconde suffisait au modèle pour classer toutes les instances. Les mêmes données sont utilisées pour valider le modèle de diagnostic; Les tableaux 4.10 et 4.11 sont établis à partir de la prédiction du modèle de diagnostic.

Tableau 4.10 *Matrice de confusion du modèle de diagnostic*

Classé comme	a	b	c	d
a = Etat sain	**26266**	0	0	0
b = Défaut de chaîne ouverte	0	**0**	0	0
c = Défaut de court-circuit	0	0	**0**	0
d = Défaut de ligne	0	0	0	**0**

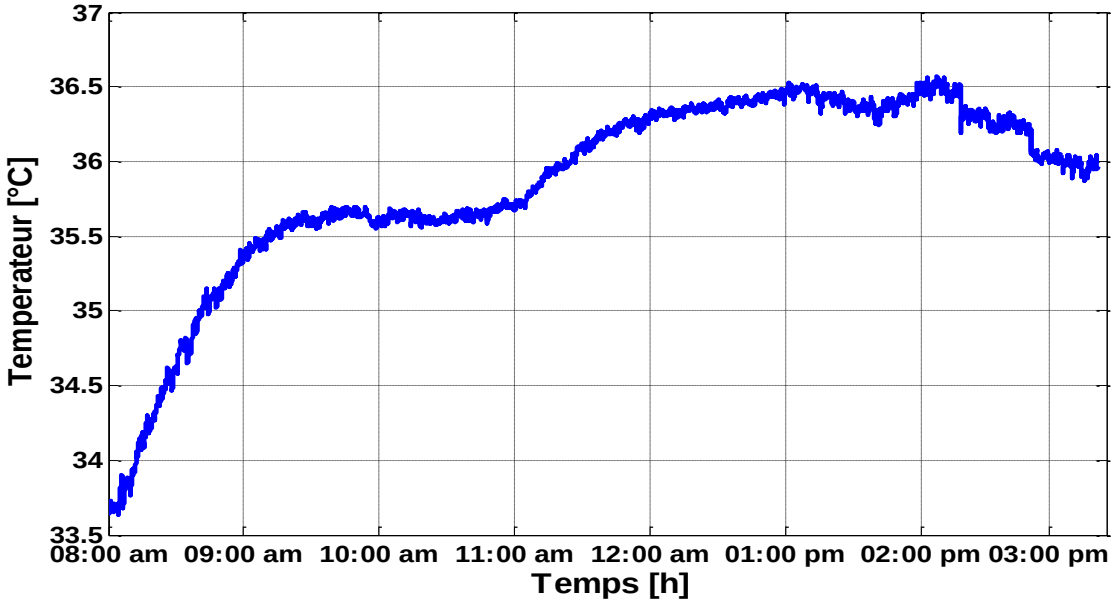
Tableau 4.11 *Résultat de test du modèle de diagnostic*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.13 seconde*
Instances correctement classées (26266)	*100 %*
Instances mal classées (0)	*0 %*
Nombre total des Instances	*26266*

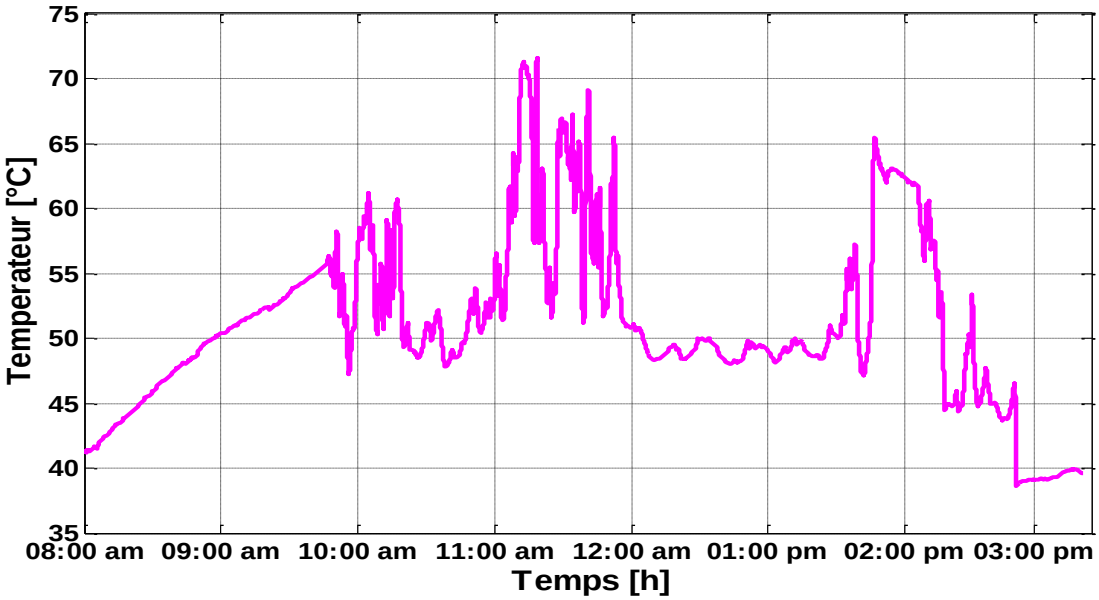
Selon les résultats présentés dans les tableaux 4.10 et 4.11, la prédiction est bien effectuée car le modèle de diagnostic prévoyait le jour en état sain avec une précision de

prédiction de 100% donc tous les échantillons sont correctement classés et prenant que 0,13 seconde d'exécution.

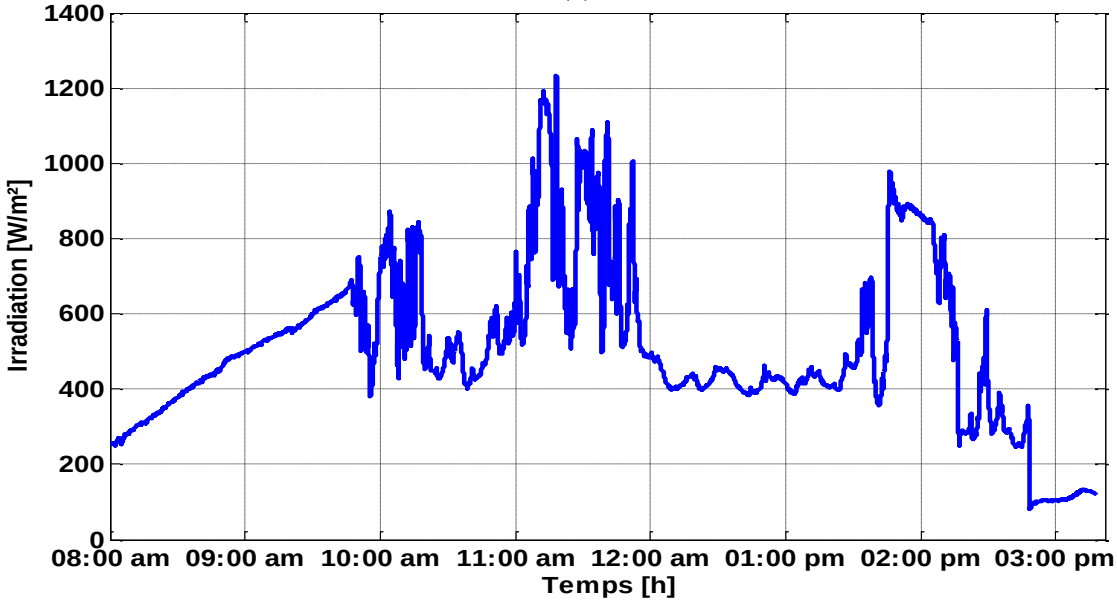
(a)



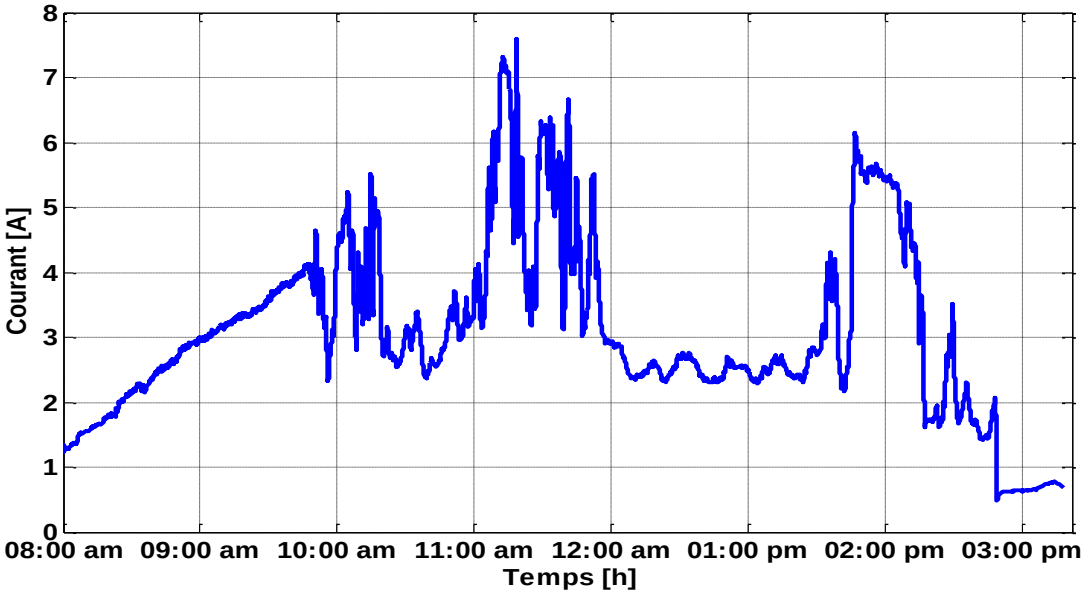
(b)



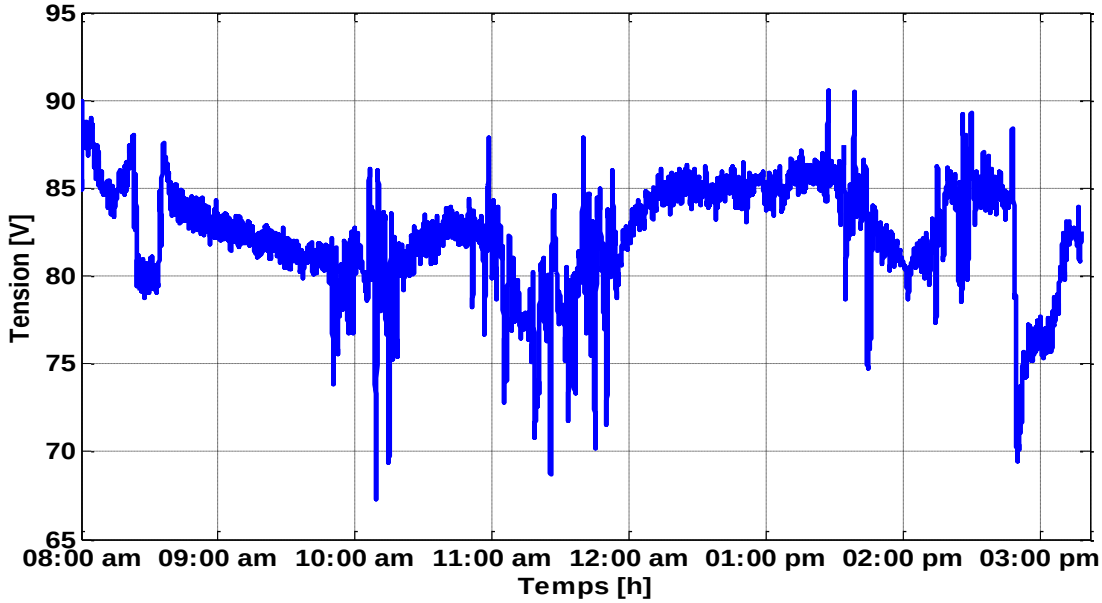
(c)


1400
1200
1000
800
600
400
200
0
Irradiation [W/m²]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

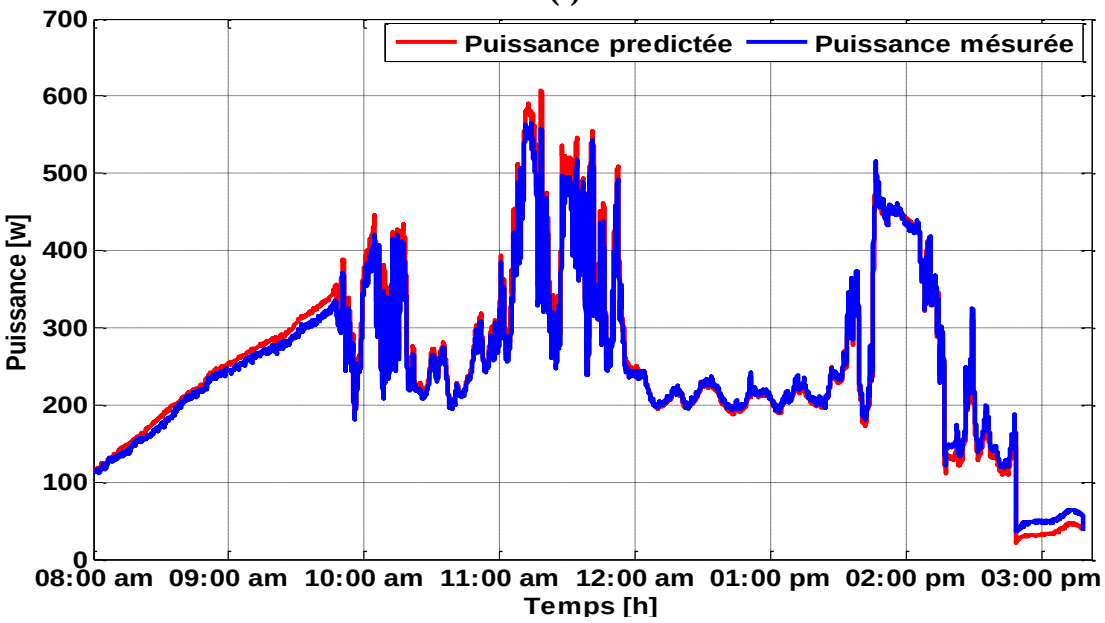
(d)


8
7
6
5
4
3
2
1
0
Courant [A]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(e)


95
90
85
80
75
70
65
Tension [V]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(f)


Puissance predictée
Puissance mésurée
700
600
500
400
300
200
100
0
Puissance [w]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

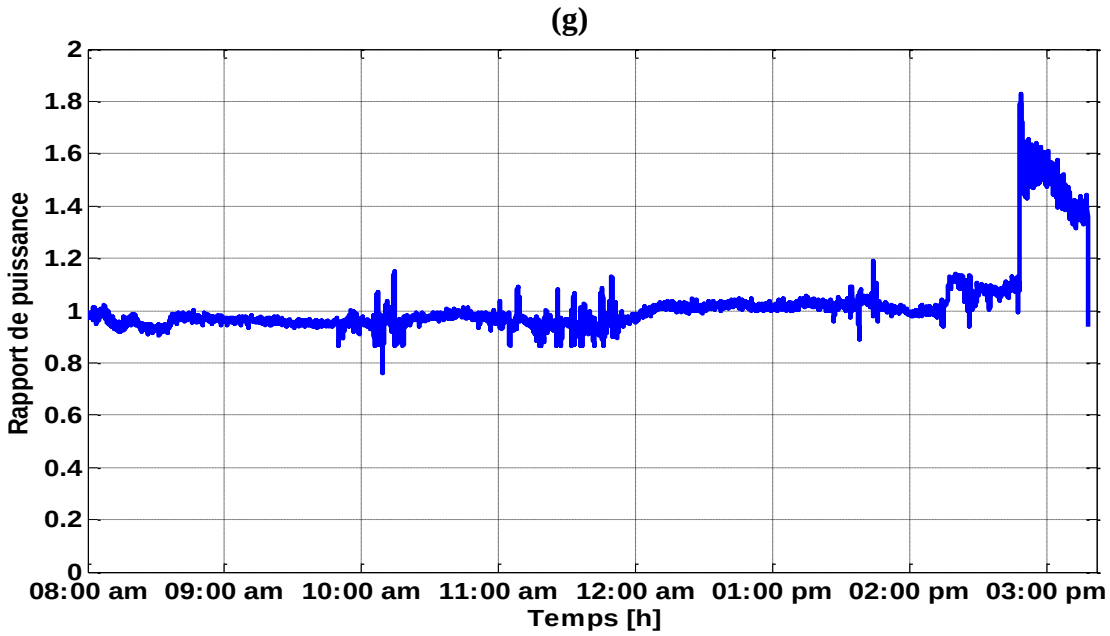


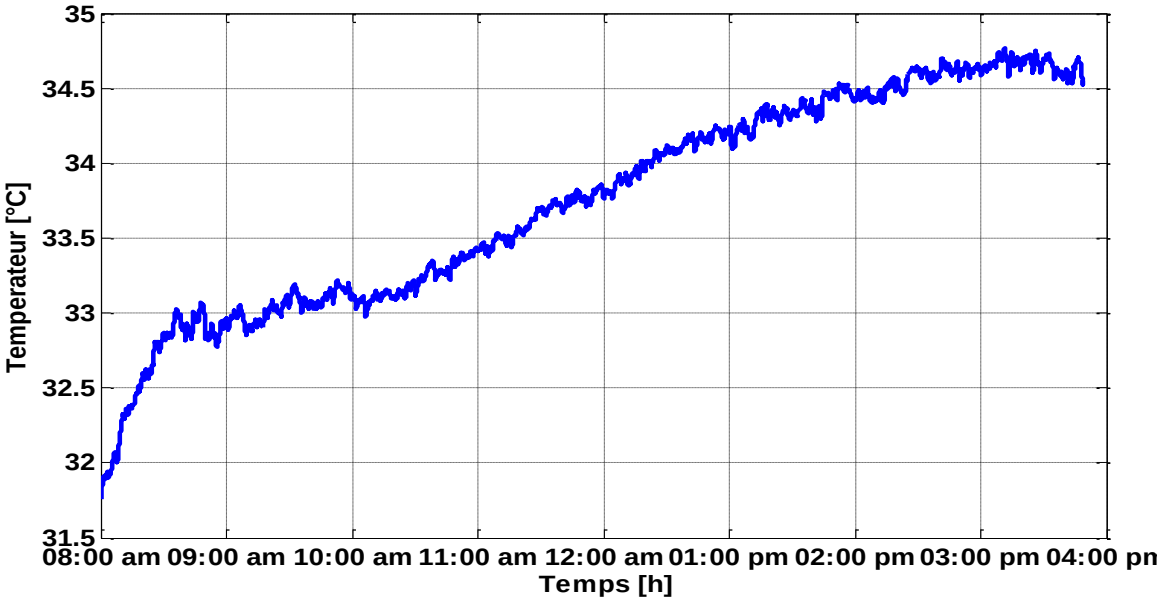
Figure 4.24 Les courbes sont : (a) la température ambiante, (b) la température estimée de la cellule, (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance.

4.6.2.2 Jour de défaut « chaîne ouverte »

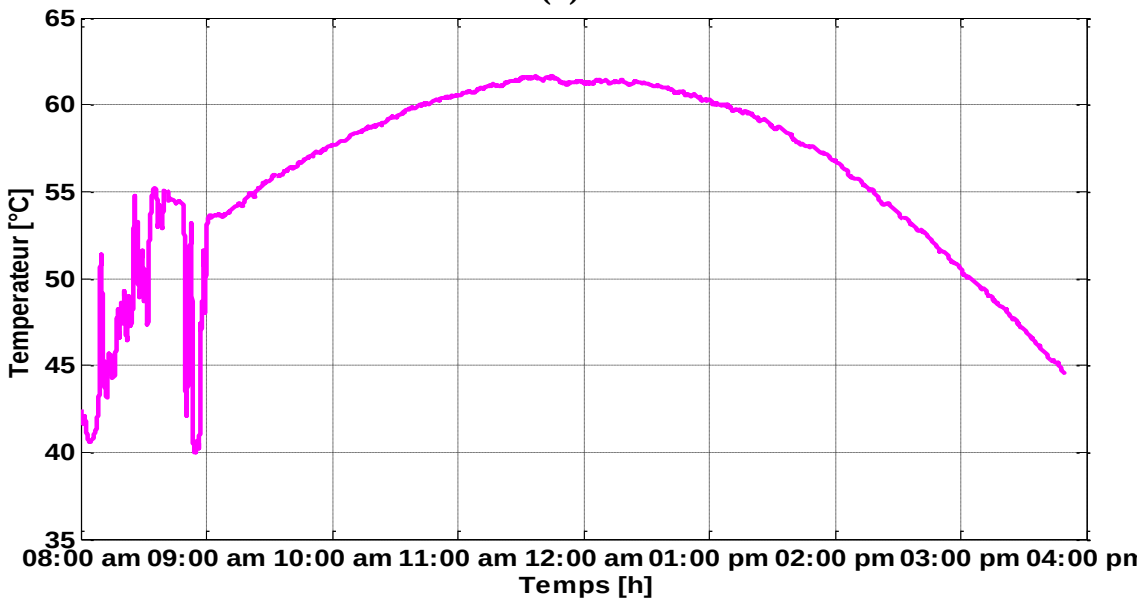
Le défaut de chaîne ouverte est créé artificiellement dans le système en déconnectant une des chaînes du champ. La figure 4.26 illustre les attributs d'entrées du modèle et les variables électriques telles que la température ambiante, la température de cellule estimée, l'irradiation, le courant, la tension et la puissance, qui sont enregistrés le jour de défaut de la chaîne ouverte avec également l'attribut de rapport de puissance.

Comme le montre la figure 4.26.e, il y a une perte de puissance importante due à une défaillance d'une chaîne. En effet, en comparant les courbes de puissance estimée et enregistrée, il apparaît que la puissance est réduite d'environ 50% lorsqu'une chaîne est déconnectée. Par conséquent, les modèles prédictifs sont utilisés pour déduire chaque classe d'instances de ces données journalières. Par la suite, les tableaux 4.12 et 4.13 détaillaient les résultats de la prévision de ce jour.

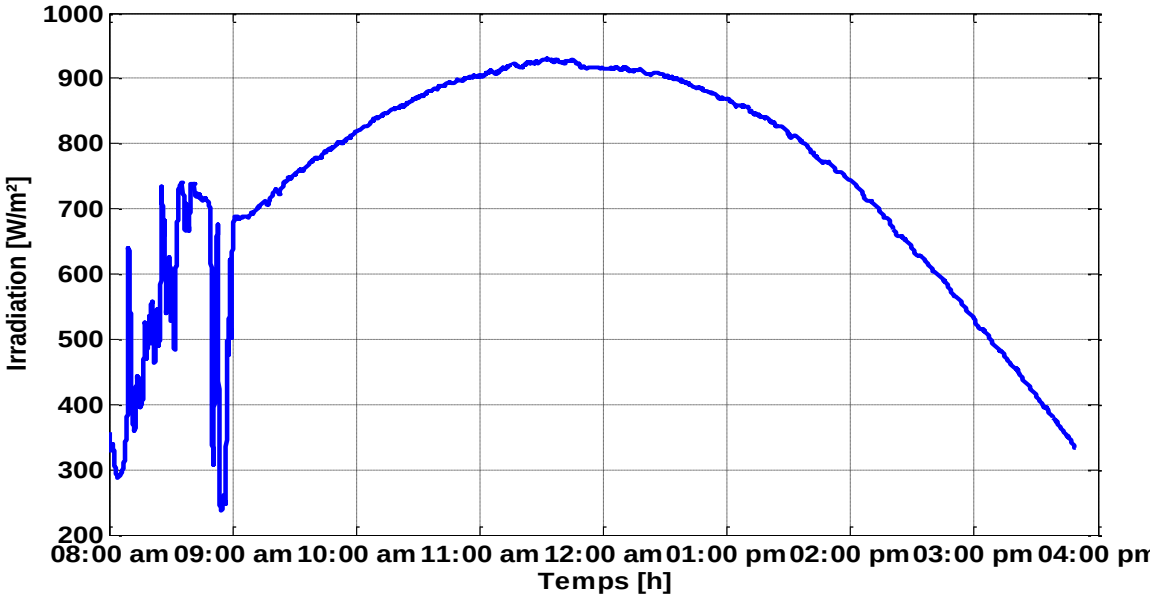
(a)


35
34.5
34
33.5
33
32.5
32
31.5
Temperateur [°C]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pn
Temps [h]

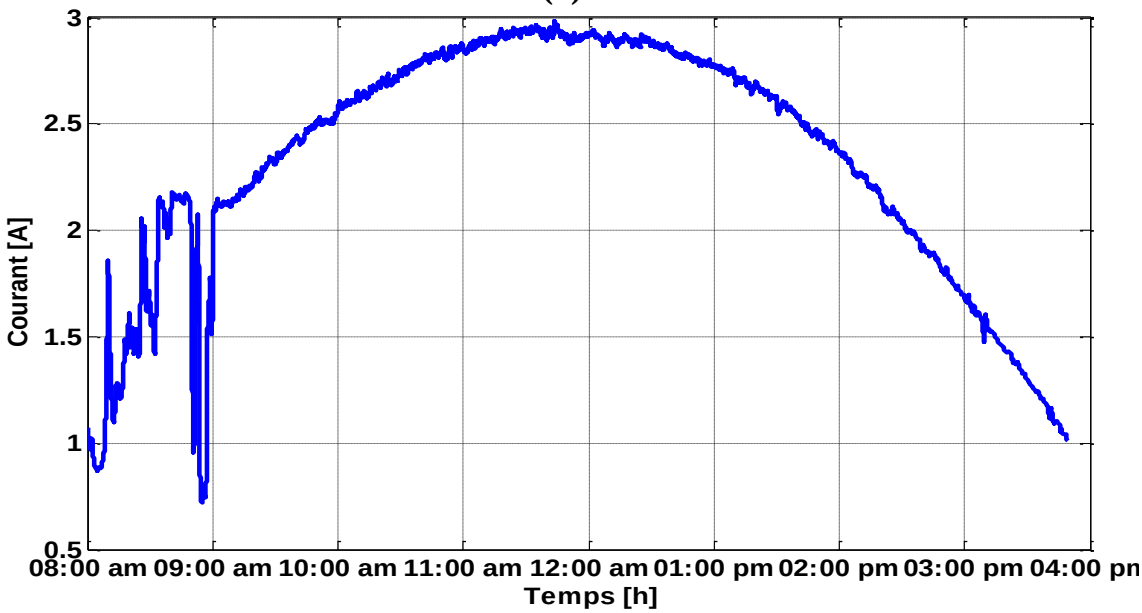
(b)


65
60
55
50
45
40
35
Temperateur [°C]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pn
Temps [h]

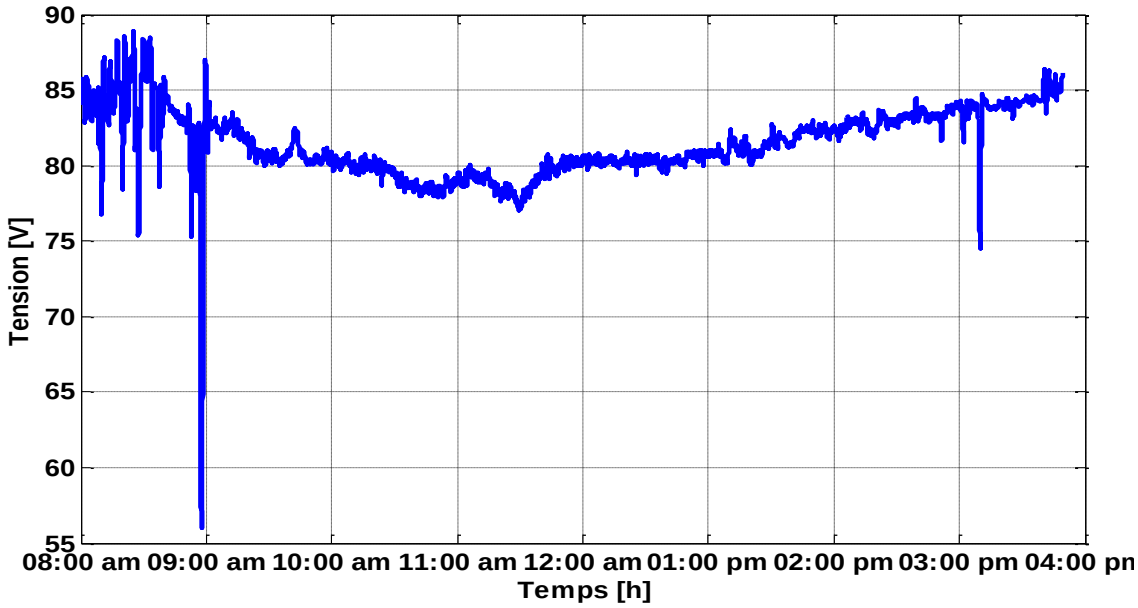
(c)


1000
900
800
700
600
500
400
300
200
Irradiation [W/m²]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
04:00 pm
Temps [h]

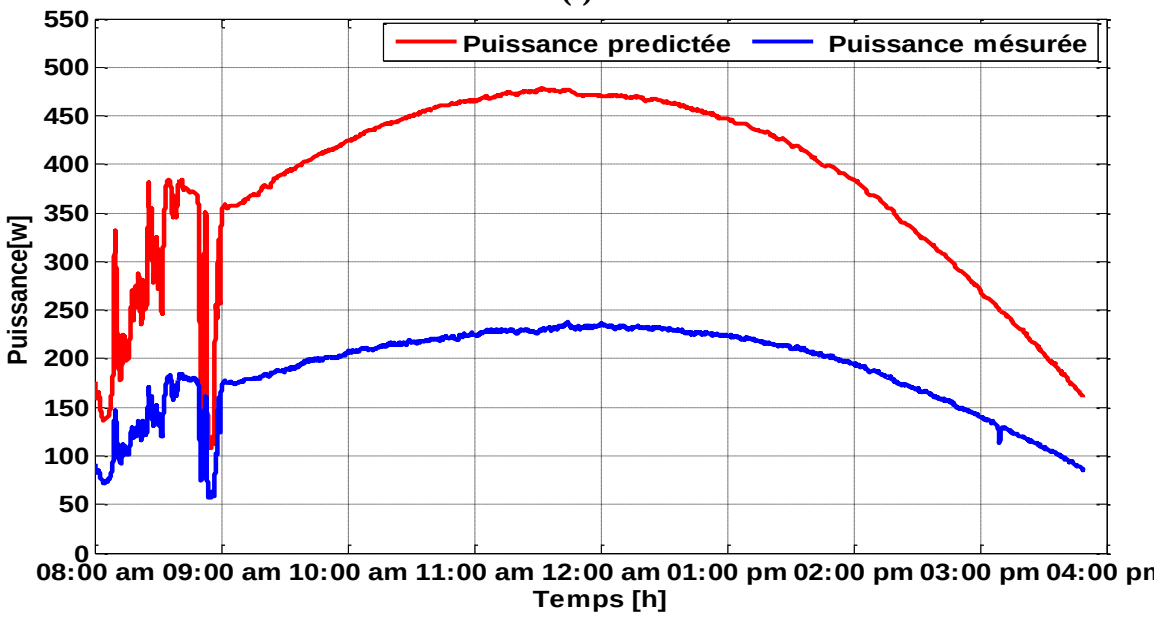
(d)


3
2.5
2
1.5
1
0.5
Courant [A]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
04:00 pm
Temps [h]

(e)


90
85
80
75
70
65
60
55
Tension [V]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

(f)


Puissance predictée
Puissance mésurée
550
500
450
400
350
300
250
200
150
100
50
0
Puissance[w]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

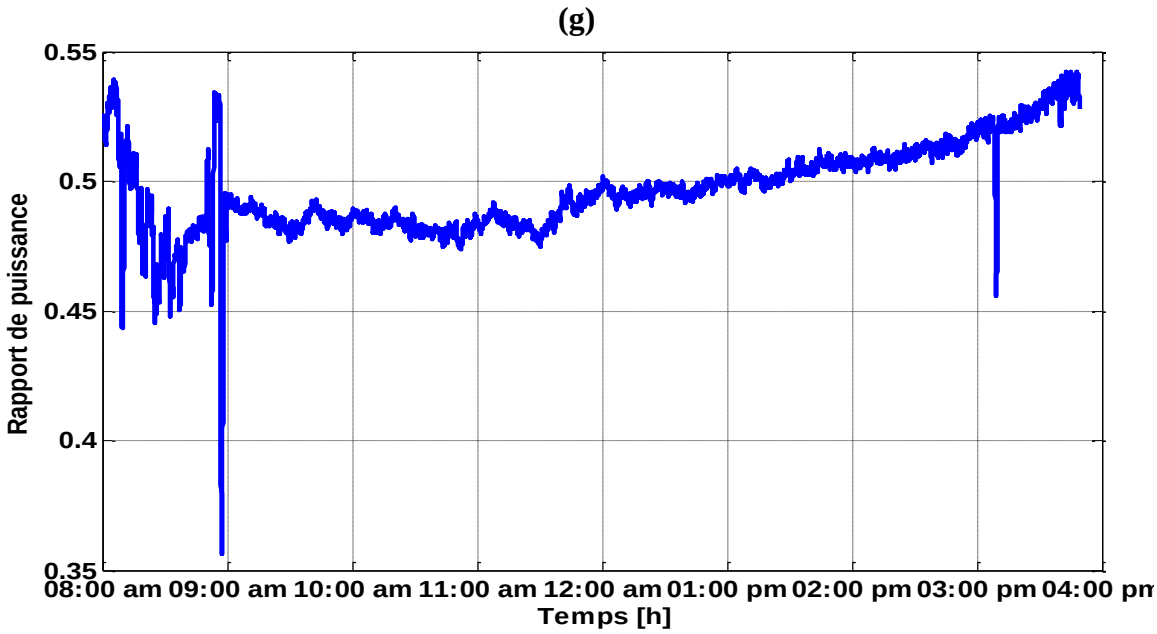


Figure 4.25 Les courbes sont : (a) la température ambiante, (b) la température estimée de la cellule, (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance.

Tableau 4.12 *Matrice de confusion du modèle de détection (Jour de chaine ouverte)*

Classé comme	**A**	**b**
a = Etat_sain	**0**	0
b = Etat_défectueux	0	**28147**

Tableau 4.13 *Résultat de test du modèle de détection*

Paramètres	**Valeur**
Temps pris pour tester le modèle	*0.16 seconde*
Instances correctement classées (28147)	*100 %*
Instances mal classées (0)	*0 %*
Nombre total des Instances	*28147*

Les résultats de détection fournis dans les tableaux 4.12 et 4.13 montrent que le modèle classe correctement toutes les instances d'états du système. Ainsi, les valeurs de précision et de temps de prédiction sont respectivement égales à 100% et 0,16 seconde. Ceci confirme l'efficacité du modèle dans la prédiction de ce type de défaut.

Le modèle de diagnostic est appliqué aux mêmes données de ce jour. Par conséquent, les informations figurant dans les tableaux 4.14 et 415 indiquent que le modèle est cohérent

pour cette anomalie de ce jour. De plus, le modèle de diagnostic donne une bonne prédiction et prend un temps d'exécution très court. Le modèle classifie parfaitement toutes les instances correctement. Donc, la précision est de 100% pendant 0,36 seconde du temps de traitement de la prévision comme est indiqué par le tableau 4.15.

Tableau 4.14 *Matrice de confusion du modèle de diagnostic.*

Classé comme	a	b	c	d
a = Etat sain	**0**	0	0	0
b = Défaut de chaîne ouverte	0	**28147**	0	0
c = Défaut de court-circuit	0	0	**0**	0
d = Défaut de ligne	0	0	0	**0**

Tableau 4.15 *Résultat de test du modèle de diagnostic*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.36 seconde*
Instances correctement classées (28147)	*100 %*
Instances mal classées (0)	*0 %*
Nombre total des Instances	*28147*

4.6.2.3 Jour de défaut « module court-circuité »

Au jour du court-circuit, le défaut est causé pour un module PV du champ d'une manière aléatoire. Les données sont collectées pendant ce jour de défaut, où les valeurs météorologiques et électriques sont tracées à la figure 4.26.

D'après la figure 4.26.e, le défaut de court-circuit affecte la puissance produite par le système, alors qu'il existe un écart entre la puissance estimée et la puissance mesurée. De plus, un module de la première chaîne est choisi pour le court-circuiter. Ce qui entraîne une dégradation des performances du système.

Comme décrit dans les tableaux 4.16 et 4.17, la matrice de confusion résultante de cette journée montre que le modèle est capable de prédire avec précision la plupart des instances à une vitesse rapide, où 26914 instances sur 27015 sont correctement prédites en 0,09 seconde. Par conséquent, la précision du modèle est égale à 99,62%.

Tableau 4.16 *Matrice de confusion du modèle de détection (Jour de court-circuit).*

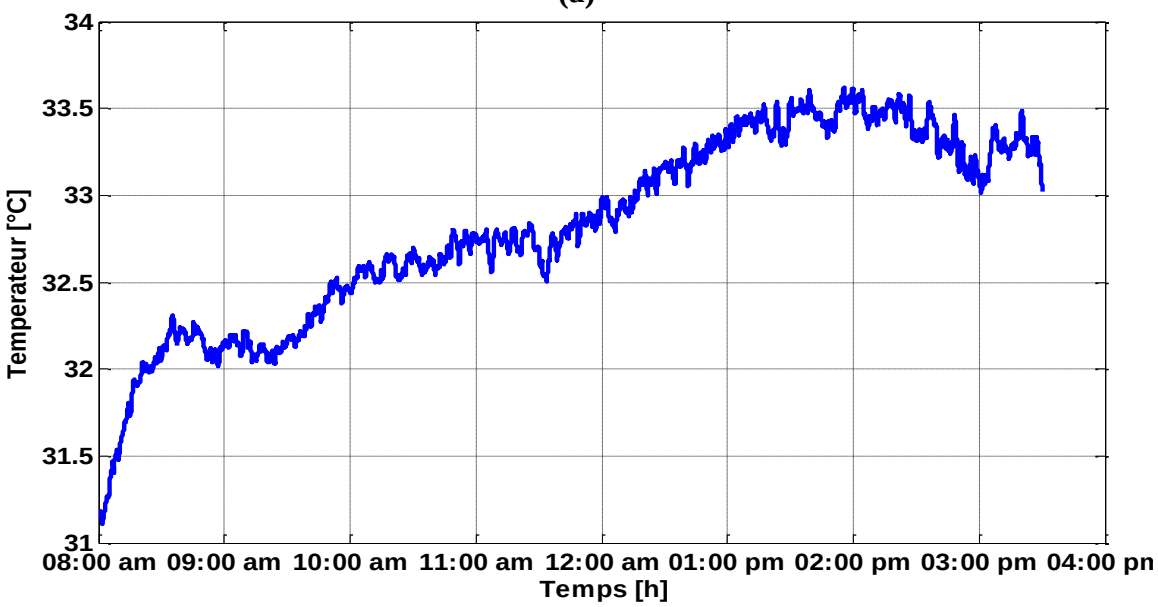
Classé comme	a	b
a = Etat_sain	**0**	0
b = Etat_défectueux	101	**26914**

Tableau 4.17 *Résultat de test du modèle de détection.*

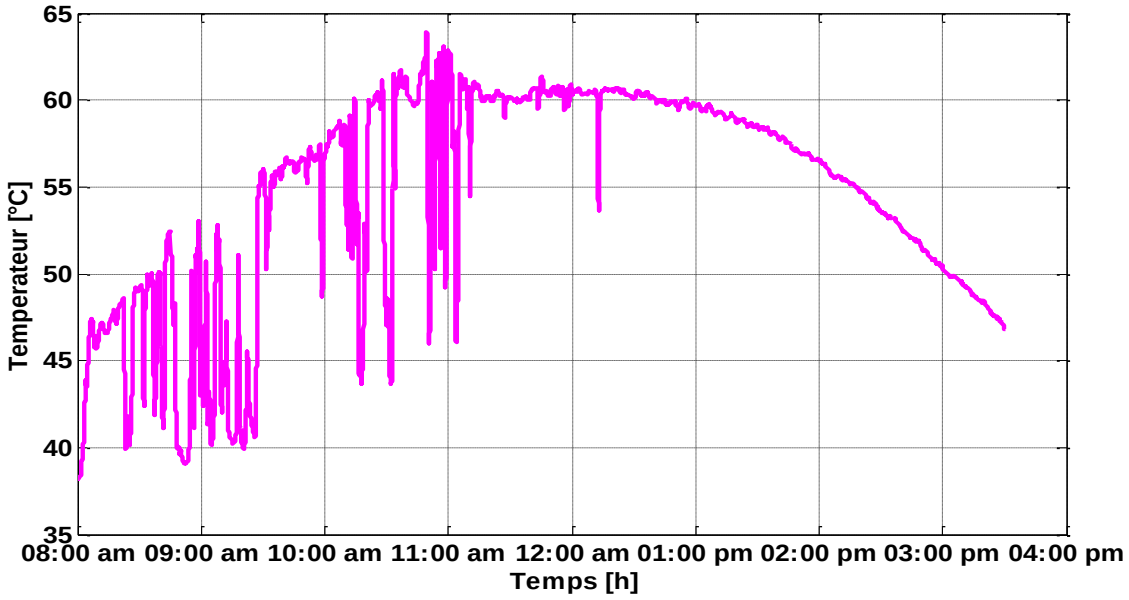
Paramètres	Valeur
Temps pris pour tester le modèle	*0.09 seconde*
Instances correctement classées (26914)	*99.6261 %*
Instances mal classées (101)	*0.3739 %*
Nombre total des Instances	*27015*

Le modèle AD du diagnostic est utilisé pour prédire la classe de défauts, le modèle ne peut pas être évité un mauvais classement de certaines instances. En revanche, la plupart des instances doivent être correctement prédites pour une bonne robustesse. En effet, les tableaux 4.18 et 4.19 montrent le résultat du modèle de classification, où la précision est très bonne et égale à 99,29%. De plus, le modèle conserve sa rapidité dans le processus de prévision avec un temps de 0,09 seconde, ce qui est considéré comme un temps très court.

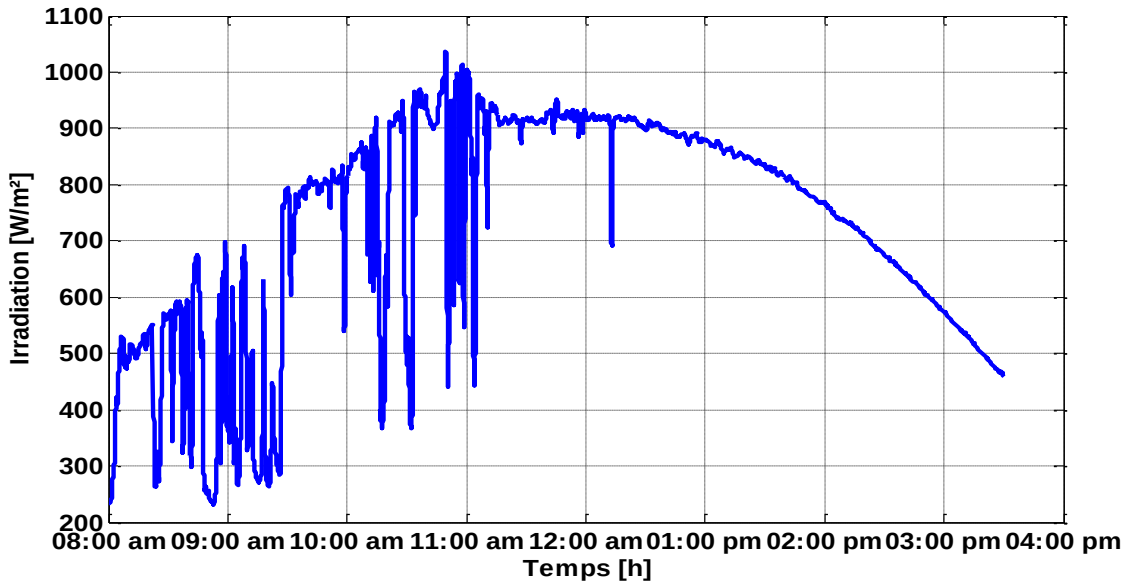
(a)



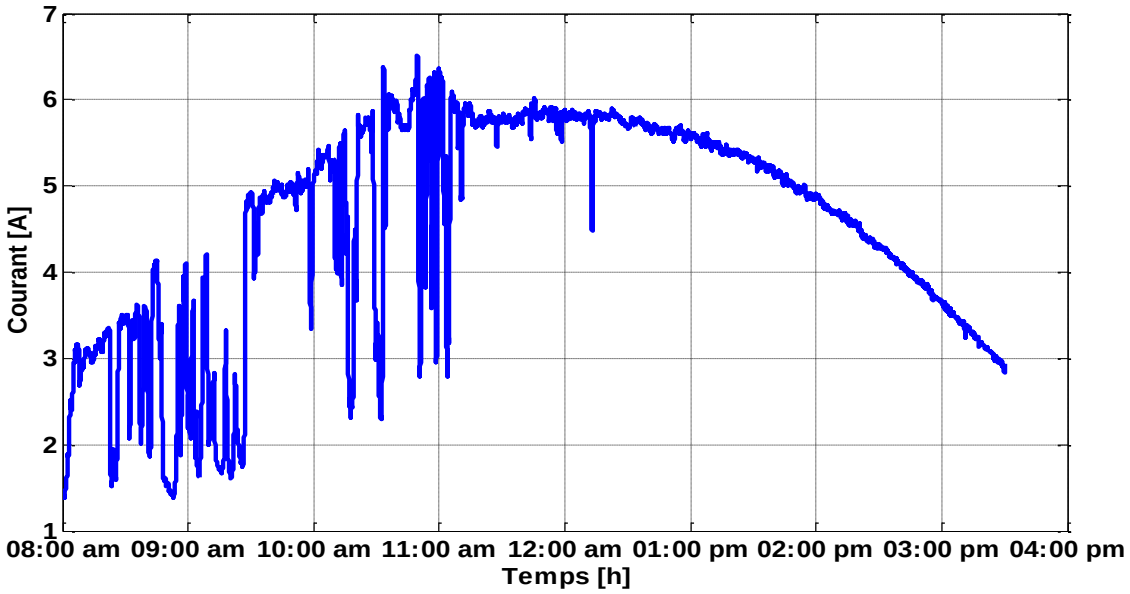
(b)


65
60
55
50
45
40
35
Temperateur [°C]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

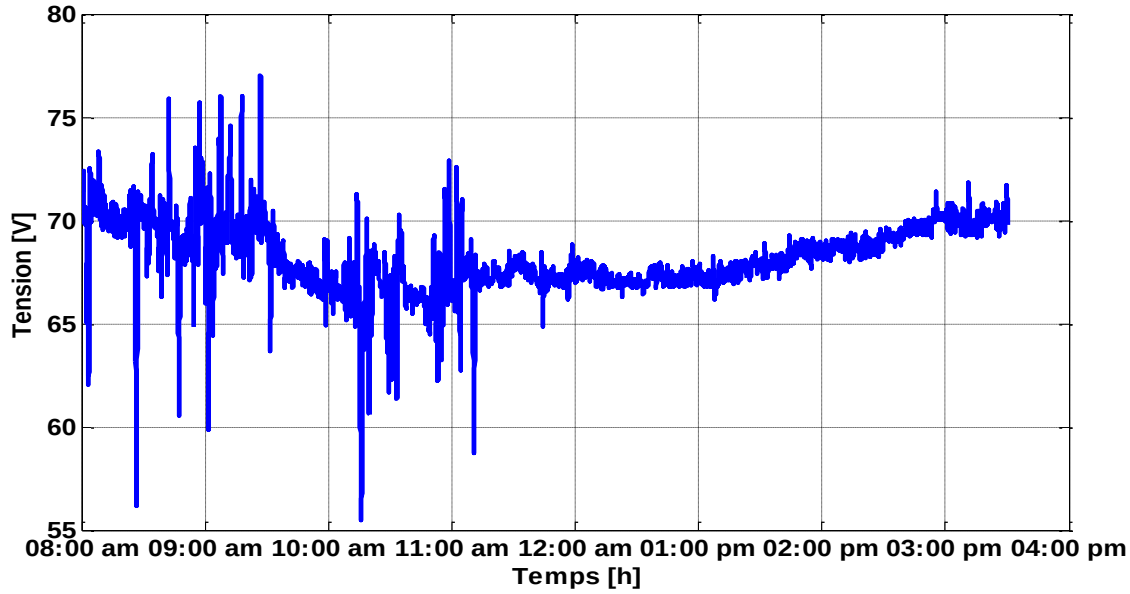
(c)


1100
1000
900
800
700
600
500
400
300
200
Irradiation [W/m²]
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

(d)


Courant [A]
7
6
5
4
3
2
1
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

(e)


Tension [V]
80
75
70
65
60
55
08:00 am 09:00 am 10:00 am 11:00 am 12:00 am 01:00 pm 02:00 pm 03:00 pm 04:00 pm
Temps [h]

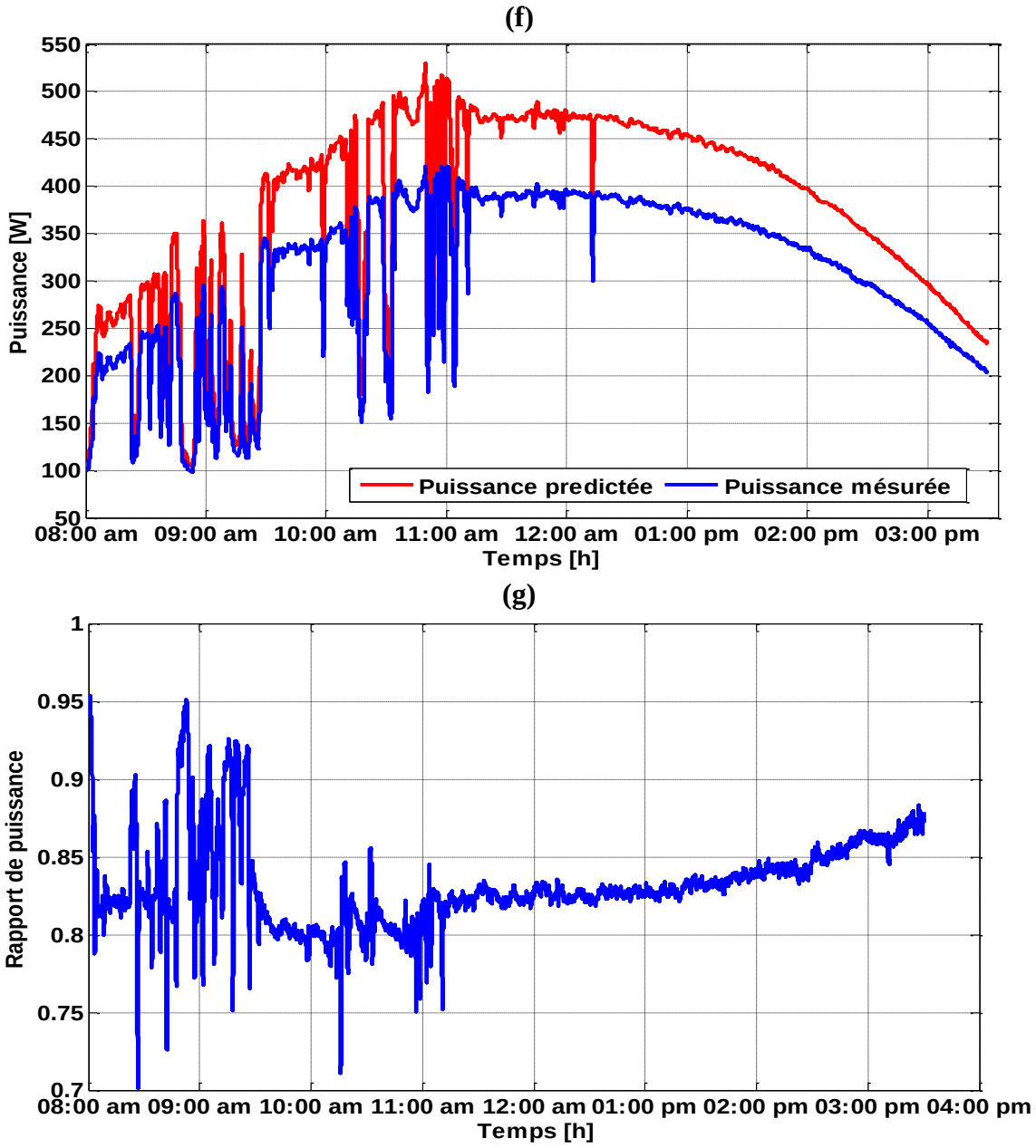


Figure 4.26 Les courbes sont : (a) la température ambiante, (b) la température estimée de la cellule, (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance.

Tableau 4.18 *Matrice de confusion du modèle de diagnostic.*

Classé comme	**a**	**b**	**c**	**d**
a = Etat sain	**0**	0	0	0
b = Défaut de chaîne ouverte	0	**0**	0	0
c = Défaut de court-circuit	129	0	**26822**	64
d = Défaut de ligne	0	0	0	**0**

Tableau 4.19 *Résultat de test du modèle de diagnostic.*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.09 seconde*
Instances correctement classées (26822)	*99.2856 %*
Instances mal classées (193)	*0.7144 %*
Nombre total des Instances	*27015*

4.6.2.4 Jour de défaut « défaut de ligne »

Finalement, le défaut de ligne est le dernier état étudié dans le système pour la validation des modèles. La figure 4.28 illustre la variation des attributs du modèle ainsi que la variation des grandeurs électriques au cours cette journée.

Les résultats obtenus à partir du modèle de détection dans ce type de défaut sont résumés dans les tableaux 4.20 et 4.21. D'après la matrice de confusion, 24121 sur 24121 instances sont correctement classées; donc la précision du modèle est de 100% alors que le processus de classification prend 0,14 seconde. De plus, le modèle de diagnostic est testé par le même ensemble. La matrice de confusion résultante est indiquée dans le tableau ci-dessous.

Tableau 4.20 *Matrice de confusion du modèle de détection (Jour de défaut de ligne).*

Classé comme	a	b
a = Etat_sain	**0**	0
b = Etat_défectueux	0	**24121**

Tableau 4.21 *Résultat de test du modèle de détection.*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.14 seconde*
Instances correctement classées (24121)	*100 %*
Instances mal classées (0)	*0 %*
Nombre total des Instances	*24121*

La matrice de confusion a prouvé que le modèle de diagnostic avait une bonne efficacité de prédiction, puis le nombre total d'instances est classé en tant que défaut de ligne. La précision résultante est donc de 100%, ce qui signifie que la prédiction de l'état du système est parfaitement correcte (voir les tableaux 4.22 et 4.23).

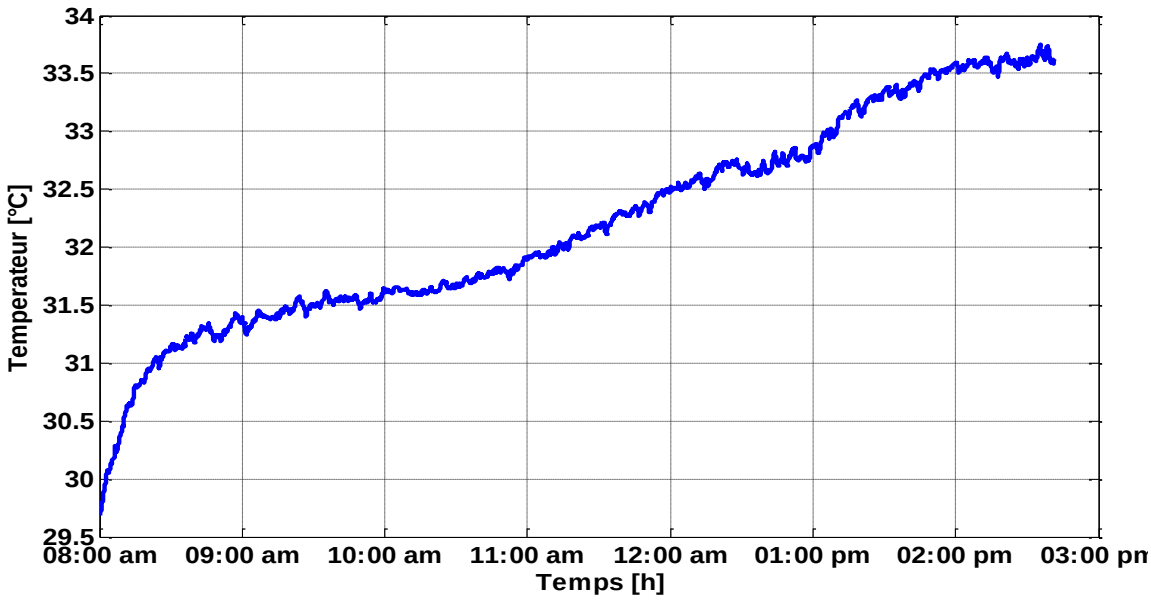
Tableau 4.22 *Matrice de confusion du modèle de diagnostic.*

Classé comme	a	b	c	d
a = Etat sain	**0**	0	0	0
b = Défaut de chaîne ouverte	0	**0**	0	0
c = Défaut de court-circuit	0	0	**0**	0
d = Défaut de ligne	0	0	0	**24121**

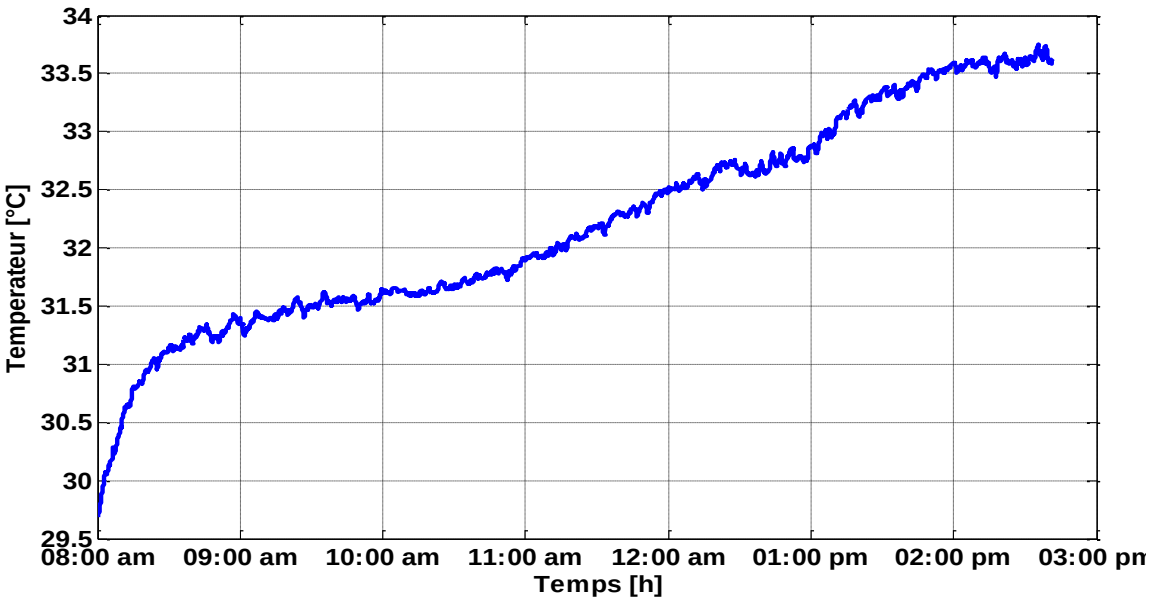
Tableau 4.23 *Résultat de test du modèle de diagnostic.*

Paramètres	Valeur
Temps pris pour tester le modèle	*0.09 seconds*
Instances correctement classées (24121)	*100 %*
Instances mal classées (0)	*0 %*
Nombre total des Instances	*24121*

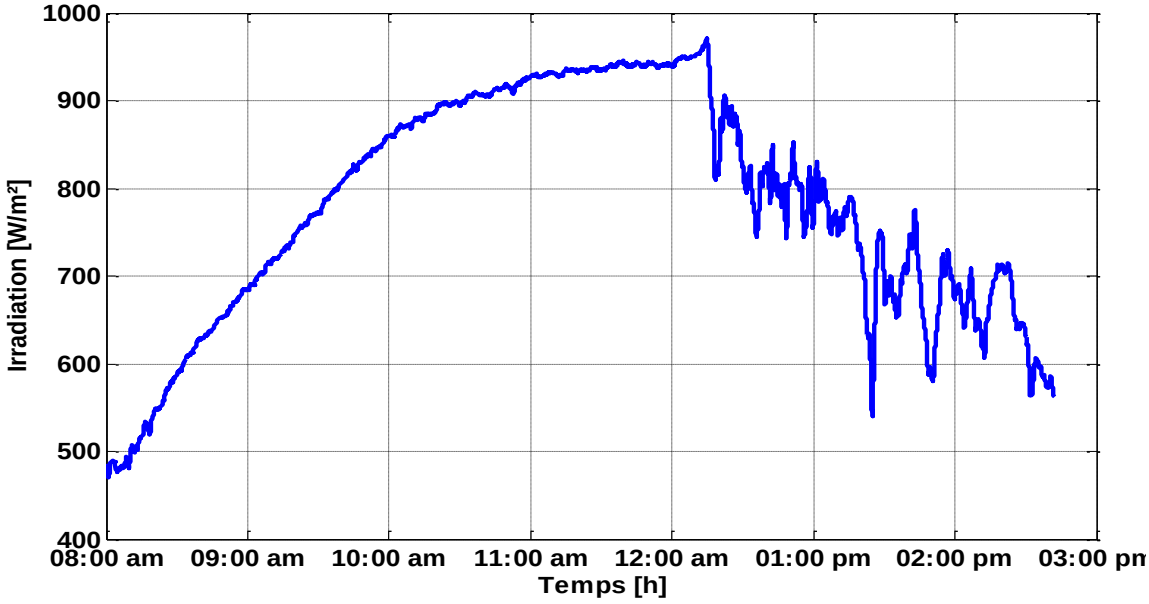
(a)



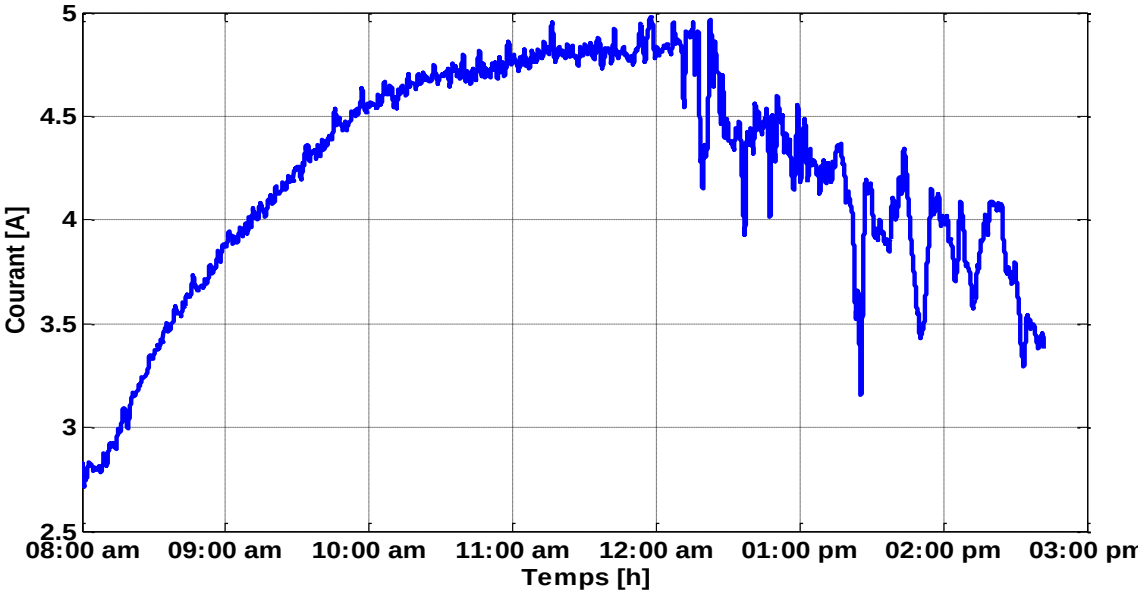
(b)


Temperateur [°C]
34
33.5
33
32.5
32
31.5
31
30.5
30
29.5
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

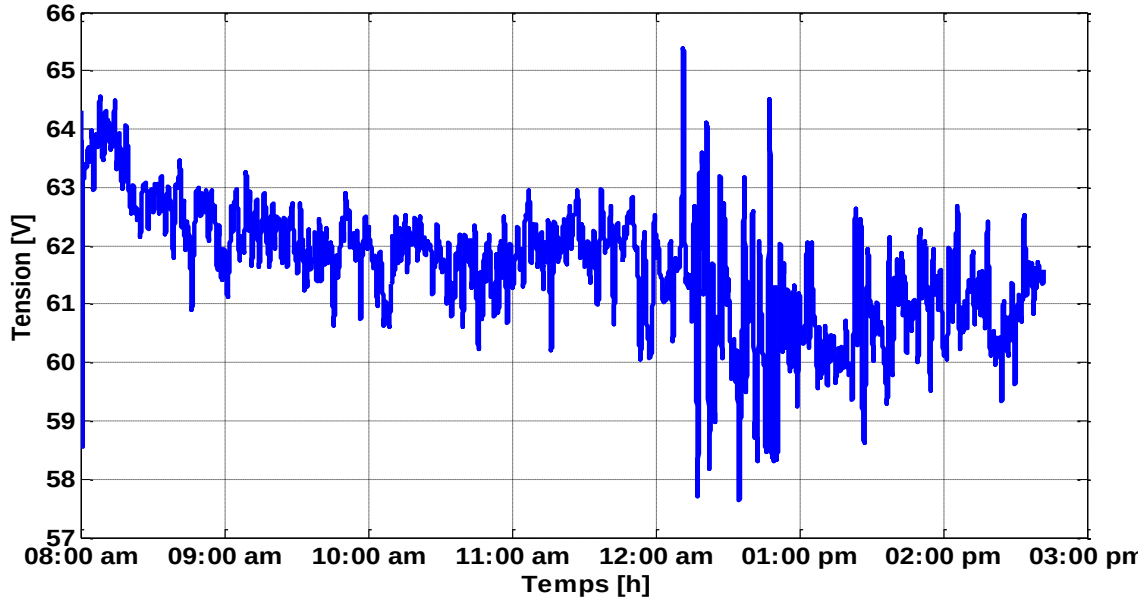
(c)


Irradiation [W/m²]
1000
900
800
700
600
500
400
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(d)


5
4.5
4
3.5
3
2.5
Courant [A]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

(e)


66
65
64
63
62
61
60
59
58
57
Tension [V]
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
Temps [h]

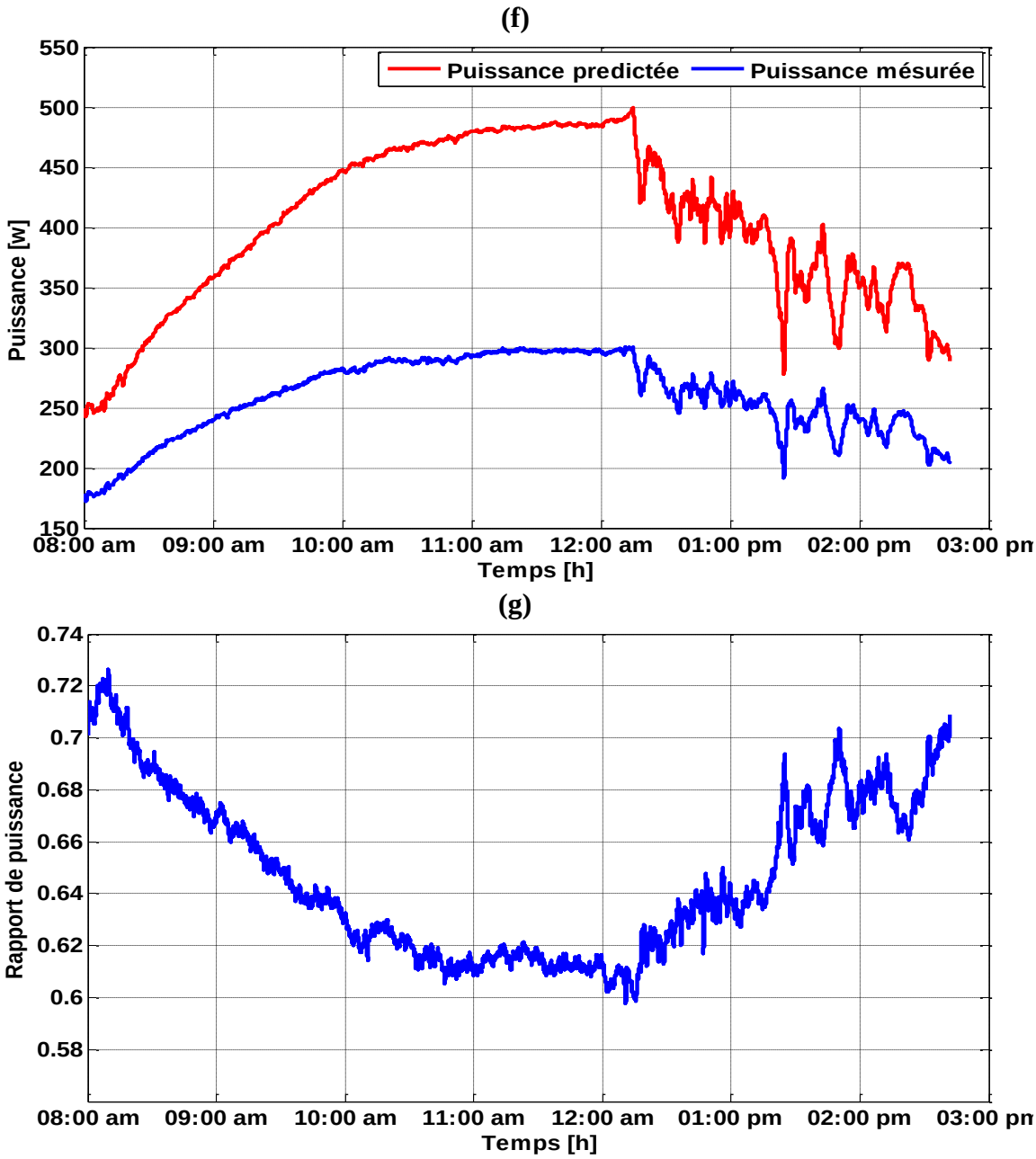


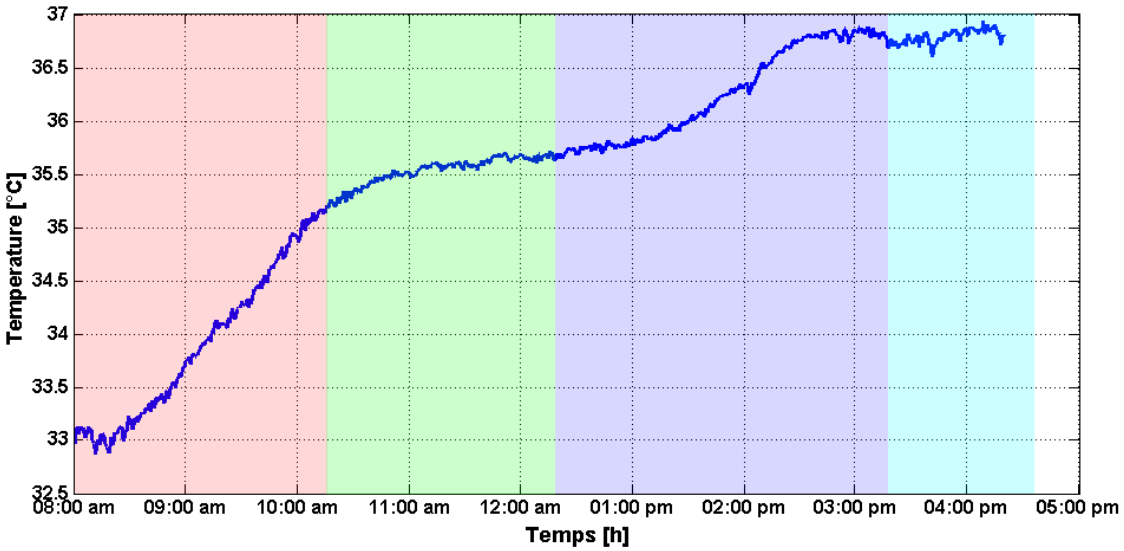
Figure 4.27 Les courbes sont : (a) la température ambiante, (b) la température estimée de la cellule, (c) l'irradiation, (d) le courant, (e) la tension, (f) la puissance estimée et mesurée et (g) le rapport de puissance.

4.6.2.5 Jour des défauts

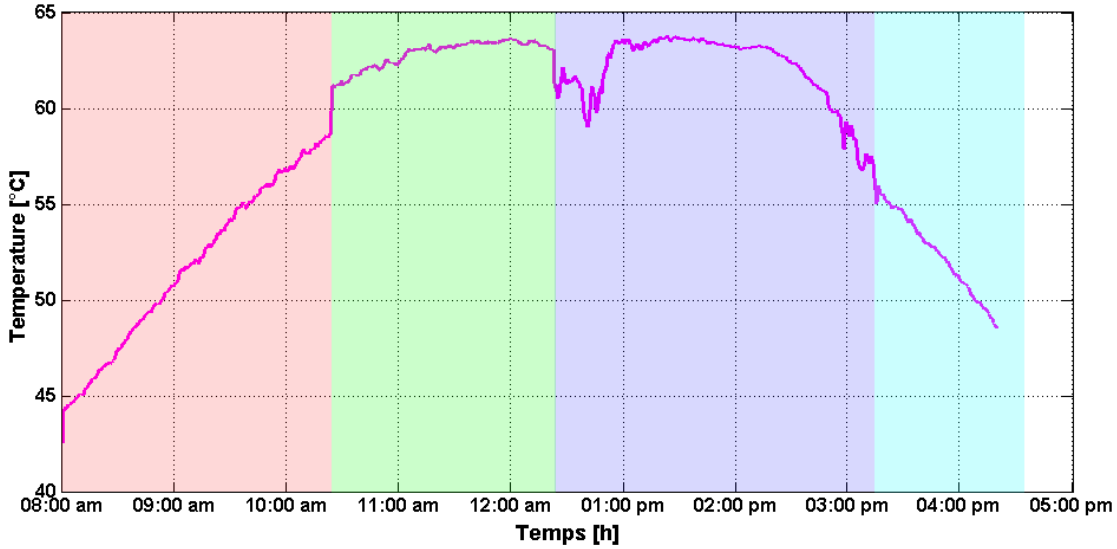
A ce jour, les quatre défauts sont introduits dans le système PV les unes après les autres pour permettre de tester les deux modèles de classification. Pour cela, on crée d'abord un court-circuit du module PV pendant 2 h 23 mn 47 s, puis une chaîne PV est déconnectée et

le système reste en cet état pendant 1 h 58 mn 40 s, puis le système revient à l'état normal (sain) pour une durée de 2 h 52 mn 21 s et finalement, un défaut de ligne est appliqué dans le système pendant une durée 1 h 4 mn 18 s de temps.

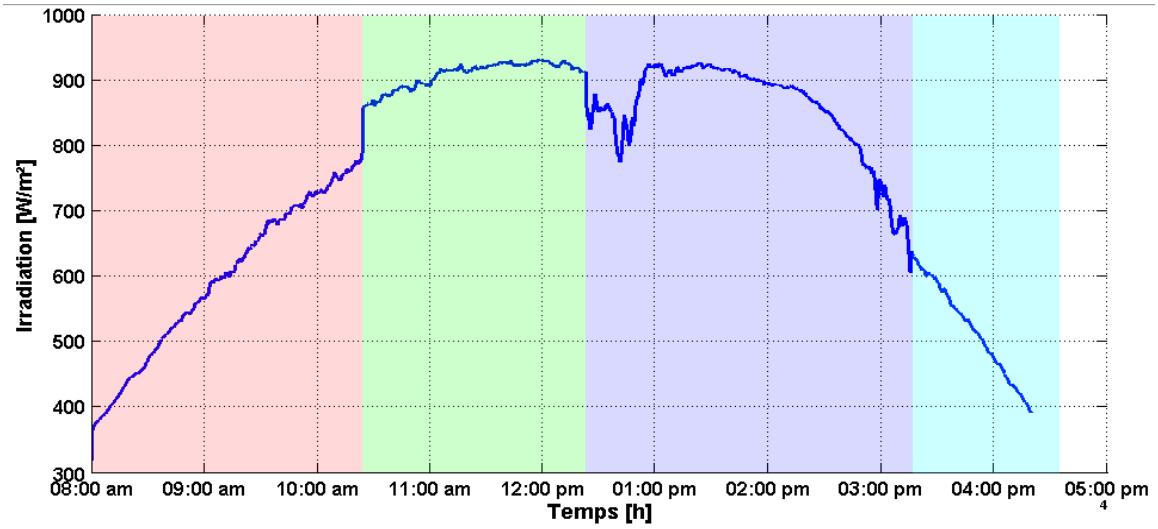
La figure 4.28 représente la variation des données météorologiques et électriques au cours de cette journée. La figure 4.28.d montre l'écart entre la sortie du modèle SANDIA et la puissance mesurée qui représente les pertes de puissance pour chaque défaut.



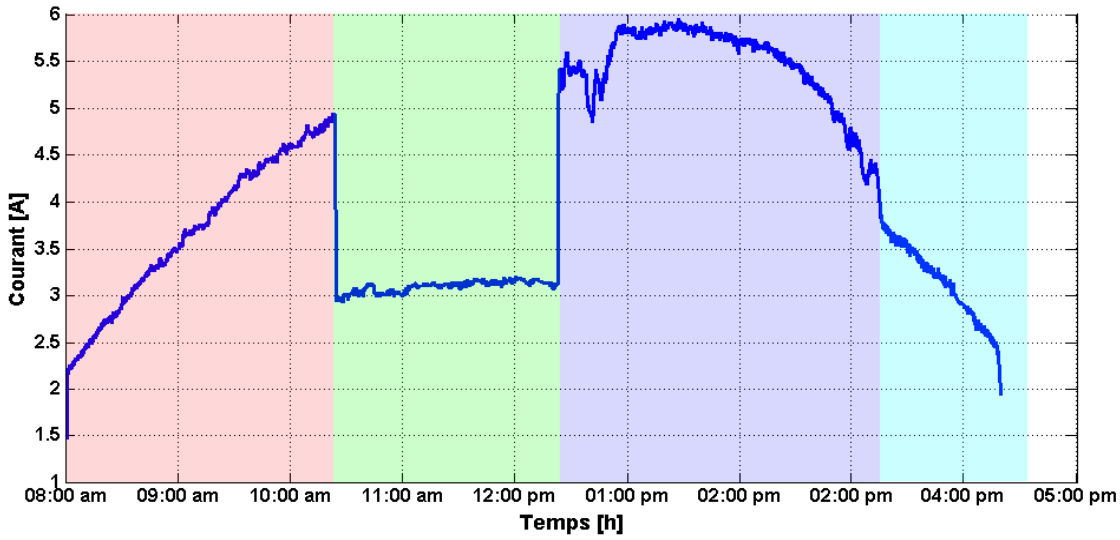
(a)



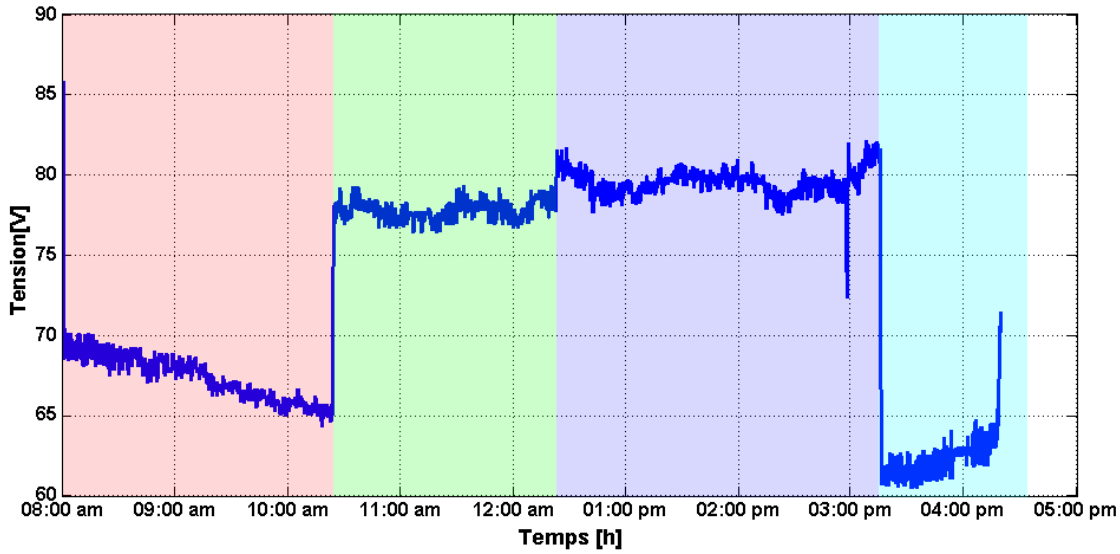
(b)


Irradiation [W/m²]
1000
900
800
700
600
500
400
300
08:00 am
09:00 am
10:00 am
11:00 am
12:00 pm
01:00 pm
02:00 pm
03:00 pm
04:00 pm
05:00 pm
Temps [h]

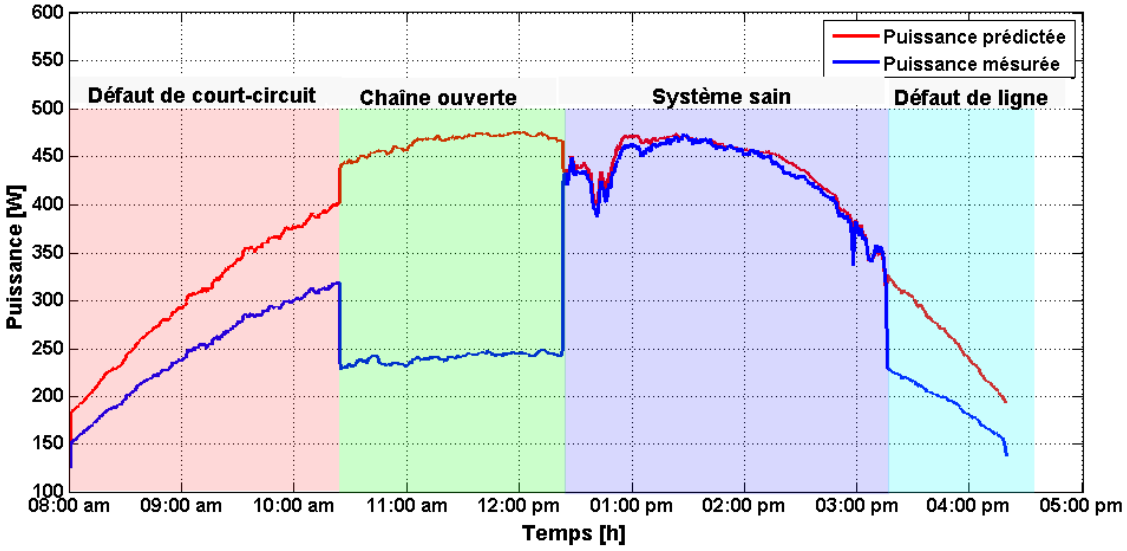
(c)


Courant [A]
6
5.5
5
4.5
4
3.5
3
2.5
2
1.5
1
08:00 am
09:00 am
10:00 am
11:00 am
12:00 pm
01:00 pm
02:00 pm
02:00 pm
04:00 pm
05:00 pm
Temps [h]

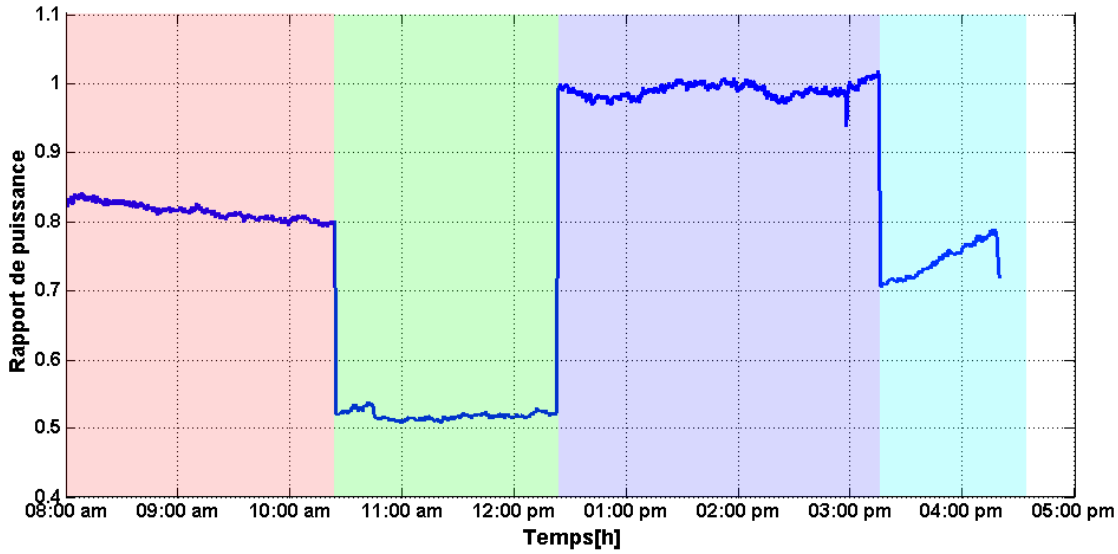
(d)


Tension[V]
90
85
80
75
70
65
60
08:00 am
09:00 am
10:00 am
11:00 am
12:00 am
01:00 pm
02:00 pm
03:00 pm
04:00 pm
05:00 pm
Temps [h]

(e)


Puissance prédictée
Puissance mésurée
Défaut de court-circuit
Chaîne ouverte
Système sain
Défaut de ligne
Puissance [W]
600
550
500
450
400
350
300
250
200
150
100
08:00 am
09:00 am
10:00 am
11:00 am
12:00 pm
01:00 pm
02:00 pm
03:00 pm
04:00 pm
05:00 pm
Temps [h]

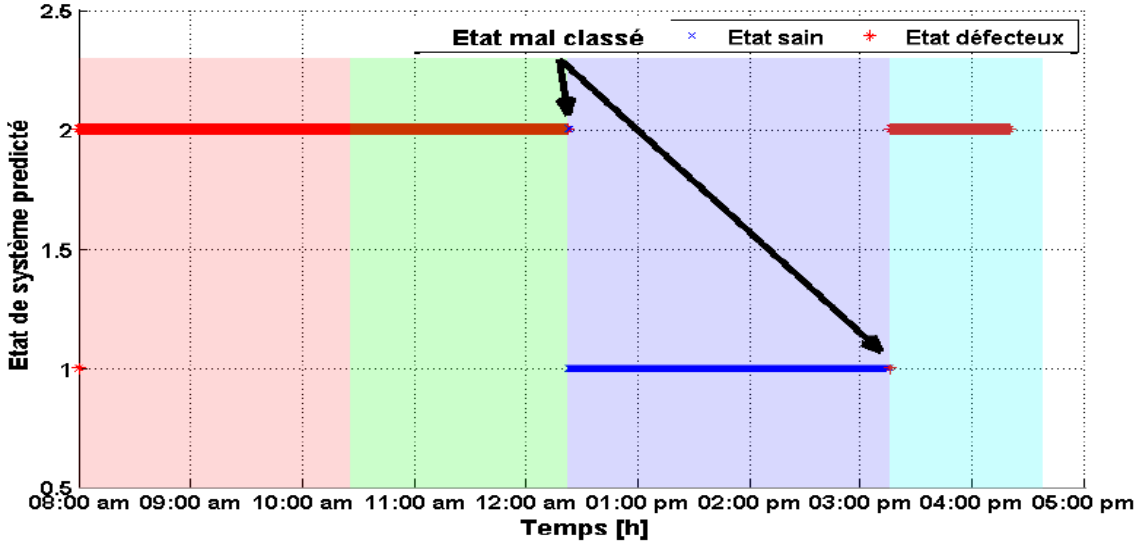
(f)



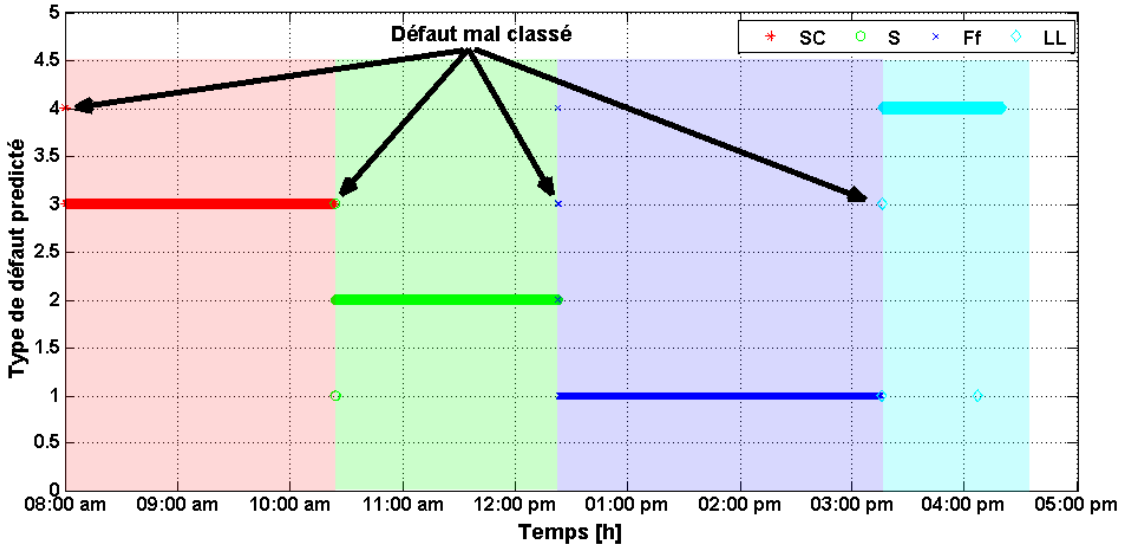
(g)

Figure 4.28 Les courbes sont : (a) la température ambiante, (b) la température estimée de la cellule, (c) l'irradiation, (d) le courant (e) la tension (f) la puissance estimée et mesurée et (g) le rapport de puissance.

Les résultats de la classification pour les modèles de détection et de diagnostic de chaque échantillon sont données par à la figure 4.29. Il est possible de remarquer sur la figure que ces modèles ont une grande précision pour connaitre l'état de système et aussi pour distinguer les différentes situations (les défauts) du système au cours de son fonctionnement.



(a)



(b)

Figure 4.29 Les résultats de la classification de (a) modèle de détection et (b) modèle de diagnostic.

Les matrices de confusion obtenues sont représentées par les tableaux 4.24 et 4.25 montrent les résultats de test pour se jour. Il est évident que le modèle a une bonne déduction. Par conséquent, la plupart des échantillons ont été correctement classés, même si le système était exposé à de multiples défauts au cours de la journée, où 29893 sur 29946 instances sont classées avec précision. Le modèle a pris un temps de prévision égale à 0,07 seconde, la précision du modèle est donc égale à 99,823%.

Tableau 4.24 *Matrice de confusion du modèle de détection (Jour des défauts).*

Classé comme	**a**	**b**
a = Etat_sain	**10318**	23
b = Etat_défectueux	30	**19575**

Tableau 4.25 *Résultat de test du modèle de détection.*

Paramètres	**Valeur**
Temps pris pour tester le modèle	*0.07 seconde*
Instances correctement classées (29893)	*99.823 %*
Instances mal classées (53)	*0.177 %*
Nombre total des Instances	*29946*

En outre, les résultats du modèle de diagnostic sont donnés dans les tableaux 4.24 et 4.25 qui montrent l'efficacité de la prévision dans ce jour où le système est soumis à

plusieurs défauts. Pour bien analyser, le modèle est correctement classé: 10311 instances sur 10341, 7093 instances sur 7120, 8611 instances sur 8627 instances et 3813 instances sur 3858 instances pour respectivement : état sain, défaut de chaîne ouverte, défaut de court-circuit et défaut de ligne. Par conséquent, la précision du modèle à ce jour est de 99,606% et le modèle a maintenu son temps de prédiction qui est toujours rapide et qui est de 0,1 seconde.

Tableau 4.26 *Matrice de confusion du modèle de diagnostic.*

Classé comme	**a**	**b**	**c**	**d**
a = Etat sain	**10311**	5	24	1
b = Défaut de chaîne ouverte	22	**7093**	5	0
c = Défaut de court-circuit	0	0	**8611**	16
d = Défaut de ligne	30	0	15	**3813**

Tableau 4.27 *Résultat de test du modèle de diagnostic.*

Paramètres	**Valeur**
Temps pris pour tester le modèle	*0.1 seconde*
Instances correctement classées (29828)	*99.606 %*
Instances mal classées (118)	*0.394 %*
Nombre total des Instances	*29946*

4.7. Conclusion

Dans ce chapitre, l'algorithme d'arbre de décision a été utilisé pour la détection et le diagnostic de défauts dans le SPVGC. L'algorithme C4.5 construit deux modèles : le premier pour la détection de l'état du système qui a deux étiquettes de classe nommées sain et défectueux. Le second pour le diagnostic de défauts, qui contient quatre classes: sain, défaut de chaîne ouverte, défaut de court-circuit et défaut de ligne. Dans le premier temps, les phases principales utilisées pour l'apprentissage de l'algorithme C4.5 ont été expliquées. L'apprentissage est basé sur des critères de fractionnement permettant de choisir le meilleur attribut de division. L'attribut ayant le rapport de gain maximal est placé au point de fractionnement, tandis que ce processus est appliqué d'une manière récursive sur le jeu de données jusqu'à la vérification de l'un des critères d'arrêt. Après la construction, une technique d'élagage est utilisée pour améliorer le modèle en supprimant les sous-arbres inutiles. Dans cet arbre, il existe trois types principaux de nœuds, le nœud

racine, les nœuds internes et les nœuds terminaux (les feuilles). Ces nœuds connectés entre eux par des branches, et aussi les nœuds représentant les attributs et les valeurs de seuil de ces attributs sont placés sur les branches marginales de chaque nœud. L'arbre se termine par des feuilles et chacune d'elles contient l'étiquette de classe. Le jeu de données a été collecté du SPVGC pendant son fonctionnement dans un état sain ou défaillant. Par conséquent, trois défauts sont réalisés dans le système, à savoir défaut chaîne ouverte, défaut de court-circuit et défaut de ligne. L'ensemble de données a été divisé en deux parties, la première partie comprend 66% de données globales et cette partie destinée pour l'apprentissage de modèle, tandis que les données restantes représentent la seconde partie. Cette dernière est utilisée pour le test de modèle. Les échantillons d'apprentissage ou de test sont choisis au hasard. Les résultats des tests ont montré l'efficacité de la classification des deux modèles, où les précisions sont égales à 99,86% et 99,80% respectivement pour le modèle de détection et de diagnostic. La validation assure l'efficacité et la robustesse (généralisation) des modèles. Le processus d'évaluation pour valider les deux modèles est basé sur des données des jours et chaque jour le système fonctionne dans un état, où le processus de prédiction a une exécution rapide avec une précision très élevée (plus de 99%) pour tous les jours.

CONCLUSION GENERALE

Conclusion générale

Le travail présenté dans le cadre de cette thèse s'inscrit dans le cadre de diagnostic et la détection des défauts. L'objectif consiste à développer une méthode, basée sur la classification par les arbres de décision, pour la détection et le diagnostic des défauts dans le SPVCR. Le travail présenté comporte deux parties importantes qui sont l'étude des éléments de système PV et les défauts qui peuvent apparaître durant son fonctionnement, et la construction de l'arbre de décision pour la classification des défauts. Bien que la construction des arbres de décisions à partir des données soit une discipline très utilisée mais peu de travaux ont été consacrés pour la détection et le diagnostic de défauts dans les systèmes PVs.

Ainsi, après avoir présenté les notions de base relatives aux systèmes photovoltaïques, à détection et au diagnostic de défauts, nous avons détaillé les points clés pour la construction d'un arbre de décision à partir de données expérimentales. Ensuite, nous avons développé deux arbres de décision, le premier pour la détection et le deuxième pour le diagnostic de défauts dans une installation photovoltaïque. Pour la construction des différents arbres, nous avons utilisé l'algorithme de classification C4.5 implémentée dans un logiciel appelé WEKA qui est un logiciel libre et téléchargeable sur internet, et un ensemble de données issues de l'installation PV considérée. Cet ensemble de données, couvrant le fonctionnement sain et avec défauts de l'installation, contient des valeurs pouvant conduire à un modèle biaisé. Ces données indésirables sont dues à un ensoleillement faible ne doivent pas être utilisées dans la phase de construction des modèles de décisions. Par conséquent, une opération de prétraitement est indispensable pour éliminer toutes les valeurs indésirables de la base de données et celles bruitées.

Afin d'évaluer l'efficacité et la précision des modèles de détection et diagnostic de défauts développés, nous avons effectué plusieurs tests en considérant différents défauts sur des données réelles récupérées durant plusieurs journées. Les différents tests ont été réalisés sur cinq jours dont quatre jours ont été dédiés à un défaut du système et le cinquième jour contient tous les défauts du système. Les résultats obtenus sont satisfaisants et concluants. Le taux de bonne classification est supérieur à 99% pour les deux modèles proposés.

Le travail réalisé est un sujet d’actualité et nous a permis d'investir des domaines très prometteurs à savoir le diagnostic des systèmes d’une façon générale et plus particulièrement le système PV connecté au réseau, et la classification des données en utilisant les arbres de décision. Les résultats probants obtenus motivent l’orientation de nos futurs travaux de recherche.

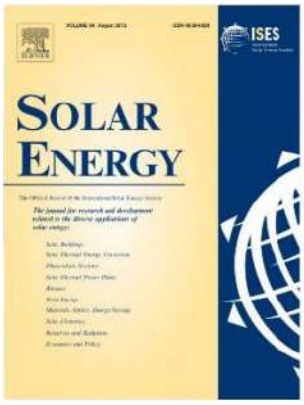
Le travail que nous avons réalisé pourrait faire l’objet de plusieurs améliorations. Nous pouvons citer, entre autre, la considération d’autres attributs et l’inclusion dans la base de données d’autres types de défauts par l’exemple l’ombrage partiel et le défaut de terre. Aussi, l'utilisation des données météorologiques satellitaires permet d'améliorer la fiabilité des modèles de détection de défauts. Une autre piste très intéressante consiste à intégrer l'onduleur dans le processus de développement de ses modèles.

PUBLICATIONS ET CONFERENCES

Publications et Conférences

1. Benkercha, R., & Moulahoum, S. (2018). Fault detection and diagnosis based on C4. 5 decision tree algorithm for grid connected PV system. Solar Energy, 173, 610-634.

2. Blaifi, S., Moulahoum, S., Benkercha, R., Taghezouit, B., Saim, A., M5P model treebased fast fuzzy maximum power point tracker, Sol. Energy 163 (January) (2018) 405-424.



CiteScore: 8.1

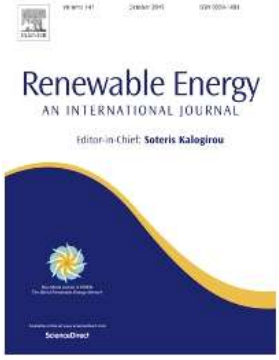
Impact Factor: 4.608 (2019 Journal Citation Reports)

5-Year Impact Factor: 4.74 (2019 Journal Citation Reports)

Source Normalized Impact per Paper (SNIP): 1.651

SCImago Journal Rank (SJR): 1.537

3. Benkercha, R., Moulahoum, S., & Taghezouit, B. (2019). Extraction of the PV modules parameters with MPP estimation using the modified flower algorithm. Renewable Energy. Jun 1.



CiteScore: 11.2

Impact Factor: 6.247 (2019 Journal Citation Reports)

5-Year Impact Factor: 5.964 (2019 Journal Citation Reports)

Source Normalized Impact per Paper (SNIP): 2.366

SCImago Journal Rank (SJR): 2.052

4. Benkercha, R., Moulahoum, S., Colak, I., & Taghezouit, B. (2016, September). PV module parameters extraction with maximum power point estimation based on flower pollination algorithm. In 2016 IEEE International Power Electronics and Motion Control Conference (PEMC) (pp. 442-449). IEEE.

5. Benkercha, R., Moulahoum, S., & Kabache, N. (2017, September). Combination of artificial neural network and flower pollination algorithm to model fuzzy logic MPPT controller for photovoltaic systems. In 2017 18th International Symposium on Electromagnetic Fields in Mechatronics, Electrical and Electronic Engineering (ISEF) Book of Abstracts (pp. 1-2). IEEE.

6. Benkercha, R., Moulahoum, S., & Colak, I. (2017, December). Modelling of Fuzzy Logic Controller of a Maximum Power Point Tracker Based on Artificial Neural Network. In 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA) (pp. 485-492). IEEE.

7. R. Benkercha, S. Moulahoum, B. Taghezouit, Parameters extraction of single and double diode model using theflower algorithm, in: IEEE International Conference on Photovoltaic Science and Technologies (PVCon), 2018, pp. 1e6. Ankara, Turkey.

REFERENCES

Références

[1] Mellit, A., Tina, G. M., & Kalogirou, S. A. (2018). Fault detection and diagnosis methods for photovoltaic systems: A review. Renewable and Sustainable Energy Reviews, 91, 1-17.

[2] Zhang, Y., Sakhuja, M., Lim, F. J., Tay, S., Tan, C., Bieri, M., ... & Reindl, T. (2017). The PV System Doctor–Comprehensive diagnosis of PV system installations. Energy Procedia, 130, 108-113.

[3] Yang, Y., Kim, K. A., Blaabjerg, F., & Sangwongwanich, A. (2018). *Advances in Grid-Connected Photovoltaic Power Conversion Systems*. Woodhead Publishing.

[4] Halwachs, M., Neumaier, L., Vollert, N., Maul, L., Dimitriadis, S., Voronko, Y., ... & Ebner, R. (2019). Statistical evaluation of PV system performance and failure data among different climate zones. Renewable Energy, 139, 1040-1060.

[5] Triki-Lahiani, A., Abdelghani, A. B. B., & Slama-Belkhodja, I. (2018). Fault detection and monitoring systems for photovoltaic installations: A review. Renewable and Sustainable Energy Reviews, 82, 2680-2692.

[6] Silvestre, S., Tahri, A., Tahri, F., Benlebna, S., & Chouder, A. (2018). Evaluation of the performance and degradation of crystalline silicon-based photovoltaic modules in the Saharan environment. Energy, 152, 57-63.

[7] Benkercha, R., Moulahoum, S., Colak, I., & Taghezouit, B. (2016, September). PV module parameters extraction with maximum power point estimation based on flower pollination algorithm. In *2016 IEEE International Power Electronics and Motion Control Conference (PEMC)* (pp. 442-449). IEEE.

[8] Benkercha, R., Moulahoum, S., & Taghezouit, B. (2018, July). Parameters Extraction of Single and Double Diode Model Using the Flower Algorithm. In *2018 International Conference on Photovoltaic Science and Technologies (PVCon)* (pp. 1-6). IEEE.

[9] Harrou, F., Sun, Y., Taghezouit, B., Saidi, A., & Hamlati, M. E. (2018). Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renewable energy*, *116*, 22-37.

[10] Harrou, F., Taghezouit, B., & Sun, Y. (2019). Improved KNN-Based Monitoring Schemes for Detecting Faults in PV Systems. *IEEE Journal of Photovoltaics*.

[11] Tadj, M., Benmouiza, K., Cheknane, A., & Silvestre, S. (2014). Improving the performance of PV systems by faults detection using GISTEL approach. *Energy conversion and management*, *80*, 298-304.

[12] Zhao, Y., Yang, L., Lehman, B., de Palma, J. F., Mosesian, J., & Lyons, R. (2012, February). Decision tree-based fault detection and classification in solar photovoltaic arrays. In *2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC)* (pp. 93-99). IEEE.

[13] Pillai, D. S., Blaabjerg, F., & Rajasekar, N. (2019). A Comparative Evaluation of Advanced Fault Detection Approaches for PV Systems. *IEEE Journal of Photovoltaics*, *9*(2), 513-527.

[14] Dhimish, M., Holmes, V., & Dales, M. (2017). Parallel fault detection algorithm for grid-connected photovoltaic plants. *Renewable energy*, *113*, 94-111.

[15] Lahdhiri, H., Said, M., Abdellafou, K. B., Taouali, O., & Harkat, M. F. (2019). Supervised process monitoring and fault diagnosis based on machine learning methods. *The International Journal of Advanced Manufacturing Technology*, 1-17.

[16] Peled, A., & Appelbaum, J. (2017). Enhancing the power output of PV modules by considering the view factor to sky effect and rearranging the interconnections of solar cells. Progress in Photovoltaics: Research and Applications, 25(9), 810-818.

[17] Madeti, S. R., & Singh, S. N. (2017). Monitoring system for photovoltaic plants: A review. Renewable and Sustainable Energy Reviews, 67, 1180-1207.

[18] Benkercha, R., Moulahoum, S., & Taghezouit, B. (2019). Extraction of the PV modules parameters with MPP estimation using the modified flower algorithm. Renewable Energy, 143, 1698-1709.

[19] Yang, X. S. (Ed.). (2017). Nature-inspired algorithms and applied optimization (Vol. 744). Springer. (Livre).

[20] Han, J., Kamber, M., & Pei, J. (2011). Data mining concepts and techniques third edition. The Morgan Kaufmann Series in Data Management Systems, 5(4), 83-124. (Livre).

[21] Sabri,N . Benkercha, R. (2013). Détection et Diagnostic des pannes dans le système photovoltaïque par les techniques intelligentes. Master Dissertation.

[22] Humada, A. M., Hojabri, M., Mekhilef, S., & Hamada, H. M. (2016). Solar cell parameters extraction based on single and double-diode models: A review. Renewable and Sustainable Energy Reviews, 56, 494-509.

[23] Dhimish, M. (2018). Fault Detection and Performance Analysis of Photovoltaic Installations (Doctoral dissertation, University of Huddersfield).

[24] Messenger, R. A., & Abtahi, A. (2010). *Photovoltaic systems engineering*. CRC press.

[25] Lumby, B. (2015). Les centrales solaires photovoltaïques commerciales: guide à l'intention des promoteurs de projets (No. 99396, pp. 1-122). The World Bank.

[26] Tang, C. W. (1986). Two-layer organic photovoltaic cell. *Applied physics letters*, *48*(2), 183-185.

[27] Castaner, L., & Silvestre, S. (2002). Modelling photovoltaic systems using PSpice. John Wiley and Sons.

[28] Voudoukis, N. F. (2018). Photovoltaic Technology and Innovative Solar Cells. European Journal of Electrical Engineering and Computer Science, 2(1).

[29] Boyle, G. (2004). Photovoltaic power. (Livre).

[30] Omran, W. (2010). Performance analysis of grid-connected photovoltaic systems. (Doctoral dissertation).

[31] Petibon, S. (2009). Nouvelles architectures distribuées de gestion et conversion de l'énergie pour les applications photovoltaïques. (Doctoral dissertation, Université Paul Sabatier-Toulouse III).

[32] Gorge, V. (2012). Caractérisations de matériaux et tests de composants des cellules solaires à base des nitrures des éléments III-V (Doctoral dissertation, Paris 11).

[32] Gallegos, C. D. R., & Alvarado, M. S. A. (2015). Analysis of the stationary and transient behavior of a photovoltaic solar array: modeling and simulation. International Journal of Computer Applications, 975, 8887.

[34] Zahedi, A. (2006). Solar photovoltaic (PV) energy; latest developments in the building integrated and hybrid PV systems. Renewable Energy, 31(5), 711-718.

[35] Notton, G., Lazarov, V., & Stoyanov, L. (2010). Optimal sizing of a grid-connected PV system for various PV module technologies and inclinations, inverter efficiency characteristics and locations. Renewable Energy, 35(2), 541-554.

[36] Hadj Arab, A., Cherfa, F., Chouder, A., Chenlo, F. (2005). Grid connected photovoltaic system at CDER-Algeria. In: Proceedings of the 20th European Solar Energy Conf. and Exhibition, June 2005, Barcelona, Spain.

[37] Macabebe, E. Q. B., Sheppard, C. J., & van Dyk, E. E. (2011). Parameter extraction from I–V characteristics of PV devices. Solar Energy, 85(1), 12-18.

[37] Jordehi, A. R. (2016). Parameter estimation of solar photovoltaic (PV) cells: A review. *Renewable and Sustainable Energy Reviews*, 61, 354-371.

[38] Silvestre, S., Boronat, A., & Chouder, A. (2009). Study of bypass diodes configuration on PV modules. *Applied Energy*, *86*(9), 1632-1640.

[39] Long, B. U. N. (2011). *'Détection et Localisation de Défauts pour un Système PV* (Doctoral dissertation, PhD thesis Electrical Engineering, University of Grenoble).

[40] Stapleton, G., & Neill, S. (2012). Grid-connected solar electric systems: the earthscan expert handbook for planning, design and installation. Routledge.

[41] Häberlin, H. (2012). Photovoltaics: system design and practice. John Wiley & Sons.

[42] Singh, P., Shukla, N., & Gaur, P. (2020). Modified variable step incremental-conductance MPPT technique for photovoltaic system. International Journal of Information Technology, 1-8.

[43] Shimakage, T., Nishioka, K., Yamane, H., Nagura, M., & Kudo, M. (2011, October). Development of fault detection system in PV system. In 2011 IEEE 33rd International Telecommunications Energy Conference (INTELEC) (pp. 1-5). IEEE.

[44] Kaabeche, A., Belhamel, M., & Ibtiouen, R. (2011). Sizing optimization of grid-independent hybrid photovoltaic/wind power generation system. Energy, 36(2), 1214-1222.

[45] G. Heilscher, M. F. (1995). On-line Diagnostic for 100 Photovoltaic Systems within de « German-1000-Roofs-Programme». European photovoltaic solar energy conference, pp 953- 956.

[46] Pillai, D. S., & Rajasekar, N. (2018). A comprehensive review on protection challenges and fault diagnosis in PV systems. *Renewable and Sustainable Energy Reviews*, *91*, 18-40.

[47] Dhoke, A., Sharma, R., & Saha, T. K. (2018). PV module degradation analysis and impact on settings of overcurrent protection devices. *Solar Energy*, *160*, 360-367.

[48] Köntges, Marc, Sarah Kurtz, C. E. Packard, Ulrike Jahn, Karl A. Berger, Kazuhiko Kato, Thomas Friesen et al. "Review of failures of photovoltaic modules." (2014): 1-140.

[49] Herrmann, W., Wiesner, W., & Vaassen, W. (1997, September). Hot spot investigations on PV modules-new concepts for a test standard and consequences for module design with respect to bypass diodes. In Conference Record of the Twenty Sixth IEEE Photovoltaic Specialists Conference-1997 (pp. 1129-1132). IEEE.

[50] Zhao, Y., & Lyons Jr, R. (2011). Ground-fault analysis and protection in PV arrays. Proc. Photovoltaic Protection, 1-4.

[51] Zhao, Y., & Lyons, R. (2011). Line-line fault analysis and protection in PV arrays. In Technical Report, Photovoltaic Protection, Note 2, Mersen, Newburyport, MA, USA, Issue 1.

[52] Zhao, Y., De Palma, J. F., Mosesian, J., Lyons, R., & Lehman, B. (2012). Line–line fault analysis and protection challenges in solar photovoltaic arrays. *IEEE transactions on Industrial Electronics*, *60*(9), 3784-3795.

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

[54] Sabbaghpur Arani, M., & Hejazi, M. A. (2016). The comprehensive study of electrical faults in PV arrays. *Journal of Electrical and Computer Engineering*, *2016*.

[55] Pei, T., & Hao, X. (2019). A fault detection method for photovoltaic systems based on voltage and current observation and evaluation. Energies, 12(9), 1712.

[56] Akram, M. N., & Lotfifard, S. (2015). Modeling and health monitoring of DC side of photovoltaic array. *IEEE Transactions on Sustainable Energy*, *6*(4), 1245-1253.

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

[58] Chokor, A., El Asmar, M., & Lokanath, S. V. (2016, October). A review of photovoltaic DC systems prognostics and health management: Challenges and Opportunities. In Proceedings of the Annual Conference of the Prognostics and Health Management Society, Denver, CO, USA (pp. 3-6).

[59] Karatepe, E., & Hiyama, T. (2011, September). Controlling of artificial neural network for fault diagnosis of photovoltaic array. In *2011 16th International Conference on Intelligent System Applications to Power Systems* (pp. 1-6). IEEE.

[60] Tadj, M., Benmouiza, K., Cheknane, A., & Silvestre, S. (2014). Improving the performance of PV systems by faults detection using GISTEL approach. *Energy conversion and management*, *80*, 298-304.

[61] Bonsignore, L., Davarifar, M., Rabhi, A., Tina, G. M., & Elhajjaji, A. (2014). Neuro-fuzzy fault detection method for photovoltaic systems. *Energy Procedia*, *62*, 431-441.

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

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

[64] Firth, S. K., Lomas, K. J., & Rees, S. J. (2010). A simple model of PV system performance and its use in fault detection. Solar energy, 84(4), 624-635.

[65] Zhao, Y. (2015). Fault detection, classification and protection in solar photovoltaic arrays. Northeastern University. (Doctoral Dissertation).

[66] Stellbogen, D. (1993, May). Use of PV circuit simulation for fault detection in PV array fields. In Conference Record of the Twenty Third IEEE Photovoltaic Specialists Conference-1993 (Cat. No. 93CH3283-9) (pp. 1302-1307). IEEE.

[67] Miwa, M., Yamanaka, S., Kawamura, H., Ohno, H., & Kawamura, H. (2006, May). Diagnosis of a power output lowering of PV array with a (-dI/dV)-V characteristic. In 2006 IEEE 4th World Conference on Photovoltaic Energy Conference (Vol. 2, pp. 2442-2445). IEEE.

[68] Hernday, P. (2011). Field Applications for IV Curve Tracers. SolarPro,(4.5).

[69] Drews, A., De Keizer, A. C., Beyer, H. G., Lorenz, E., Betcke, J., Van Sark, W. G. J. H. M., ... & Heinemann, D. (2007). Monitoring and remote failure detection of grid-connected PV systems based on satellite observations. Solar energy, 81(4), 548-564.

[70] International Electrotechnical Commission. (1998). Photovoltaic system performance monitoring-guidelines for measurement, data exchange and analysis. IEC 61724.

[71] Haeberlin, H., & Beutler, C. (1995, October). Normalized representation of energy and power for analysis of performance and on-line error detection in PV-systems. In Proc. 13th EU PV Conf., Nice (p. 934).

[72] Marion, B., Adelstein, J., Boyle, K. E., Hayden, H., Hammond, B., Fletcher, T., ... & Townsend, T. (2005, January). Performance parameters for grid-connected PV systems. In Conference Record of the Thirty-first IEEE Photovoltaic Specialists Conference, 2005. (pp. 1601-1606). IEEE.

[73] Blaesser, G., & Munro, D. (1993). Guidelines for the assessment of photovoltaic plants. Analysis and presentation of monitoring data. In ECSC-E C-EAEC.

[74] Chouder, A., & Silvestre, S. (2010). Automatic supervision and fault detection of PV systems based on power losses analysis. *Energy conversion and Management*, *51*(10), 1929-1937.

[75] Silvestre, S., Chouder, A., & Karatepe, E. (2013). Automatic fault detection in grid connected PV systems. Solar energy, 94, 119-127.

[76] Chouder, A., Silvestre, S., Taghezouit, B., & Karatepe, E. (2013). Monitoring, modelling and simulation of PV systems using LabVIEW. Solar Energy, 91, 337-349.

[77] Silvestre, S., da Silva, M. A., Chouder, A., Guasch, D., & Karatepe, E. (2014). New procedure for fault detection in grid connected PV systems based on the evaluation of current and voltage indicators. Energy Conversion and Management, 86, 241-249.

[78] Schirone, L., Califano, F. P., Moschella, U., & Rocca, U. (1994, December). Fault finding in a 1 MW photovoltaic plant by reflectometry. In Proceedings of 1994 IEEE 1st World Conference on Photovoltaic Energy Conversion-WCPEC (A Joint Conference of PVSC, PVSEC and PSEC) (Vol. 1, pp. 846-849). IEEE.

[79] Takashima, T., Yamaguchi, J., Otani, K., Kato, K., & Ishida, M. (2006, May). Experimental studies of failure detection methods in PV module strings. In 2006 IEEE 4th World Conference on Photovoltaic Energy Conference (Vol. 2, pp. 2227-2230). IEEE.

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

[81] Smith, P., Furse, C., & Gunther, J. (2005). Analysis of spread spectrum time domain reflectometry for wire fault location. IEEE sensors journal, 5(6), 1469-1478.

[82] Alam, M. K., Khan, F., Johnson, J., & Flicker, J. (2013, September). PV ground-fault detection using spread spectrum time domain reflectometry (SSTDR). In 2013 IEEE energy conversion congress and exposition (pp. 1015-102). IEEE.

[83] Ancuta, F., & Cepisca, C. (2011). Failure analysis capabilities for PV systems. Recent Res. Energy, Environ. Entrep. Innov, 109-115.

[84] Spagnolo, G. S., Del Vecchio, P., Makary, G., Papalillo, D., & Martocchia, A. (2012, May). A review of IR thermography applied to PV systems. In 2012 11th International Conference on Environment and Electrical Engineering (pp. 879-884). IEEE.

[85] Yang, B. S., Oh, M. S., & Tan, A. C. C. (2009). Fault diagnosis of induction motor based on decision trees and adaptive neuro-fuzzy inference. Expert Systems with Applications, 36(2), 1840-1849.

[86] Onduleur connecté au Réseau Solis-Mini 700 www.ginlong.com.

[87] Capteur de Courant ACS712 Http://www.alldatasheet.com/view.jsp?Searchword=Acs712%20datasheet&gclid=EAIaIQobChMIwIilr4b60wIVxJ8bCh2jggPmEAAYASAAEgJBPvD_BwE. [Dernière consultation du site est le 27/6/2021].

[88] Capteur de Tension LEV LM 25-P Http://www.alldatasheet.com/view.jsp?Searchword=Lv25p&gclid=EAIaIQobChMIoerMwIf60wIVEY0bCh2tRg6mEAAYASAAEgLLHfD_BwE. [Dernière consultation du site est le 27/6/2021].

[89] Capteur de l'Irradiation SPEKTRON 200 http://www.tritec-energy.com/en/control-and-measuring-instruments/spektron-irradiation-sensors-c-66/. [Dernière consultation du site est le 27/6/2021].

[90] Capteur de Température LM35 www.ti.com/lit/ds/symlink/lm35.pdf. [Dernière consultation du site est le 27/6/2021].

[91] http://sine.ni.com/psp/app/doc/p/id/psp-117/lang/en. [Dernière consultation du site est le 27/6/2021].

[92] Jordehi, A. R. (2016). Parameter estimation of solar photovoltaic (PV) cells: A review. Renewable and Sustainable Energy Reviews, 61, 354-371.

[93] El-Naggar, K. M., AlRashidi, M. R., AlHajri, M. F., & Al-Othman, A. K. (2012). Simulated annealing algorithm for photovoltaic parameters identification. Solar Energy, 86(1), 266-274.

[94] Bell, A. D., & Bryan, A. (2008). Plant form: an illustrated guide to flowering plant morphology. Timber Press.

[95] Cronquist, A., & Takhtadzhi͡an, A. L. (1981). An integrated system of classification of flowering plants. Columbia university press.

[96] Glover, B. J. (2007). Understanding flowers and flowering: an integrated approach (Vol. 277). Oxford: Oxford University Press.

[97] Yang, X. S. (2012, September). Flower pollination algorithm for global optimization. In International conference on unconventional computing and natural computation (pp. 240-249). Springer, Berlin, Heidelberg.

[98] Yang, X. S. (2010). Nature-inspired metaheuristic algorithms. Luniver press.

[99] Wagner, A. (2000, June). Peak-power and internal series resistance measurement under natural ambient conditions. In Proceedings EuroSun (Vol. 5).

[100] Louzazni, M., Khouya, A., Amechnoue, K., Gandelli, A., Mussetta, M., & Crăciunescu, A. (2018). Metaheuristic algorithm for photovoltaic parameters: comparative study and prediction with a firefly algorithm. Applied Sciences, 8(3), 339.

[101] Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized Simulated Annealing for Global Optimization: The GenSA Package. R Journal, 5(1).

[102] El-Naggar, K. M., AlRashidi, M. R., AlHajri, M. F., & Al-Othman, A. K. (2012). Simulated annealing algorithm for photovoltaic parameters identification. Solar Energy, 86(1), 266-274.

[103] Rawat, R., Kaushik, S. C., & Lamba, R. (2016). A review on modeling, design methodology and size optimization of photovoltaic based water pumping, standalone and grid connected system. Renewable and Sustainable Energy Reviews, 57, 1506-1519.

[104] Kratochvil, J. A., Boyson, W. E., & King, D. L. (2004). Photovoltaic array performance model (No. SAND2004-3535). Sandia Laboratories.

[105] De Soto, W., Klein, S. A., & Beckman, W. A. (2006). Improvement and validation of a model for photovoltaic array performance. Solar energy, 80(1), 78-88.

[106] Olukan, T. A., & Emziane, M. (2014). A comparative analysis of pv module temperature models. Energy Procedia, 62, 694-703.

[107] I. H. Witten and E. Frank. (2000). Data Mining: Practical Machine Learning Tools and Techniques: Morgan Kaufmann.

[108] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

[109] Mitchell, T. (1997). Machine Learning. McGraw-Hill.

[110] Tan, P. N., Steinbach, M., & Kumar, V. (2008). Introduction to data mining.

[111] I. H. Witten, E. Frank: 'Data Mining: Practical Machine Learning Tools and Techniques', Morgan Kaufmann, 2 édition, 2005. (Livre).

[112] Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81-106.

[113] Morgan, J. N., & Messenger, R. C. (1973). THAID, a sequential analysis program for the analysis of nominal scale dependent variables.

[114] Quinlan, J. R. (1993). C4. 5: programs for machine learning. Morgan kaufmann publishers. (Livre).

[115] Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. Journal of the Royal Statistical Society: Series C (Applied Statistics), 29(2), 119-127.

[116] Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees, Wadsworth International Group, Belmont, California (1984).

[117] Quinlan, J. R. (1979). Discovering rules by induction from large collections of examples. Expert systems in the microelectronics age.

[118] Help Center of Mathwork « Classification » 'https://www.mathworks.com/help/stats/classification-example.html'

[119] Hunt, E. B. (1962). Concept learning: An information processing problem. JW\. S. Inc, Ed.

[120] Lavoie, B. (2006). Arbres de décisions–Synthèse de lectures–Séminaire sur l'apprentissage automatique. Programme de Doctorat en Informatique Cognitive, Université du Québec à Montréal, 15.

Annexe A

Exemple

Un exemple sera introduit pour plus de précisions sur la construction d'un modèle AD par l'algorithme C4.5. L'ensemble d'apprentissage de l'exemple est présenté dans le tableau 1.

Tableau 1 Ensemble d'apprentissage pour l'exemple d'algorithme C4.5.

V_{dc} (V)	I_{dc} (A)	Etat de système
80.65	5.15	**Etat sain**
78.92	6.3	**Etat sain**
79	5.08	**Etat sain**
66.1	3.3	**Etat défectueux**
79.9	1.3	**Etat défectueux**
60.3	2.42	**Etat défectueux**

Par conséquent, le modèle est créé en suivant les étapes suivantes:

Étape 1: l'entropie de la classe est calculée à l'aide du tableau des fréquences de l'étiquette de classe D (tableau 2), où D est l'état du système dans ce cas;

Tableau 2 Le tableau des fréquences de la classe D.

Etat de système (D)		Nombre total d'instances
Etat sain	**Etat défectueux**	
3	3	6

$$Info(D) = Info(3,3) = -\frac{3}{6}\log_2(\frac{3}{6}) - \frac{3}{6}\log_2(\frac{3}{6})$$

$$Info(D) = 1bit$$

Étape 2: l'organisation dans un ordre croissant les valeurs de chaque attribut;

Étape 3: Calculez le rapport de gain pour chaque valeur d'attribut séparément pour choisir le meilleur attribut de division.

➢ **Pour l'attribut Vdc:**

Tableau 3 Calcul de l'information pour l'attribut Vdc.

60.3		**66.1**		**78.92**		**79**		**79.9**		**80.65**		**V_{dc} (V)**
≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	
0	3	0	3	1	2	2	1	2	1	3	0	Etat sain
1	2	2	1	2	1	2	1	3	0	3	0	Etat défectueux
0.8091		**0.5409**		0.9183		1		0.8091		1		*Info (bits)*
0.1909		**0.4591**		0.0817		0		0.1909		0		*Gain*
0.65		**0.9183**		1		0.9183		0.65		1		*Split Info*
0.2936		**0.5**		0.0817		0		0.2936		0		*Rapport de Gain*

Le tableau 3 donne les résultats du rapport de gain pour chaque valeur d'attribut de Vdc, notez que chaque valeur de Vdc est considérée comme un seuil et les critères de division sont appliqués pour calculer le rapport de gain pour ces seuils, puis la valeur d'attribut qui a un gain maximum ratio est choisi. De plus, comme mentionné dans le chapitre 4 section 5, les expressions 4.9 à 4.13 sont utilisées pour montrer le processus de construction du modèle, dont le calcul est le suivant:

Pour Vdc = 60,3 V:

$$Info(V_{dc},D)=\frac{1}{6}\times Info\left[V_{dc}\le 60{,}3\right]+\frac{5}{6}\times Info\left[V_{dc}>60{,}3\right]$$

$$=\frac{1}{6}\times\left[-\frac{0}{1}\times\log_2(\frac{0}{1})-\frac{1}{1}\times\log_2(\frac{1}{1})\right]-\frac{5}{6}\times\left[-\frac{3}{5}\times\log_2(\frac{3}{5})-\frac{2}{5}\times\log_2(\frac{2}{5})\right]=0.8091$$

$$Gain(V_{dc},D)=Info(D)-Info(V_{dc},D)=1-0.8091=0.1909$$

$$SplitInfo=-\frac{1}{6}\times\log_2(\frac{1}{6})-\frac{5}{6}\times\log_2(\frac{5}{6})=0.65$$

$$Rapport_de_Gain=\frac{Gain\left(V_{dc},D\right)}{Split_Info\left(V_{dc},D\right)}=\frac{0.1909}{0.65}=0.2936$$

Comme décrit dans le tableau 3, le seuil qui a le grand rapport de gain est Vdc = 66,1 V, alors cette valeur est prise comme meilleure fractionnée par l'attribut Vdc, la même procédure est effectuée pour l'attribut Idc.

- **Pour l'attribut Idc:**

Comme il est remarqué dans le tableau 4, cet attribut a un seuil de partage parfait dans Idc = 3,3 A, où pour cette valeur le rapport de gain est égal à un.

Tableau 4 Calcul de l'information pour l'attribut Idc.

1.3		**2.42**		**3.3**		**5.08**		**5.15**		**6.08**		**I_{dc} (A)**
≤	>	≤	>	≤	>	≤	>	≤	>	≤	>	
0	3	0	3	0	3	1	2	2	1	3	0	Etat sain
1	2	2	1	3	0	3	0	3	0	3	0	Etat défectueux
0.8091		0.5409		**0**		0.5409		0.8091		1		*Info (bits)*
0.1909		0.4591		**1**		0.4591		0.1909		0		*Gain*
0.65		0.9183		**1**		0.9183		0.65		1		*Split Info*
0.2936		0.5		**1**		0.5		0.2936		0		*Rapport de Gain*

Étape 4: Choisissez le meilleur attribut de fractionnement

Tableau 5 Les meilleures valeurs des seuils pour chaque attribut.

L'attribut	V_{dc} (V)	I_{dc} (A)
Seuil	60.4	**3.3**
Rapport de Gain	0.5	**1**

Le tableau 5 montre la meilleure valeur de rapport de gain pour chaque attribut, où les seuils d'attribut présentés ont le grand rapport de gain par rapport aux autres valeurs. Dans cette étape, l'attribut qui a le rapport de gain maximal par rapport aux autres attributs est sélectionné pour être placé en tant que un nœud racine de l'arbre. De plus, selon le tableau

5, l'attribut qui a le gain de rapport le plus élevé est l'attribut **Idc**. Par conséquent, cet attribut présente le nœud racine comme illustré dans la figure 1.

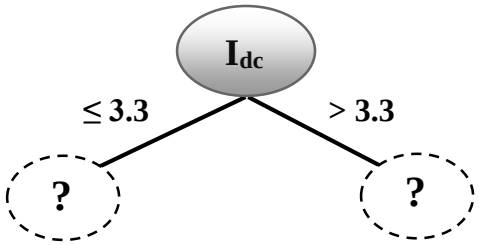


Figure 1 Le nœud racine de l'exemple de modèle AD.

De plus, le nœud racine a deux branches droite et gauche, où les arêtes sous la branche sont respectivement Idc> 3,3 A et Idc ≤ 3,3 A, une condition est faite pour décider si le nœud suivant est une feuille ou un nœud interne, si le la classe majoritaire sur ce nœud est supérieure ou égale à deux, alors ce nœud est considéré comme une feuille. De plus, la branche gauche possède une classe majoritaire et elle est supérieure à 2 (voir le tableau 5 pour la valeur Idc 3.3 A), donc cette branche se termine par l'étiquette d'état défectueux dans la feuille comme illustré à la figure 2. La même condition est appliquée sur la branche droite, il ressort clairement du tableau 5 de la colonne Idc> 3.3 que la classe majoritaire supérieure à 2 est l'état sain, donc la branche droite se termine par cette étiquette d'état dans la feuille, voir la figure 3.

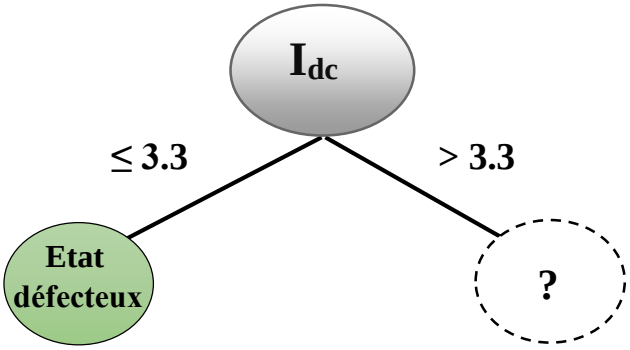


Figure 2 Le nœud racine avec une feuille gauche.

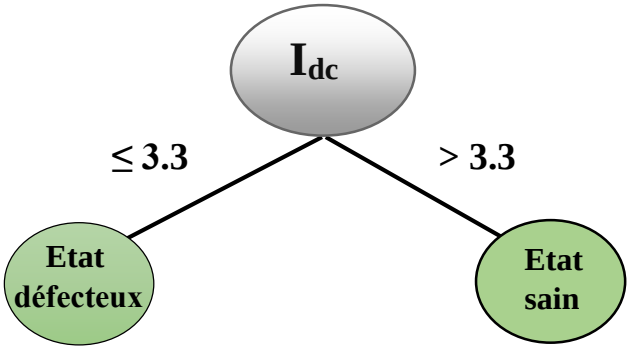


Figure 3 Le modèle final de l'arbre.

Ce modèle peut être traduit en règles de décision:

Si Idc ≤ 3.3 **alors** la décision est état défectueux, **sinon** Idc> 3.3 **alors** la décision est état sain.

Étape 5: le processus d'élagage est appliqué après la construction de l'arbre du modèle à l'aide de l'expression d'élagage 4.14; la même procédure est appliquée dans cette étape.

- L'erreur attendue pour les deux feuilles, la feuille droite N = 3 et E = 0 donc $e_d(3{,}0/3, \alpha)$= 0,14, pour la feuille gauche N = 3 et E = 0 donc $e_g(3{,}0/3, \alpha)$ = 0,14;
- L'erreur globale attendue de ces feuilles est: $3 \times e_r + 3 \times e_l = 0.84$;
- Pour calculer l'erreur prédite à partir du nœud racine de ce sous-arbre, alors que le nœud racine considéré comme une feuille avec l'étiquette de la classe majoritaire atteint par ce nœud, pour $e(6{,}3/6, \alpha) = 0.64$ donc l'erreur globale prédite par ce nœud est $6 \times e_g = 3.81$.

En comparant les deux erreurs, nous avons trouvé que l'erreur globale obtenue par le nœud racine est supérieure à la somme des erreurs obtenues par les feuilles. Par conséquent, l'arbre est conservé dans sa forme actuelle.

Application pratique :

Cette fois on utilise le logiciel WEKA pour construire le modèle. L'algorithme C4.5 dans WEKA nommé J48 qui est la 8ème version de l'algorithme C4.5.

Les données d'apprentissage sont enregistrées dans fichier d'extension Arff et se la forme suivant :

```
@relation Données_de_l_exemple
@attribute Vdc numeric
@attribute Idc numeric
@attribute Classe {Etat_sain, Etat_défectueux }
@data
% Vdc ,  Idc  , la classe
80.65,5.15,Etat_sain
78.92,6.3,Etat_sain
79,5.08,Etat_sain
66.1,3.3,Etat_défectueux
79.9,1.3,Etat_défectueux
60.3,2.42,Etat_défectueux
```

Pour construire le modèle on suivre les étapes suivants

1. D'abord, nous ouvrons le logiciel WEKA

2. Ensuite, nous choisissons Explorer comme indique sur la figure 4



Figure 4 Interface frontal de logiciel WEKA.

3. On clique sur **ouvrir un fichier** et on choisit l'emplacement de fichier de données souhaitable

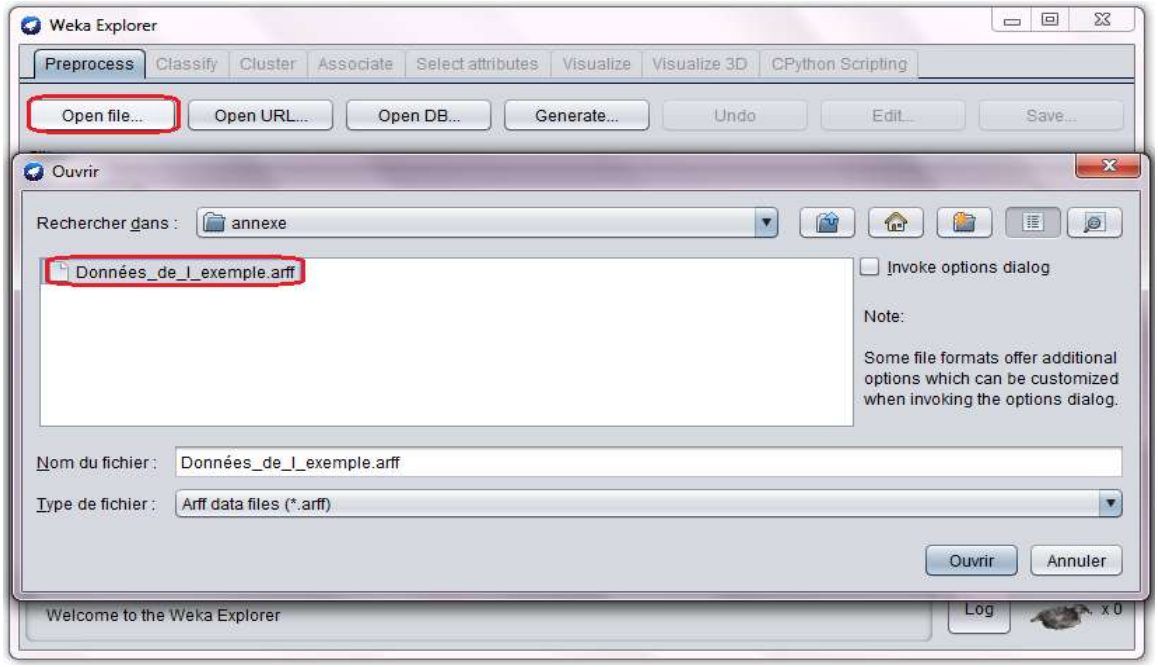


Figure 5 Fenêtre de logiciel WEKA (pour ouvrir le fichier de données).

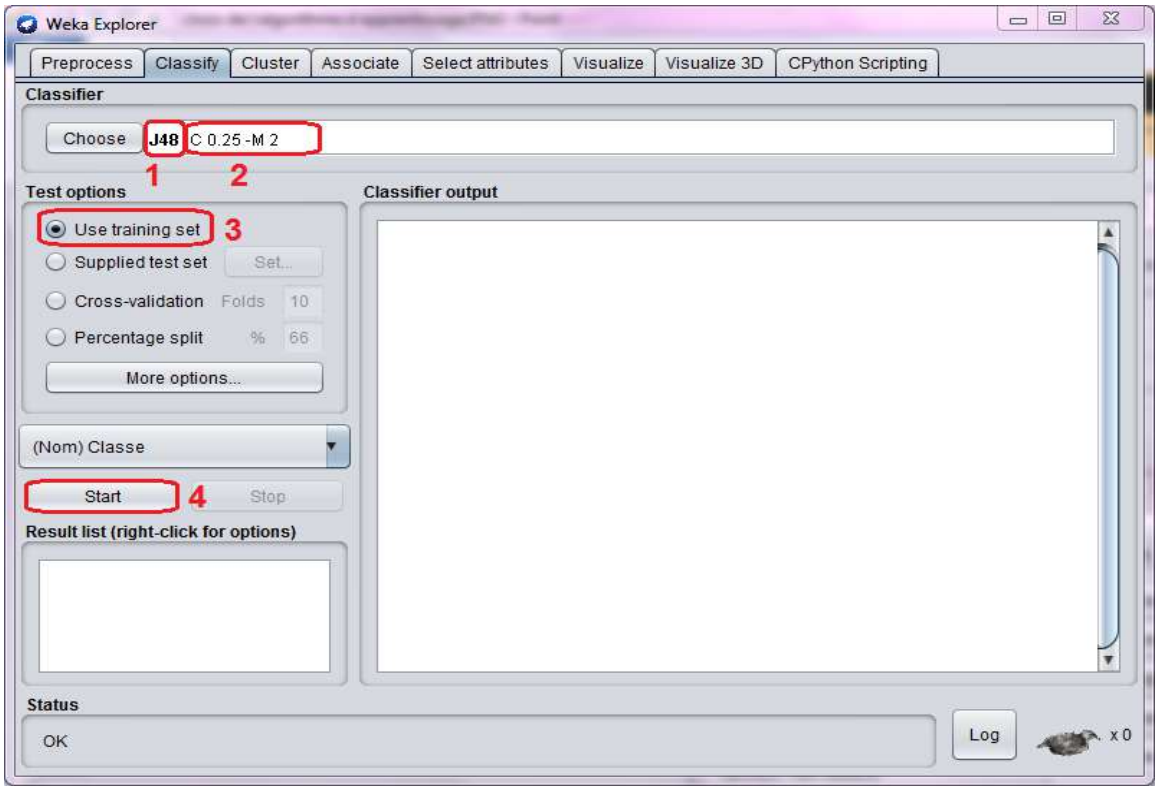


Figure 6 Fenêtre de choix de l'algorithme avec ses paramètres et le mode de test dans le software WEKA.

D'où : **1** Le nom de l'algorithme choisit.

2 Les paramètres de l'algorithme.

3 Choix de test dans ce cas les données de l'apprentissage sont les mêmes utilisées pour le test de modèle obtenue

4 Pour démarrer le processus de l'apprentissage.

4. Dès qu'on ouvre le fichier de données la sous-fenêtre de classification devient accessible. On choisit l'algorithme d'apprentissage et leurs paramètres, ainsi que le type de test comme illustré dans la figure 6.

5. Les résultats de l'apprentissage et de test sont montrés dans les figures 7 et 8

```
J48 pruned tree
------------------

Vdc <= 3.3: Etat_defectueux (3.0)
Vdc > 3.3: Etat_sain (3.0)

Number of Leaves  :     2

Size of the tree :      3


Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances           6               100      %
Incorrectly Classified Instances         0                 0      %
Kappa statistic                          1
Mean absolute error                      0
Root mean squared error                  0
Relative absolute error                  0      %
Root relative squared error              0      %
Total Number of Instances                6
```

Figure 7 Les résultats de l'apprentissage et le test.

```
=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000     Etat_sain
                 1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000     Etat_defectueux
Weighted Avg.    1,000    0,000    1,000      1,000    1,000      1,000    1,000     1,000

=== Confusion Matrix ===

 a b   <-- classified as
 3 0 | a = Etat_sain
 0 3 | b = Etat_defectueux
```

Figure 8 La matrice de confusion.

La figure 9 représente le modèle AD obtenue pour cet ensemble de données

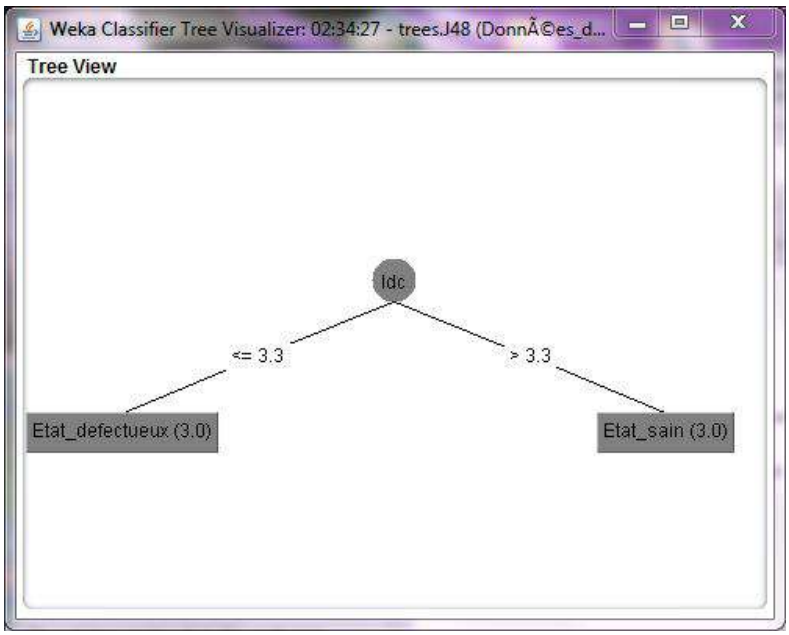
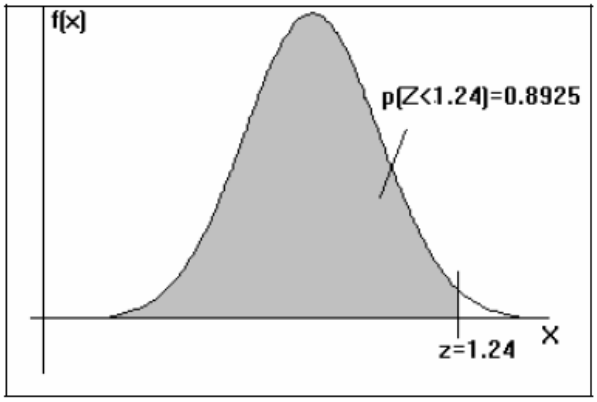


Figure 9 Modèle AD résultant.

On remarque que les modèles obtenues dans le deux cas (théorique et par WEKA) sont identique.

TABLE DE LA LOI NORMALE CENTREE REDUITE

Lecture de la table: Pour z=1.24 (intersection de la ligne 1.2 et de la colonne 0.04), on a la proportion P(Z < 1,24) = 0.8925



P(Z > 1,96) = 0,025
P(Z > 2,58) = 0,005
P(Z > 3,29) = 0,0005

Rappels:

1/ P(Z > z) = 1 - P(Z < z) et 2/ P(Z < -z) = P(Z > z)

Exemple: Sachant P(Z < 1,24) = 0,8925, on en déduit:

1/ (P(Z > 1,24) = 1 - P(Z < 1,24) = 1- 0,8925 = 0,1075

2/ P(Z < -1,24) = P(Z > 1,24) = 0,1075

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,99865	0,99869	0,99874	0,99878	0,99882	0,99886	0,99889	0,99893	0,99896	0,99900
3,1	0,99903	0,99906	0,99910	0,99913	0,99916	0,99918	0,99921	0,99924	0,99926	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,99940	0,99942	0,99944	0,99946	0,99948	0,99950
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,99960	0,99961	0,99962	0,99964	0,99965
3,4	0,99966	0,99968	0,99969	0,99970	0,99971	0,99972	0,99973	0,99974	0,99975	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,99980	0,99981	0,99981	0,99982	0,99983	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,99987	0,99987	0,99988	0,99988	0,99989
3,7	0,99989	0,99990	0,99990	0,99990	0,99991	0,99991	0,99992	0,99992	0,99992	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,99994	0,99994	0,99995	0,99995	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	0,99996	0,99996	0,99996	0,99997	0,99997
4,0	0,99997	0,99997	0,99997	0,99997	0,99997	0,99997	0,99998	0,99998	0,99998	0,99998

Tableau 6 Table de la loi Normale Centré Réduite.